IA Title: Scalable System Packet Interface Implementation Agreement: System Packet Interface Capable of Operating as an Adaptation Layer for Serial Data Links

IA # OIF-SPI-S-01.0

November 17, 2006

Implementation Agreement created and approved
by the Optical Internetworking Forum
www.oiforum.com

The OIF is an international non-profit organization with over 100 member companies, including the world's leading carriers and vendors. Being an industry group uniting representatives of the data and optical worlds, OIFs purpose is to accelerate the deployment of interoperable, cost-effective and robust optical internetworks and their associated technologies. Optical internetworks are data networks composed of routers and data switches interconnected by optical networking elements.

With the goal of promoting worldwide compatibility of optical internetworking products, the OIF actively supports and extends the work of national and international standards bodies. Formal liaisons have been established with The ATM Forum, IEEE 802.3, IETF, ITU-T Study Group 13, ITU-T Study Group 15, MEF, NPF, T1M1, T1X1, TMF, UXPi and the XFP MSA Group.

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

www.oiforum.com

Notice: This Technical Document has been created by the Optical Internetworking Forum (OIF). This document is offered to the OIF Membership solely as a basis for agreement and is not a binding proposal on the companies listed as resources above. The OIF reserves the rights to at any time to add, amend, or withdraw statements contained herein. Nothing in this document is in any way binding on the OIF or any of its members.

The user's attention is called to the possibility that implementation of the OIF implementation agreement contained herein may require the use of inventions covered by the patent rights held by third parties. By publication of this OIF implementation agreement, the OIF makes no representation or warranty whatsoever, whether expressed or implied, that implementation of the specification will not infringe any third party rights, nor does the OIF make any representation or warranty whatsoever, whether expressed or implied, with respect to any claim that has been or may be asserted by any third party, the validity of any patent rights related to any such claim, or the extent to which a license to use any such rights may or may not be available or the terms hereof.

Working Group: Physical and Link Layer Working Group

Title: **Scalable System Packet Interface (SPI-S)**
System Packet Interface capable of operating as an adaptation layer for serial data links using OIF CEI protocol or over serial data links using 64B/66B framing.

Source:

Technical Editor
Christopher Ebeling
Xilinx, Inc
2100 Logic Drive
San Jose, CA
USA
Phone: +1 408 879-4849
Email: christopher.ebeling@xilinx.com

Working Group Chair
David R. Stauffer
IBM Corp.
1000 River Road MS862J
Essex Junction, VT
USA
Phone: +1 802 769-6914
Email: dstauffe@us.ibm.com

Date: November 17, 2006

Abstract: This specification defines the Scalable System Packet Interface implementation agreement (SPI-S). The interface supports the transfer of network traffic between a pair of network processing elements (NPEs); these include physical layer (PHY) devices (e.g., SONET framers or mappers), network processors, network coprocessors, and switch fabrics or link layer devices (e.g., Ethernet MACs). The target functionality is passing packets or cells between a network processing element (NPE) and an adjacent physical layer, link layer, coprocessor, or switch fabric device. The interface is not targeted to a particular bandwidth it is intended to support OC-192, OC-786 line rates and beyond. The interface also supports the transfer of status information in band with the forward Payload Data Transfers. This can either be Credit Pool Status information or Signaling Data Transfers, utilizing NPF Messaging.

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

1 Table of Contents

2 LIST OF Tables and Figures

3 Document Revision History

Revision	Date	Description
OIF-SPI-S-01.0	November 2006	IA Text derived from Draft IA of document oif2006.114.07 accepted by OIF Technical Committee principal ballot #46

4 Scope and Purpose

This document specifies the Scalable System Packet Interface (SPI-S), an interface for the interconnection of network processing elements (NPEs). Included in the definition of network processing elements are framers, switch fabrics, network processors, and network coprocessors (e.g., classification or encryption coprocessors).

SPI-S is the successor to both the NPF Streaming Interface (SI) and the OIF System Packet Interface (SPI). The protocol is intended to operate over either 64/66B encoded serial link(s) or OIF CEI Protocol encoded link(s). The interface is not tied to a particular physical interface or any fixed bandwidth and is capable of operating over individual serial links or multi-lane aggregated links. Typical applications may involve the transfer of data between two devices at an aggregate of 10-40Gbps (OC-192 to OC-768 rates), and beyond. It is a point-to-point interface with support for addressing and flow control for multiple framer channels, switch fabric and/or coprocessor data contexts, endpoints and/or classes, as well as multicast traffic.

The SPI-S specification defines the link-level requirements, including data framing and packet delineation, flow control, address formats, and error detection. Configuration of SPI-S may be implemented either in-band or out of band. The interface is defined to operate in a self-healing manner in the face of errors through the use of soft state, but does not include a retry recovery mechanism for payload protection.

4.1 System Block Diagram

Figure 4-1, a reference diagram, illustrates a typical application utilizing the SPI-S.

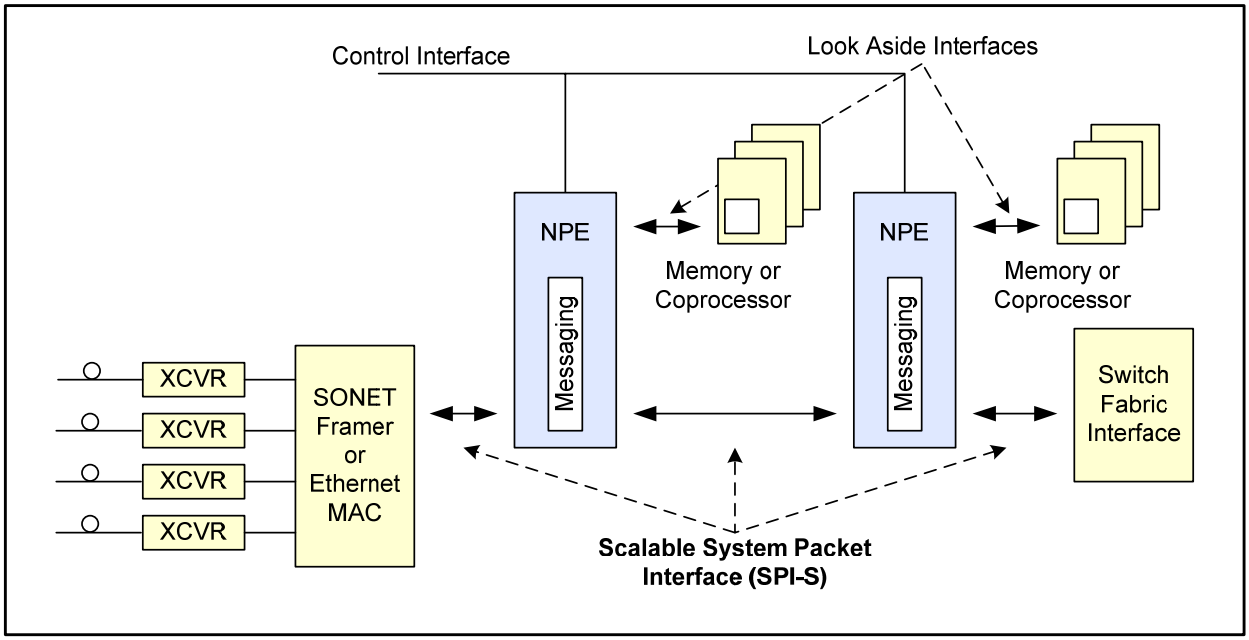


Figure 4-1: System Block Diagram

4.2 References

The following documents contain provisions, which through reference in this text constitute provisions of this specification. At the time of publication, the editions indicated were valid. All referenced documents are subject to revision, and parties implementing to agreements based on this specification are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below.

[1] OIF-SPI5-01.0, System Packet Interface Level 5 (SPI-5): OC-768 System Interface for Physical and Link Layer Devices, Optical Internetworking Forum Implementation Agreement, September 2002[1].

[2] NPF2001.121.25, Network Processing Forum Streaming Interface (NPSI) implementation agreement, October 18, 2002[2].

[3] System Packet Interface Level 4 (SPI-4) Phase 2 Revision 1: OC-192 System Interface for Physical and Link Layer Devices, Optical Internetworking Forum Implementation Agreement, October 15, 2003[3].

[4] NPF2002.429.05, NPF Message Layer Implementation Agreement, November 11, 2003[4].

[5] OIF-CEI-P—1.0, Common Electrical I/O – Protocol (CEI-P), Optical Internetworking Forum Implementation Agreement, Mach 2005[5].

[6] IEEE802.3ae-2002, Media Access Control (MAC) Parameters, Physical Layers, and Management Parameters for 10 Gb/s Operation, August 30, 2002[6].

Portions of the OIF SPI-5 [1] specification and NPF NPSI[2] specification were used to create this document.

[1] At the time of publication of this document, the SPI-5 specification was located at the following http:// address: http://www.oiforum.com/public/documents/OIF-SPI5-01.1.pdf

[2] At the time of publication of this document, the NPSI specification located at the following http:// address: http://www.oiforum.com/public/documents/HWStreamingIA.pdf

[3] At the time of publication of this document, the SPI-4 Phase 2 specification was located at the following http:// address: http://www.oiforum.com/public/documents/OIF-SPI4-2.01v2.pdf

[4] At the time of publication of this document, the NPF Message Layer specification was located at the following http:// address: http://www.oiforum.com/public/documents/Messaging_IA.pdf

[5] At the time of publication of this document, the CEI-P specification was located at the following http:// address: http://www.oiforum.com/public/documents/CEI_P-01.0.pdf

[6] At the time of publication of this document, the IEEE 10 Gbps Ethernet specification was located at the following http:// address: http://standards.ieee.org/getieee802/index.html

4.3 Specification Conventions

This specification uses the following terminology protocol:

• SHALL indicates that the item is a requirement for conformance to the NPSI specification.

• MAY indicates that the item is optional.

• SHOULD indicates that the item is not required by this specification, but is offered as implementation guidance.

In addition:

• Annex A is an integral part of the specification. If there is a conflict between the annex and the text of this document then Annex A takes precedence over the text.

• Figures (except state diagrams) are informative only

• Footnotes used in this specification are informative only.

• A reserved field or bit shall be transmitted as zero and ignored on reception.

4.4 Definitions

The following terms are used throughout this specification:

Block:	A Block (64 bits) is the basic unit of information transfer for the SPI-S protocol and the unit of striping across multiple lanes. There are two Block types, Control Block and Data Block, differentiated by the associated Tag identifier.
Burst:	Data Transfer occurs as Bursts. Burst are always initiated and terminated by a Control Word. The initiating control word contains information about the type and context of the data in the Burst.
Class:	A sub-field of the channel address may be used to discriminate and manage traffic flows. The mechanism for differentiation of services among multiple classes is implementation specific and beyond the scope of this specification.
Control Block:	A Block containing a 32-bit control word plus a 32-bit data word (Quad Data Word).
Data Block:	A Block containing 64 bits of Data (Octal Data Word).
Data Transfer	In this document is a generic term for data transfers that are either a Payload Data Transfer or a Signaling Data Transfer.
Look Aside Interface:	The Look-Aside interface is intended for devices located adjacent to a network processing element (NPE) that offload certain tasks.

Network Processing Element (NPE): A component that is involved in the production, transfer, or processing of data within a network. Examples of NPEs include physical and link layer devices network processors, network coprocessors, network search engines and switch fabrics. SPI-S is used to transfer packet or cell-oriented data streams between adjacent NPEs.

Octal Data Word: The data portion of a Data Block that contains eight data bytes.

Packet: A service data unit that is conveyed via the SPI-S. Packets may contain Payload or Signaling Data.

Packet Termination: Packets are terminated by control word flags. End of Packet, End of Packet with PAD bytes and ABORT terminate packets.

Payload Data: Data on the SPI-S link that is part of a Packet associated with a specific Port.

Pool: An aggregation of ports for flow control.

Port: An addressable endpoint of the SPI-S protocol.

Quad Data Word: The data portion of a Control block that contains data four data bytes.

Status Data: Data that is not specifically Payload Data, e.g. Signaling Data, Credit Pool Status and Idle data are classes of Status Data.

Segment: The payload of a protocol data unit that is transferred across a port. For SPI-S, the smallest message unit is a Segment. Except for end-of-packet (EOP) Segments, Segments have a minimum and maximum size. Payload packets are transferred as one or more Segment Transfers.

Signaling Data: Packets of control information not associated with any port or payload data flow. These non-segmented packets are encoded using NPF Messaging.

Tag Identifier: The Tag bit(s) designate if the associated Block of data is Control Block or a Data Block. The physical format of the Tag field depends on the coding one bit for CEI links, two bits (encoded) for 64/66B links.

5 Architectural Overview

The SPI-S provides for the transfer of data traffic and flow control information between two adjacent network processing devices. The two framing formats that are supported for the data framing are described in the following section.

5.1 SPI-S Reference Model

The SPI-S supports two framing formats, defined below and illustrated in Figure 5-1.

- Mode 1 - 64B/66B: 10 Gigabit Ethernet compatible framing and scrambling. A mandatory mode of operation utilizing 64B/66B coding and scrambling, as defined in IEEE802.3ae-2002:
    - Clause 49.1.4.1 - PCS introduction
    - Clause 49.2 - Physical Coding Sub layer
- Mode 2 – CEI-P: An optional mode of operation utilizing OIF CEI-P framing, scrambling, and FEC. This Description of fields in the Control Word mode is recommended for use with interfaces requiring the use of Forward Error Correction (FEC).

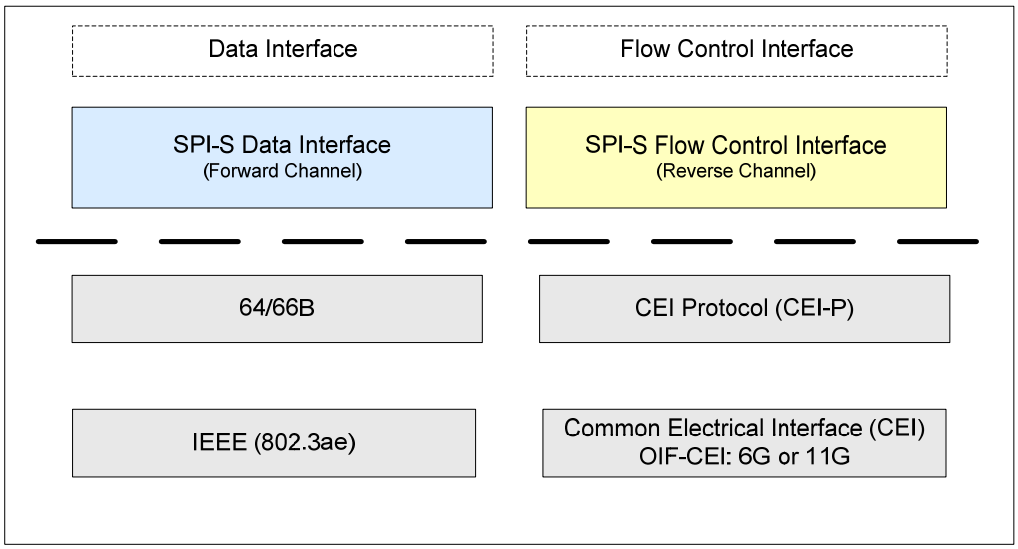


Figure 5-1: Reference Model

In Figure 5-1 the relationships to IEEE802.3ae and OIF CEI physical data links are shown as examples of typical applications, however use of either of these links is neither mandated nor presumed by this implementation agreement. As a scalable protocol, SPI-S is capable of operating over any physical link(s) that can support 64B/66B or CEI-P encoding. For example, it is also possible to use 64B/66B encoding over OIF SxI-5 links or CEI links or XAUI links. Likewise, it is also possible to use CEI-P encoding over SxI-5 links or XAUI links.

5.2 Features of SPI-S

The SPI-S has the following characteristics:

- It can efficiently transfer common packet formats, such as:
    - Ethernet packets (≥ 64 bytes)
    - ATM cells (48 or 52 bytes)
    - IP packets (≥ 40 bytes)
- It has provisions for a control packet context (utilizing NPF Messaging).
- It can operate very efficiently over both 64B/66B links and CEI-P links.
- The logical operation is based on the OIF SPI-5 specification.
- SPI-5 Channel based flow control
- SPI-S Links maybe uni-directional, bi-directional or asymmetric

5.3 Interface Description

An SPI-S interface is not limited to a particular physical interface. The physical format can use any number of links operating at any application specific frequency. Information transfer may be uni-directional as well as bi-directional. Bi-directional links need not be symmetric, as shown in Figure 5-4. In a full-duplex implementation, both the media side and system side interfaces carries a forward channel (Payload Data) and optionally reverse channel (Signaling) information associated with received and transmitted packets as illustrated in Figure 5.3. Payload Data Transfers are delimited by Start and End of Packet Control Words. These packets may be further divided into segments and flow control information for the reverse channel may be optionally interleaved with the Payload Data.

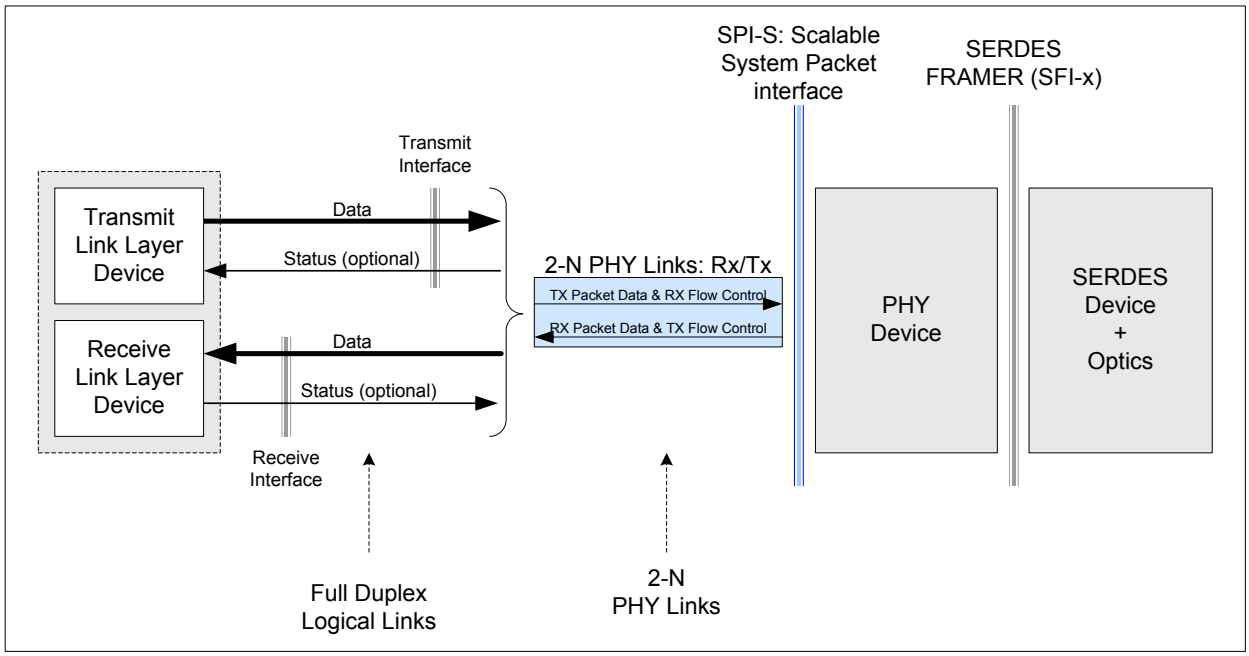


Figure 5-2: SPI-S Bi-directional

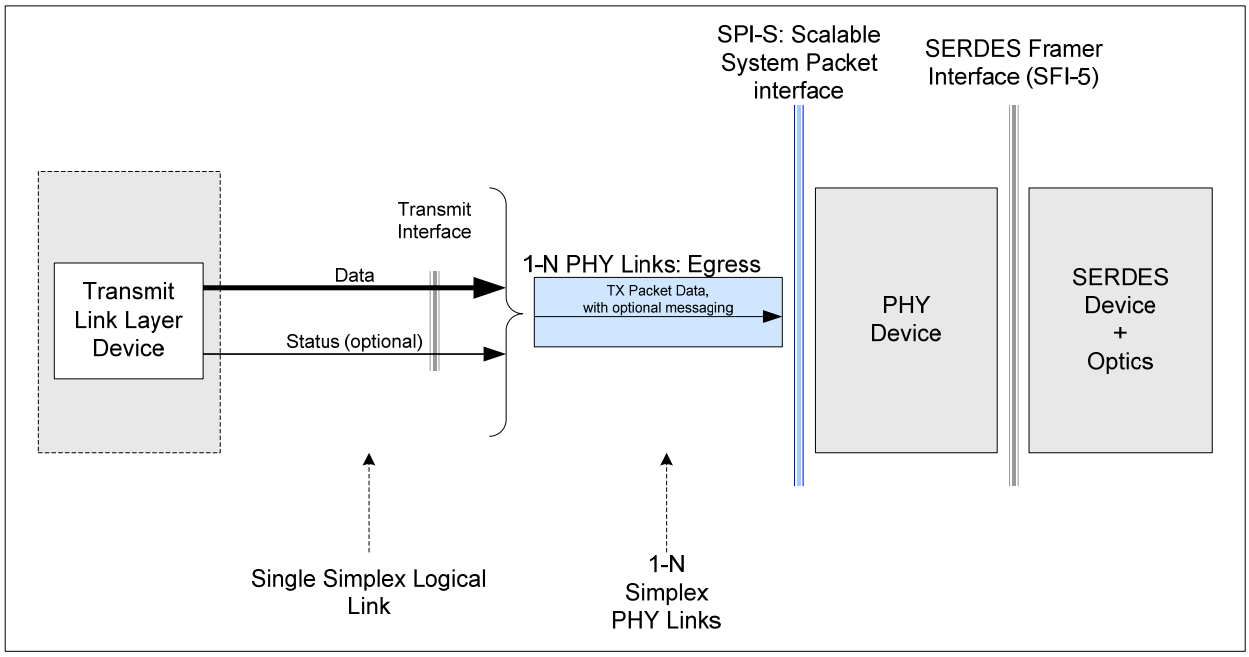


Figure 5-3: SPI-S Uni-directional

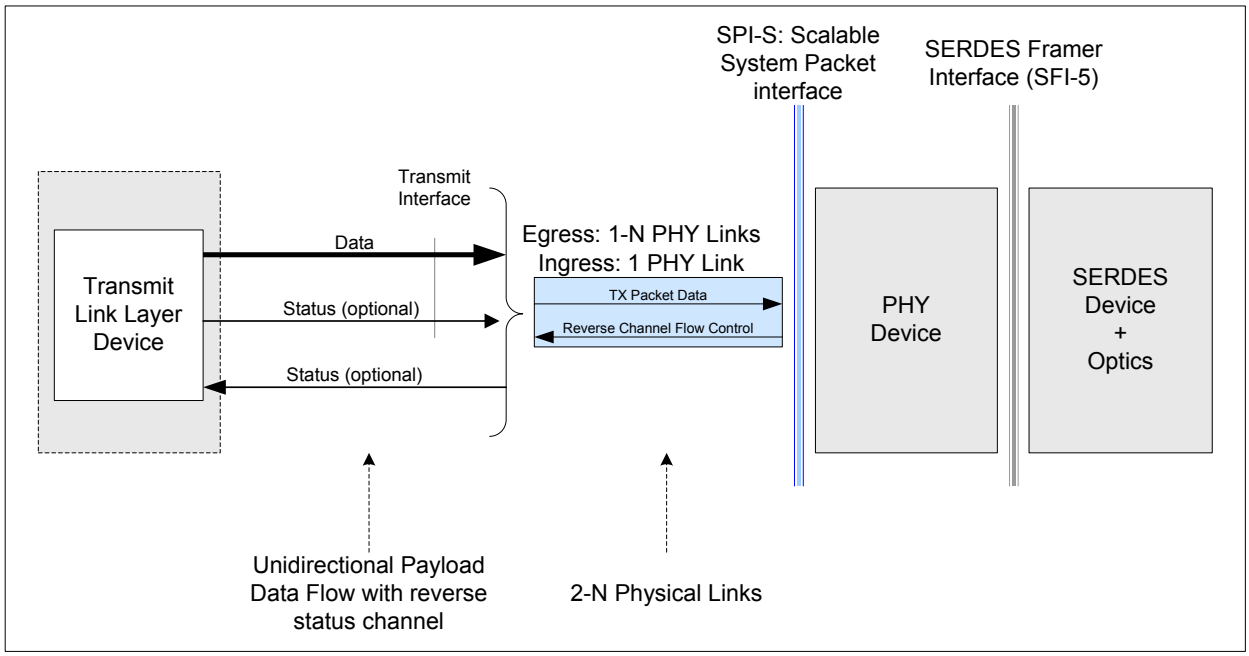


Figure 5-4: SPI-S Asymmetric

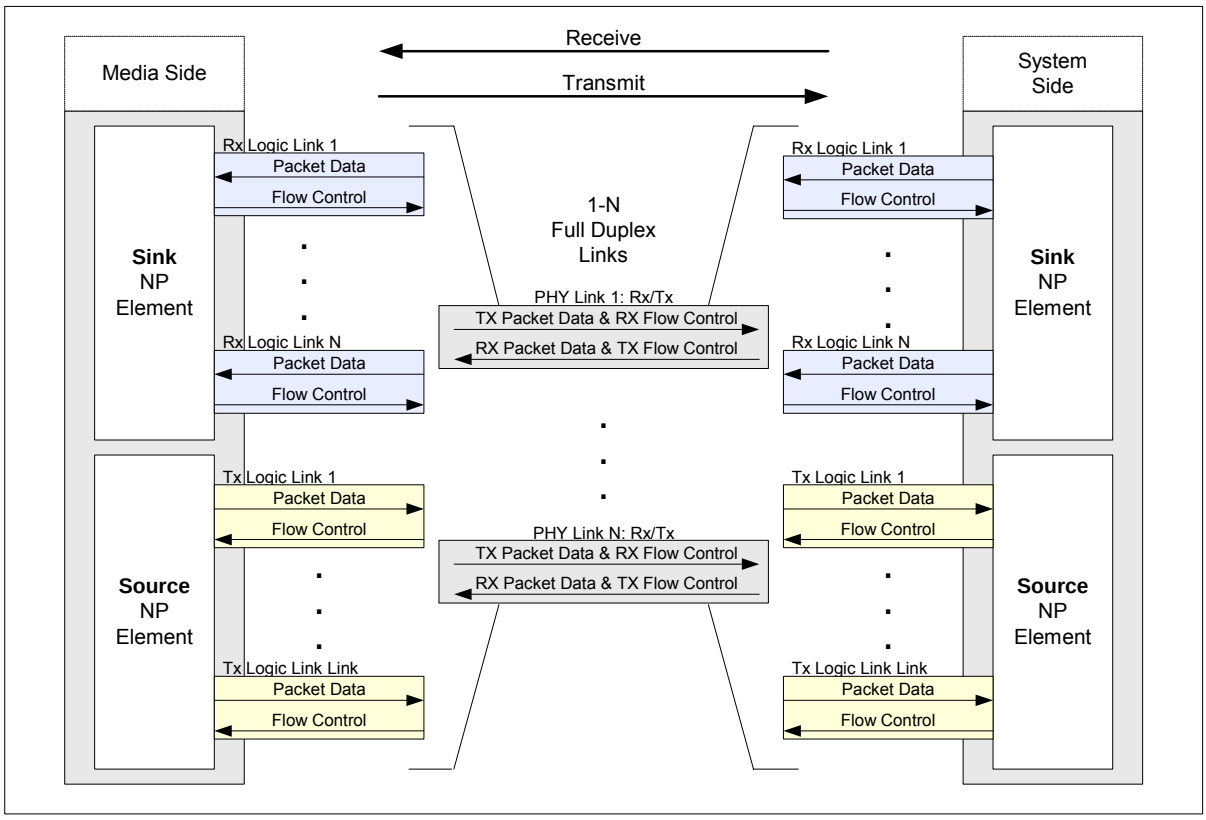


Figure 5-5: Multiple Physical Links between NPEs

6 DATA PATH OPERATION

Data is transferred over the SPI-S bus in Bursts. A Burst is always delimited by control words. All the data in a burst shall be of the same transfer type (Payload Data or Signal Data). Payload Data Transfers within a Burst are associated with the channel addressed by the initiating Control Word. The Burst data may be Payload Data Transfer (PDT) with an associated address, or a Signaling Data Transfer (SDT) with no associated address, or Pool Status (in band, reverse channel flow control information), or Idle data (null data). Payload Data Transfers and Signal Data Transfers are packet based, with Start of Packet and End of Packet information. The Start Of Packet is always aligned to the beginning of Burst. Payload Data Transfers may be further segmented into single or multiple segments. In Segment Interleave Mode the segments composing a packet need not be transmitted consecutively. Segmentation of Payload Data Transfers shall only occur at segment boundaries. A segment is of a defined length SEGMENT_LEN. SEGMENT_LEN is a link provision-able parameter, which is an integer multiple (m) of 16 bytes (4 <= m <= 31). All implementations must support L = 64 (m = 4) i.e. 64 byte segmentation. Consecutive segments of the same packet may be concatenated into a single data burst with or without segmentation, but shall not exceed the length specified by MAX_BURST

Signaling Data Transfers may not be segmented and are limited to a maximum length of one segment (SEGMENT_LEN). The Control Word encoding leverages this fact to start Signal Data Transfers with an implied, rather than explicit, Start of Packet, see Annex A, Link State Transition Matrix.

The packet size of a Payload Data Transfers may be any integer multiple of bytes. This is accommodated by termination of the packet with End of Packet Padded (EPAD). The number of pad bytes is indicated in the last data byte of the packet and is dependent on both the size of the packet and the format (Quad or Octal Data Word) of the final data transfer unit. Efficient transmitters will only require Packet Size (in bytes) modulus 4 of pad bytes per packet. Worst case pad byte overhead would be Packet Size (in bytes) modulus 8.

6.1 Interleaving of Segments

The SPI-S supports two types segment interleaving, Packet Mode and Segment Mode. In *Packet Interleave Mode* (SEGMENT_INTERLEAVE = False) only one active or paused Payload Data Transfer is allowed. If a Payload Data Channel (designated by the Control Word Address field) is active or paused (Start-of-Packet without a corresponding End-of-Packet) then a second packet shall not be initiated until the currently active or paused PDT is completed with End Of Packet (or other packet termination). Therefore, the new PDT (for any channel) may only be initiated when the currently active (or paused) PDT is first completed or terminated. Conversely, in *Segment Mode* (SEGMENT_INTERLEAVE = True), supports multiple active/paused Payload Data Transfers (multiple active segmentation and re-assembly contexts) and Payload Data Transfers corresponding to different channels may be freely interleaved at any segment boundary. In Segment Interleave Mode each End of Segment (EOS) within the SPI-S stream provides an opportunity to pause an active PDT for a given channel and resume (or initiate) a new PDT for a different channel. Packet Interleave Mode shall be supported by SPI-S sources and sinks. If a SPI-S sink supports Segment Interleave Mode then SEG_INTERLEAVE may be enabled at the source (if supported).

Signaling Data Transfers may be interleaved with Payload Data Transfers at any segment boundary in both Packet Mode and Segment Mode. When Signaling Data Transfers are interleaved with Payload Data Transfers in Packet Mode, Payload Data transactions must resume with the currently paused Payload Data Transfer.

Both Payload and Signaling Data Transfers may be suspended, which is different from paused in two respects. First a Data Transfer maybe be suspended at any time without regard for segment boundaries. Second if a Data Transfer is suspended it must be resumed before any other Data Transfer can become active. The second rule insures that that SUSPEND is consistent with the segmentation rules even though it may occur at any time in an SPI-S data stream.

Figures 6-1a depicts the segmentation of a packet into two data bursts with a Suspend interleaved. In this example, Packet 1 is interrupted with a Suspend Control Block. Suspend insertion is not restricted to segment boundaries.

Figures 6-1b is an example of Segment Interleave Mode. The second Control Word pauses the PDT for packet 2 and initiates a new (SOP) PDT on a different channel (packet 3). In this case the Control Word may only be inserted at a segment boundary (EOS) of packet 2. This scenario would only be valid in Packet Interleave Mode if packet 3 was a Signaling Data Transfer and not a Payload Data Transfer.

Figure 6.1a: Single Suspend Control Word

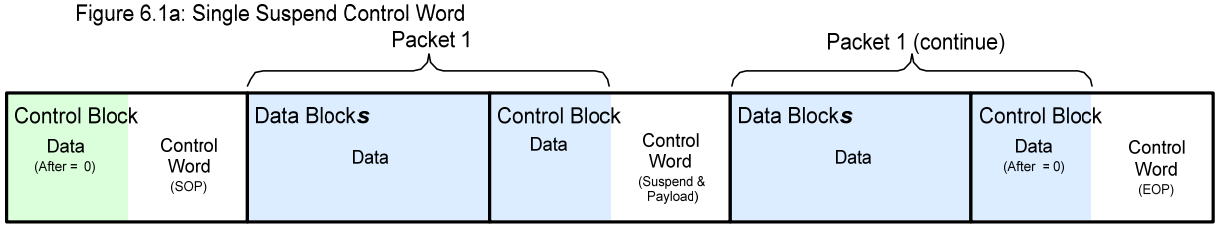
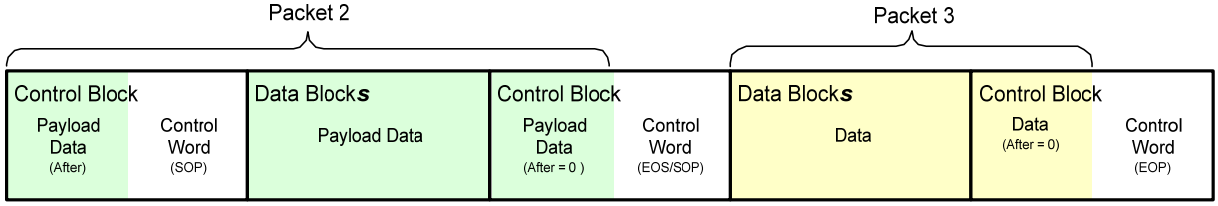


Figure 6.1b: Packet 2 Paused (Segment Interleave Mode Only)



Figures 6-1 a & b: Suspend and Paused Payload Data Transfers

6.2 Block Format

The data of the SPI-S Interface is logically organized blocks (64 bits of information). Blocks are either solely data “Data Block” or may contain both data and control information “Control Block.” The block type is indicated by the Tag field associated with each block.

For the 64B/66B frame format, the tag bit is double-coded in the sync bits (64B/66B synchronization preamble - 2 bits per 8 octets):

- Tag = 0 (Sync = 01): Data Block: Octal Data Word
- Tag = 1 (Sync = 10): Control Block: Quad Data Word + Control Word

For the CEI-P frame format, a dedicated Tag bit is transmitted:

- Tag = 0: Data Block: Octal Data Word
- Tag = 1: Control Block: Quad Data Word + Control Word

The remainder of this document will simply reference the Tag bit with the implication that in the case of 64/66B framing the Tag bit is encoded as described previously.

- If the TAG equals zero, then the 8 bytes is referred to as a *Data Block*. A Data Block is 8 bytes of segment data and is referred to as an Octal Data Word.
- If the TAG equals one, then the 8 bytes is referred to as a *Control Block*. A Control Block is 4 bytes of data, payload, signaling or status (Quad Data Word) followed by 4 bytes of control information (Control Word).

Figure 6-2 illustrates these two formats. Note that Big-endian (MSB first) is utilized within blocks and words for bit and byte order.

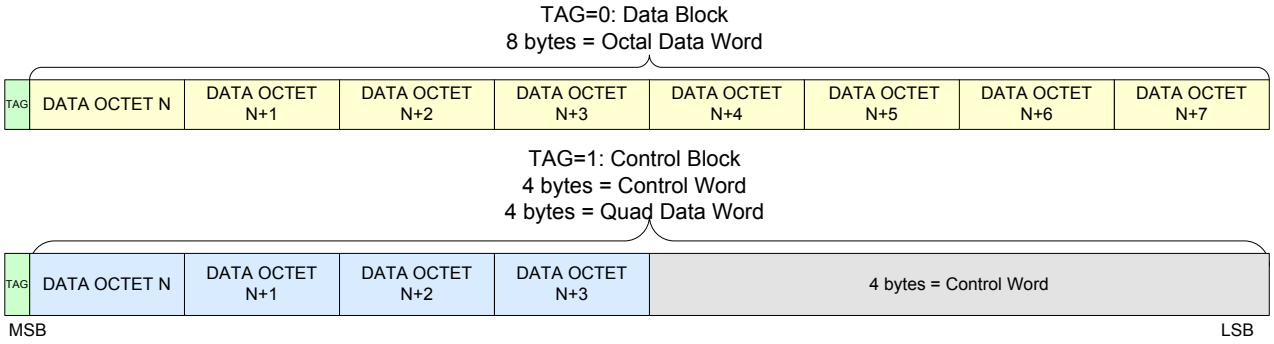


Figure 6-2: Block Formats

6.3 Control Block Format

As shown in Figure 6-3 the Control Block contains a Quad Data Word followed by a 4-byte Control Word. Both are described in detail below. The Control Word contains five control flags [4:0], an address field [14:0], and a CRC field [11:0]. The bit positions, of these fields, is shown in Figure 6-3 and are included for reference in Table 1. The rest of the document will refer to the individual sub-fields of the Control Word and not reference their bit position within the Control Word or the Control Block. The Quad Data Word contained in the control block may be associated with the current data context (logically precedes the Control Word) or may be the start of a new context (logically after the control word). the control block may be associated with the current data context (logically precedes the Control Word) or may be the start of a new context (logically after the control word).

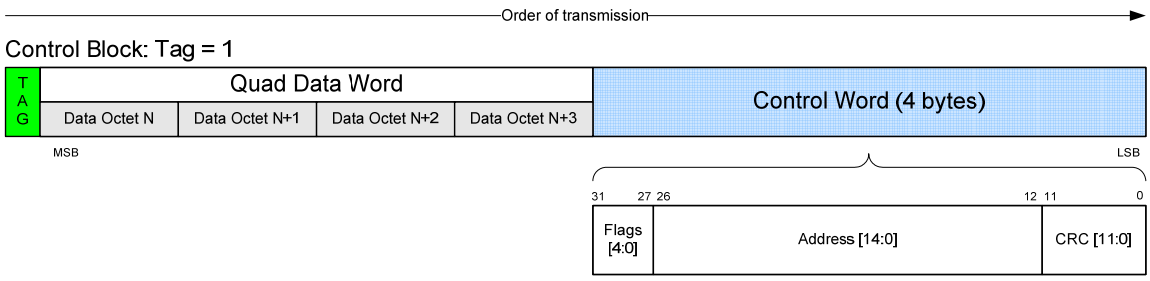


Figure 6-3: Control Block Format

Table 1: Description of Fields in the Control Word

Bit Position	Function	Description
31	AFTER	Control Word - Flag 4 Indicates if the Quad Data Word in the Control Block is part of a new stream of data (payload or signaling) - AFTER = 1: Quad Data Word is beginning of the new stream of data words, logically after (subsequent to) the Control Word. - AFTER = 0: Quad Data Word is a continuation of the current data stream, logically before the control word. (e.g. end of the currently active stream of data words) Flag 4 has an additional definition for EOT Suspend/Abort (see below)
30:29	End of Transfer Status (EOTS)	Control Word - Flag 3:2 - EOT = 00: End of Segment (EOS) - EOT = 10: End of Packet (EOP) with no pad byte(s) - EOT = 01: Suspend when AFTER = 0 and Abort when AFTER = 1 (SUS/ABT) - EOT = 11: End of Packet with pad byte(s) (EPAD)
28	PAYLOAD	Control Word – Flag 1 - PAYLOAD = 1: Indicates that associated data is a Payload Data Transfer - PAYLOAD = 0: Indicates that associated data is not a Payload Data Transfer
27	SOP-PDR	Control Word - Flag 0 If the associated data is a Payload Data Transfer (PAYLOAD = 1), then SOP-PDR is used to indicate a start of packet (SOP). - SOP = 1: Data segment is the start of a new packet - SOP = 0: Data segment is a continuation of a packet already being transferred across the interface If the associated data is not a Payload Data Transfer (PAYLOAD = 0), then SOP-PPDR is used to indicate Payload Data Ready (PDR). PDR throttles payload data transmission from the device's link partner: PDR = 0: Requests the link partner to complete the segment transfer in progress, and then cease Payload Data Transfers PDR = 1: Allows the link partner to resume a Payload Data Transfer.
[26:12]	ADDRESS[14:0]	Port Address 15-bit address field, where ADDRESS[14] is the MSB and ADDRESS[0] is the LSB
[11:0]	CRC	Cyclic Redundancy Check 12-bit CRC field, where CRC[11] is the MSB and CRC[0] is the LSB (and last bit transmitted in the Control Block).

Table 2: Type List for Control Flags

CONTROL FLAG[4:0]						bit4	bit[3:2]	bit1	bit0
	BIT 4	BIT 3	BIT 2	BIT 1	BIT 0	AFTER LOGICAL ORDER	EOTS	PAYLOAD SUBSEQUENT DATA IS:	SOP-PDR
0	0	0	0	0	0	Data precedes Control	End of Segment	Status Data	NOT Data Ready
1	0	0	0	0	1	Data precedes Control	End of Segment	Status Data	Data Ready
2	0	0	0	1	0	Data precedes Control	End of Segment	Packet Data	Packet Continue
3	0	0	0	1	1	Data precedes Control	End of Segment	Packet Data	Start of Packet
4	0	0	1	0	0	Data precedes Control	Suspend	Status Data	NOT Data Ready
5	0	0	1	0	1	Data precedes Control	Suspend	Status Data	Data Ready
6	0	0	1	1	0	Data precedes Control	Suspend	Packet Data	**NOT Data Ready**
7	0	0	1	1	1	Data precedes Control	Suspend	Packet Data	**Data Ready**
8	0	1	0	0	0	Data precedes Control	End of Packet (No Pad)	Status Data	NOT Data Ready
9	0	1	0	0	1	Data precedes Control	End of Packet (No Pad)	Status Data	Data Ready
10	0	1	0	1	0	Data precedes Control	End of Packet (No Pad)	Packet Data	Packet Continue
11	0	1	0	1	1	Data precedes Control	End of Packet (No Pad)	Packet Data	Start of Packet
12	0	1	1	0	0	Data precedes Control	End of Packet (With Pad)	Status Data	NOT Data Ready
13	0	1	1	0	1	Data precedes Control	End of Packet (With Pad)	Status Data	Data Ready
14	0	1	1	1	0	Data precedes Control	End of Packet (With Pad)	Packet Data	Packet Continue
15	0	1	1	1	1	Data precedes Control	End of Packet (With Pad)	Packet Data	Start of Packet

CONTROL FLAG[4:0]						bit4	bit[3:2]	bit1	bit0
	BIT 4	BIT 3	BIT 2	BIT 1	BIT 0	AFTER LOGICAL ORDER	EOTS	PAYLOAD SUBSEQUENT DATA IS:	SOP-PDR
16	1	0	0	0	0	Data is after Control	End of Segment	Status Data	NOT Data Ready
17	1	0	0	0	1	Data is after Control	End of Segment	Status Data	Data Ready
18	1	0	0	1	0	Data is after Control	End of Segment	Packet Data	Packet Continue
19	1	0	0	1	1	Data is after Control	End of Segment	Packet Data	Start of Packet
20	1	0	1	0	0	Data is after Control	Packet Abort	Status Data	NOT Data Ready
21	1	0	1	0	1	Data is after Control	Packet Abort or SYNC	Status Data	Data Ready
22	1	0	1	1	0	Data is after Control	Packet Abort	Packet Data	Packet Continue
23	1	0	1	1	1	Data is after Control	Packet Abort	Packet Data	Start of Packet
24	1	1	0	0	0	Data is after Control	End of Packet (No Pad)	Status Data	NOT Data Ready
25	1	1	0	0	1	Data is after Control	End of Packet (No Pad)	Status Data	Data Ready
26	1	1	0	1	0	Data is after Control	End of Packet (No Pad)	Packet Data	Packet Continue
27	1	1	0	1	1	Data is after Control	End of Packet (No Pad)	Packet Data	Start of Packet
28	1	1	1	0	0	Data is after Control	End of Packet (With Pad)	Status Data	NOT Data Ready
29	1	1	1	0	1	Data is after Control	End of Packet (With Pad)	Status Data	Data Ready
30	1	1	1	1	0	Data is after Control	End of Packet (With Pad)	Packet Data	Packet Continue
31	1	1	1	1	1	Data is after Control	End of Packet (With Pad)	Packet Data	Start of Packet

Summary of control encoding for more information on state transition, see annex A.

6.3.1 AFTER Flag

To allow the most efficient use of channel bandwidth and to minimize fragmentation losses, the most significant 4 bytes (Quad Data Word) of the Control Block may be utilized to transmit Data, as shown Figures 6-1a. Figures 6-1 depicts the segmentation of a packet into two data bursts with a Suspend interleaved. In this example, Packet 1 is interrupted with a Suspend Control Block.

Figures 6-1b is an example of Segment Interleave Mode. The second Control Word pauses the PDT for packet 2 and initiates a new (SOP) PDT on a different channel (packet 3). In this case, the Control Word may only be inserted at a segment boundary (EOS) of packet 2. This scenario would only be valid in Packet Interleave Mode if packet 3 was a Signaling Data Transfer and not a Payload Data Transfer.

A Control Block always contains four bytes of data (Quad Data Word). This data can either be associated with a new Data Transfer or the current Data Transfer. This is delineated by the AFTER flag in the Control Word. A Control Word with AFTER active (=1), is an indication that the data contained the Control Block is logically after the Control Word and hence is the start of a new Data Transfer (or Status Data) indicated by the Control Word contents. If the AFTER Flag is inactive (=0), this indicates the associated Quad Data word is simply a continuation of the currently data transfer, i.e. the data logically precedes the Control Word.

Figure 6-4: AFTER Flag = 0

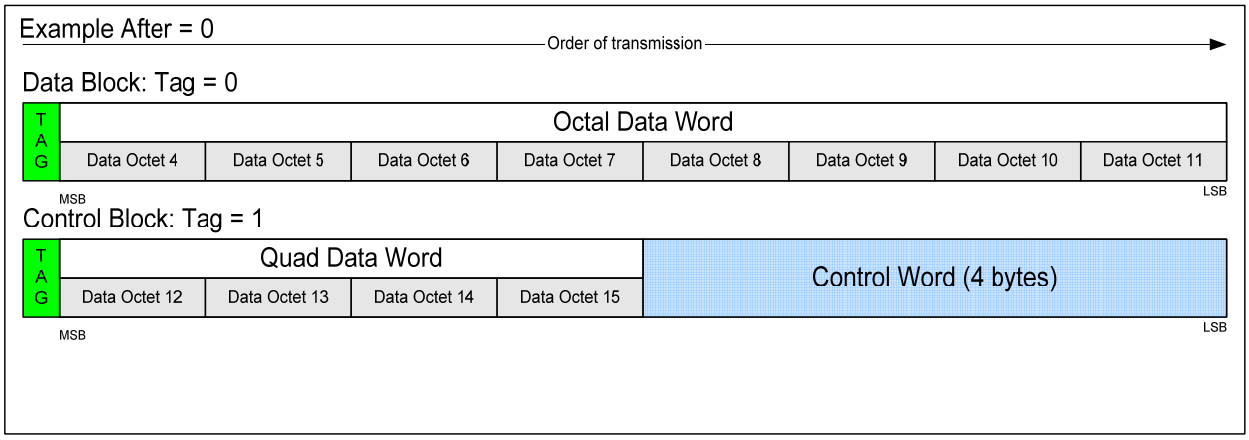


Figure 6-5: AFTER Flag = 1

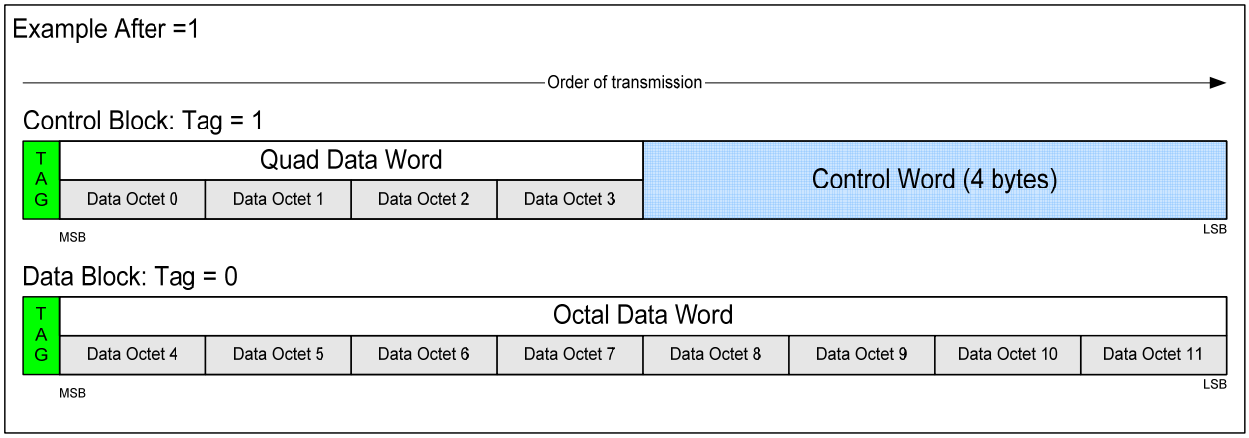
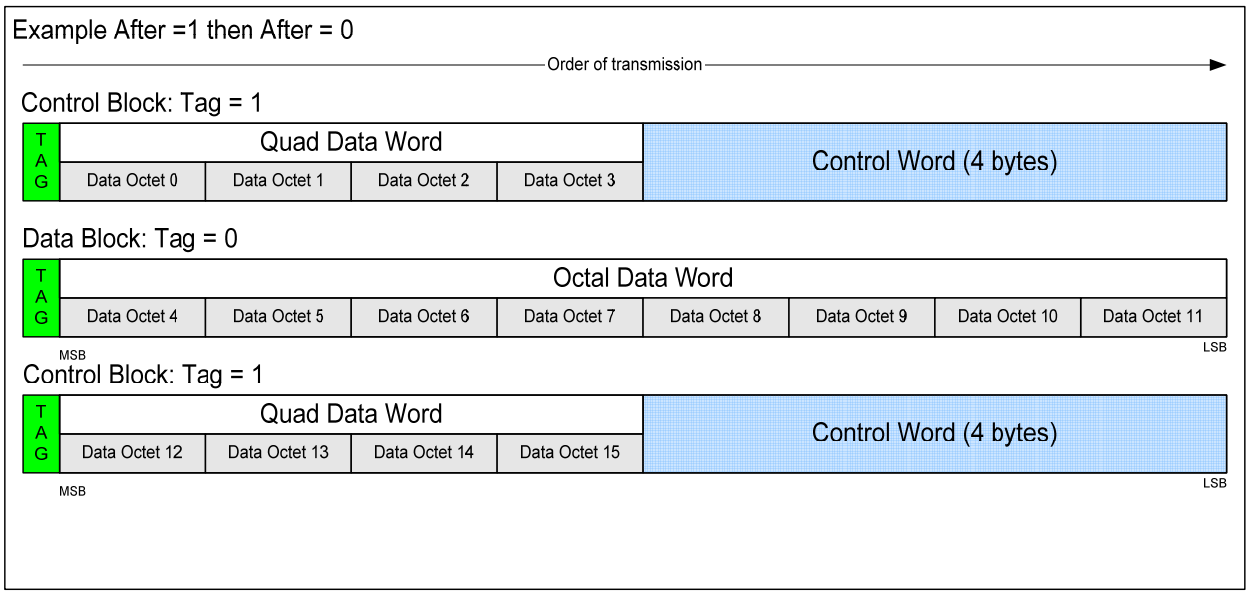


Figure 6-6: 16-Byte Packet (AFTER 1 then 0)



6.3.2 EOTS Flags

End of Transfer/Suspend (EOT) is a two bit field that indicates the termination state of the current burst. For Payload and Signal Data transfers all four possibilities are valid, End of Packet, EPAD and ABORT complete the current Data Transfer and cause that channel to enter the inactive state. Conversely End of Segment (EOS) places the active channel into a paused state (note that same Control word may pause and immediately re-activate the same channel). Information transfers that are not initiated with Start of Packet (e.g. other than PDT or SDT) shall be terminated with End of Segment encoding of the EOTS flags. The encoding for this flag is defined in Table 6.3.

When a Data Transfer is terminated with EPAD the last byte of this data transfer shall indicate the number of pad bytes present in the PDT. Note that the number of pad bytes reported includes the pad count byte. Pad bytes other than the pad count byte shall be should be zero filled and shall be ignored upon receipt.

Table 3: EOTS Definitions

Bits [3:2]	Description
0 0	EOS: End Of Segment (not end of packet)
1 0	EOP: End of packet without any pad bytes
0 1	Indicates either packet ABORT(AFTER=0) or SUSPEND(AFTER=1). This state of EOT also includes a one special SYNCHRONIZATION pattern
1 1	EPAD: End of Packet with pad bytes. Range of pad bytes depends on associated AFTER bit field (AFTER = 0, pad is 1, 2, or 3 bytes. AFTER = 1 pad is 1,2,3,4,5,6 or 7 bytes)

6.3.3 Payload Flag

The data transferred over the SPI-S interface can be either payload data or status data. Status data is further subdivided into different types based on the least significant nibble of the address field (Address[3:0]). Status data types are Void (no data content), Credit Pool Status, Signaling Data utilizing the NPF Messaging formats (see reference 4) or a Synchronization Block.

- PAYLOAD = 1: The subsequent data is a Payload Data Transfer.
- PAYLOAD = 0: The subsequent data may be a Signal Data Transfer, Pools Status or void (e.g. Idle)

6.3.4 SOP-PDR Flag

The SOP-PDR (Start of Packet/Payload Data Ready) flag is overloaded. If the PAYLOAD bit of the Control Word is active then the Control Word is initiating a new Payload Data Transfer (at a segment boundary) and the SOP-PDR bit field indicates Start of Packet. If the Control Word PAYLOAD bit is not active the associated data is not Payload Data and therefore SOP is either implied (for Signal Data Transfers) or not required. In these cases the SOP-PDR bit indicates Payload Data Ready for the reverse channel. There is one exception for the PAYLOAD active case, if the control word is a SUSPEND (EOTS = 01) then link state does not change and hence a Start of Packet would not be valid. This case also uses SOP-PDR for Payload Data Ready signaling even though PAYLOAD = 1 in Suspend control words. This exception provides a low latency means to toggle the reverse channel Payload Data Ready state during Payload Data of Signal Data Transfers. The SOP-PDR flag shall be ignored at the receiver for SYNC control blocks.

6.3.4.1 Start Of Packet (SOP)

When the SOP-PDR flag indicates Start Of Packet an active state initiates a new Payload Data Transfer in the channel indicated by the address field.

- SOP = 1: The subsequent Payload Data segement is the start of a packet.
- SOP = 0: The subsequent Payload Data segment is a continuation of a packet already being transferred across the interface.

In Packet Interleave Mode, a Start Of Packet (SOP) shall not be initiated unless there are no currently paused Payload Data Transfers. In Segment Interleave Mode, a Payload Data Transfer may be initiated (SOP) for any payload data context (address) that is currently inactive or is terminated (EOP/EPAD/ABORT) in the same Control Word.

6.3.4.2 Payload Data Ready (PDR)

Payload Data Ready signaling performs flow control for the reverse channel. Payload Data Ready signaling only applies to Payload Data Transfers, it does not effect other types of transfers. The default state of the link is always NOT Payload Data Ready until updated by the link partner. Implementations without a reverse link will need to activate PDR at least once during link initialization.

- PDR = 0: Requests the link partner to complete the segment transfer in progress. Once the segment in progress is completed the link partner shall not initiate new Data transfers as long as the PDR=0 condition persists

- PDR = 1: Allows the link partner to resume a Payload Data Transfers.

When the SOP-PDR flag indicates the Payload Ready state, it is indicating the reverse channels capability to accept additional Payload Data Transfers. Payload Data Ready provides a low latency throttling mechanism for the reverse channel, the link that receives a NOT Payload Data Ready flag may complete an current segment, but shall not initiate new Payload Data Transfers as long at the PDR=0 condition persists. The Payload Data Ready state of the link supersedes other flow control mechanism, but it does not revoke Credits granted by the Pool Status or Tokens generated by the Token-Bucket mechanisms. Once a link enters the NOT Payload Data Ready state, no Payload Data Transfers shall occur until the link state is changed by the reception of an active PDR flag. Signaling Data Transfers or Pool Status may still be transmitted by a link that is in the NOT Payload Data Ready state. When no reverse channel exists, the SOP-PDR flag should always be inactive when it is signaling the PDR state.

6.3.5 ADDRESS Field of Control Word

The function of the address field is determined by the state of the Flag bits in the same control word. If Payload Flag is active then the associated data is a Payload Data Transfer and the full address field [14:0] is used to indicate the Payload context. The number of Credit pools is 2^p, where p <= 15 (number of address bits). Scheduling can be implemented by using lower order pool bits (c) to create 2^c services classes c <= p. So the address hierarchy for Payload data is:

0 <= c (class bits) <= p (pool bits) <= 15

Un-provisioned Credit Pools should not be addressed, e.g. if p is 14 then ADDRESS[14] would always be zero.

For Status Data the Control Word Address Field is used to differentiate between the various types of Status Data and is summarized in Table 4.

		ADDRESS	
DATA TYPE	**CONTROL WORD FLAG(s)**	**ADDRESS [3:0]**	**ADDRESS[14:4]**
PAYLOAD	PAYLOAD	Payload context identifier	
IDLE	NOT PAYLOAD	ADDRESS [3:0] = 0000	ADDRESS[14:4] =“00000000000”
POOL STATUS	NOT PAYLOAD	ADDRESS [3:0] = 0010	ADDRESS[14:4] = Points to first credit pool for Pool Status
SYNC	FLAG[4:0] = “10101”	ADDRESS[0] = 1 ADDRESS[7:1] = Number of Bit Lanes in Link ADDRESS[14:8] = Current Bit Lane (starting with 1)	

Table 4: Status Data Address Decoding

6.4 Link States

The state of a SPI-S link determines the context (if any) of the subsequent data transfers on the link. The link state is only changed by Control Words. Note that some Control Words do not change the link state, but rather maintain the current state. The current Link State places restrictions on what of subsequent Control Words are valid, see annex A for the Link State Transition Matrix.

The logical point that the Link State changes is dependent on the state of the AFTER flag. When the AFTER flag is 1 the link state changes with the Control Block and the accompanying Quad DATA Word is associated with the new link state. If the AFTER flag is 0 then the Link State doesn't change until the beginning of the next Block (Data or Control) and the accompanying Quad Data Word is associated with the current Link State.

6.4.1 PAYLOAD DATA STATE (PDS)

Payload Data Transfers occur when the SPI-S link is in the Payload Data State. PDTs are initiated or resumed by Control Words with the PAYLOAD flag = 1. The link can support up to 32K unique channels using the 15 bit address field in the Control Word that initiates a PDT. Any number, including zero, of these channels may be provisioned. An active PDT may be paused by a SDT or another PDT addressed to a different channel (Segment Interleave Mode only) at any segment boundary. The PDT may also be suspended at any time if the link state is changed to Pool Status, IDLE or SYNC.

6.4.2 SIGNALING DATA STATE (SDS)

Signaling Data State is a separate link state reserved for transferring control packets encoded using NPF Messaging. Signaling Data Transfers are initiated or resumed by Control Words with the PAYLOAD flag = 0, ADDRESS[14:0] = x0003. There is only one SDT context. SDTs are not segmented, which implies that they shall not be terminated with End Of Segment (EOS) control words. SDT may be suspended and the Link State may transition from SDS to Pool Status, IDLE or SYNC. Non-segmentation of SDTs also requires that once suspended, the next Data Transfer on the link must resume the suspended SDT. Start of Packets (SOP) for SDT can be inferred from the current link state, and knowledge of the presence of a suspended SDT. This permits the SOP-PDR flag to indicate the PDR state of the reverse channel, as opposed to SOP.

6.4.3 POOL STATUS STATE (PSTAT)

The Pool Status state is used to transfer Pool Status for the reverse channel. Pool Status data transfers are paused not terminated (i.e. that pool status context is never inactive). Therefore Pool Status data transfers require no Start of Packet, End of Packet information Control Word signaling.

The Pool Status state may be entered at segment boundaries, or at non-segment boundaries using SUSPEND. The suspended Data Transfer must be resumed when the exiting the Pool Status state and returning to a Payload or Signaling Data state. The Pool Status and Idle states may be interleaved freely while a Data Transfer is suspended.

The associated ADDRESS field in the Control Word designates which Credit Pool the first two bits of Pool Status are associated with. See Section 7 Status Path Operation for more information. In the Pool Status state all data is associated with Credit Pool information (or is null if the indexing exceeds the provisioned SPI-S address space).

6.4.4 IDLE STATE

A link in the Idle State sends no information in the data fields. The Idle state is similar to the Pool Status State, in that it may be entered without regard for segment boundaries when transitioning from a Payload Data or Signaling Data state. However, if this creates a paused PDT or SDT that Data Transfer must be resumed when the exiting the Pool Status (or IDLE) state and returning to a Payload or Signaling Data state. As opposed to the PSTAT state, the associated data is null and need not be processed.

Encoding Flags [4:0] as 0100x with ADDRESS[14:0] = 0000h will place the link in the Idle State. All status data is null, any status value may be transmitted and shall be ignored by the receiver.

6.4.5 LINK SUSPENSION (SUS)

Suspend Control Words provide a low latency mechanism for communication of the Payload Data Ready (PDR) state or Credit Pool Status on the reverse channel. Suspend Control Words also allow the transmitter to suspend the current data burst at the next Block boundary when the Token Bucket of the Burst Admission Procedure becomes exhausted (see 6.5). Suspend Control Words also provide a reliable method of handling internal FIFO under-runs in the Transmitter, relieving the internal data path in the Transmitter of the requirement to support worst-case data bandwidth of the SPI-S interface.

Suspend control words may occur during Payload or Signaling Data Transfers without regard for segment (or packet) boundaries. To maintain consistency with segmentation rules and the implemented interleaving mode, context switching with a Suspend Control Word when a

suspended Data Transfer exists is not allowed. Other than altering the CRC codes in the subsequent control words insertion of one or multiple Suspend Control Blocks into the SPI-S stream shall not change the active Data Transfer. Idle Control Blocks or Credit Pool Status may be sent while a Payload or Signaling Data Transfer has been suspended. The first subsequent Data Transfer (Payload/Signal) and context (channel) must resume the suspend Data Transfer (e.g. resumed with a Control Word of the same PAYLOAD type and with the same ADDRESS of the Suspended Data Transfer). Data Transfers may enter a Suspend State, but the Suspend is not a link state. Suspends that do not alter the Link State are used to toggle the state of a Data Transfer between active and suspended.

Suspend Control Words that do not change the Link State allow low-latency low overhead PDR signaling. The PAYLOAD flag and ADDRESS field must match the current Data Transfer context. The Data Transfer is only suspended for the 4 bytes of the Control Word and immediately returns to the active Data State with the subsequent data. This provides a mechanism to signal PDR of the link partner with low latency and only 4 bytes of overhead.

Suspend Control Words may also be used to suspend an active Payload or Signaling Data Transfer for longer periods. The suspension of a Data Transfer differs from a paused transfer in that a suspended Data Transfer must be resumed when the link to transitions back to a Data Transfer state (it must be the same state and context that was suspended), regardless of the Interleave Mode. In this case, the Suspend Control Word transitions the link from a Payload or Signaling Data State to the Pool Status or Idle State. In the same fashion as PDR signaling, a suspend Control Word can be used to suspend an active Data Transfer and transfer Pool Status information, without waiting for a segment boundary.

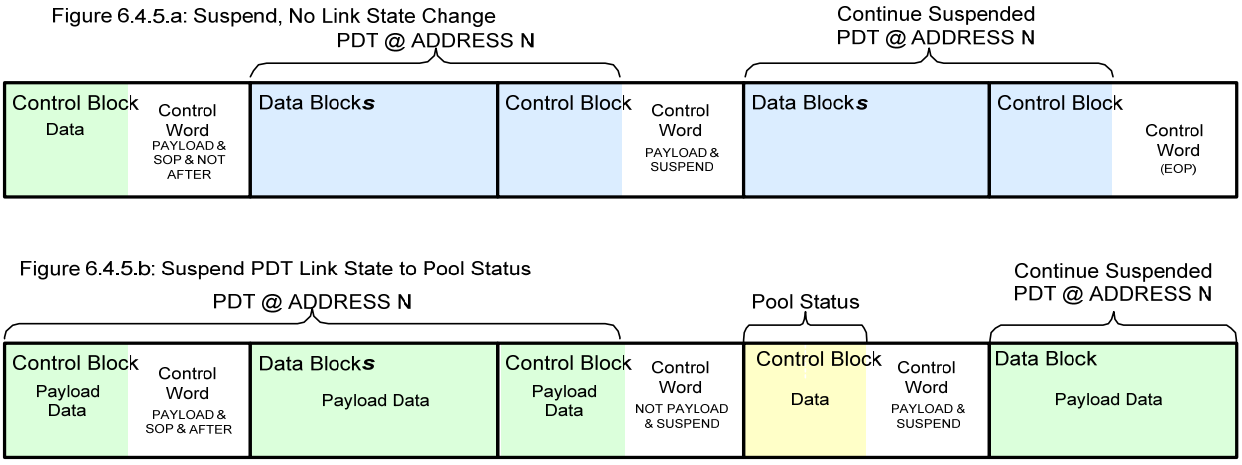


Figure 6-7: Example of Suspend Control (Link State no change/change)

6.4.6 LINK SYNCHONIZATION PATTERN

Transmission of the Link Synchronization Pattern (SYNC) is optional for single lane SPI-S links. Multi-lane SPI-S implementations shall transmit SYNC patterns. SYNC Control Blocks shall be inserted contiguously in all lanes of the SPI-S link beginning with lane 1. Table 6.4 defines a specific set of flags and address bits for synchronization of multi-lane SPI-S links that allows each lane to be uniquely identified. The content of the Quad data word associated with the synchronization control word (SYNC) is implementation specific. The minimum interval between SYNC is 256 Blocks per lane and the maximum is 4096 blocks per lane. SYNC may be inserted at any at any Block boundary beginning in lane 1

6.5 Burst Admission Procedure

The length of a Burst is a minimum of one segment (64 bytes or larger), with the exception of a Burst with an End of a Packet. Small packets, or a Burst that is terminated with an End of Packet that can be any length. This allows an SPI-S interface to efficiently transfer packets of any length.

SPI-S devices may operate internally at lower clock rates than the external SPI-S signals. In such devices, full bandwidth operation requires the internal bus carrying signals to or from the SPI-S interface to be wider than the external bus. This implies that burst boundaries can occur at more than one position on the internal bus. To facilitate further processing, these bursts may be separated and realigned. This introduces some inefficiency in the use of the internal bus, because it effectively rounds bursts up to the nearest multiple of the internal bus size. Consequently, it may be difficult to design a data path that can handle an arbitrarily long stream of short bursts.

Although the output side of an SPI-S data interface has the freedom to insert Idle data as needed to accommodate internal bus alignment constraints, the input side of the interface must accommodate all legal bursts generated by the sender. The Burst Admission Procedure places restrictions on the maximum rate of data transfer that may be generated by the sender. This mechanism allows the SPI-S source to be self throttling so that the SPI-S sink is not overwhelmed by continuous short burst transfers.

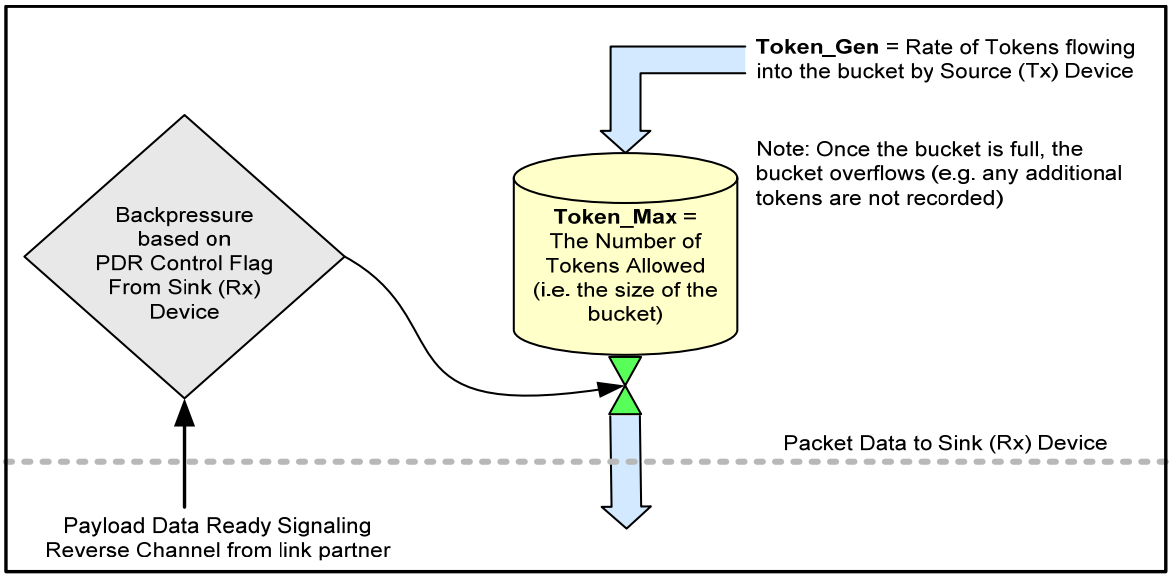
The Burst Admission Procedure regulates consecutive burst transfers across a data interface, regardless of Credit Pools, to avoid overloading the Sinks data path bandwidth. By comparison, the Pool Status regulates burst transfers on a per-Pool basis, regardless of sending order, to avoid overloading the input data FIFO for each Pool. In order for a burst transfer to begin, both the Burst Admission Procedure and the Pool Status must allow it

.

The Burst Admission Procedure uses a token bucket algorithm at the output side of the interface, as illustrated in Figure 6-8.

This algorithm uses per-block tokens that are generated at a fixed rate, stored in a finite container and consumed by both Payload and Signal Data Transfers. Each Token corresponds to an allocated amount of Data Transfer bandwidth. The parameter TOKEN_GRAN specifies the number of Quad Data Words granted by each Token. Tokens are consumed as Payload or Signal Data is sent by a SPI-S source. If TOKEN_GRAN was 1 a 64-byte PDT Segment would consume 16 (64/4) Tokens, but only 8 (64/(4*2)) Tokens if TOKEN_GRAN was 2. Pad bytes present in padded end of packet (EPAD) bursts are included in Token consumption calculations.

Figure 6-8: Token Bucket Algorithm



Tokens are stored in an up/down counter that can hold from 0 to TOKEN_MAX tokens, where TOKEN_MAX is a parameter. When a token is generated the counter is incremented; when tokens are consumed the counter is decremented. Increments shall not cause the counter to exceed TOKEN_MAX. Decrements shall not cause the counter to fall below zero. The value of TOKEN_MAX is a design parameter at the input side of the interface and a provisional parameter at the output side of the interface. The minimum value of TOKEN_MAX is the smaller of values A and B, where:

A = the number of provisioned lanes of a multi-lane link

B = the number of Blocks in a Segment = SEGMENT_LEN/8

Tokens are generated based on the Block rate of the SPI-S interface. Every TOKEN_GEN Block cycles, TOKEN_ADD tokens are added to the token bucket (where TOKEN_GEN and TOKEN_ADD are parameters). The ratio of TOKEN_ADD*TOKEN_GEN/TOKEN_GRAN sets an aggregate maximum rate for Data transmission by a source. The setting of these parameters is implementation specific and both over-speed and under-speed configurations are permitted. Configuration of these parameters on uni-directional links should always use under-speed to prevent FIFO overflow at the Sink. It is not recommended that a Source initiate a Payload or Signaling Data packet until a Segment worth of tokens are available.

6.6 Checksum

The CRC coverage is aligned with block boundaries, and covers all payload bits from consecutive blocks. A new CRC calculation begins with the first data word that follows a Control Block and ends with the CRC field of next Control Word. CRC is calculated using the transmitted bit order and hence is not affected by the state of the AFTER bit in the Control Block. It is pre-loaded with ones to detect variable number of leading zero Data Words.

The polynomial for the CRC is X^12 + X^7 + X^6 + X^4 + 1.

6.7 Striping

For interfaces using multiple serial bit lanes per link, striping across multiple lanes occurs at a Block granularity, starting with the lowest number lane (i.e. lane 1) and proceeding in increasing order of lane number.

6.8 Credit Pools

An SPI-S interface may have up to 32K channels (flow control contexts). Reverse channel Pools Status (if present) is transmitted in-band with data transfers. To accommodate a large number of channels, multiple channels may be collected into a Pool. Credits are granted and consumed on a per-Pool basis and all channels in a Pool share a common bank of credits. The pool number associated with a port address shall be uniquely identified by the lower order channel address bit values Address[POOL_LEN-1:0]. Applications with a small number of ports may choose to have one channel per pool, giving a simple one-to-one mapping between channels and credit pools.

Credits are granted by data path sink devices based on their ability to accept data. Each Pool Status update grants credits to the data source device, allowing it to transfer data for the corresponding Pool(s) across the interface. Each unit of credit grants the Source the ability to send a Segment (SEGMENT_LEN) to any one of the channels sharing the Pool.

Credits are consumed by data path source device for each Payload Data Transfer segment. A credit is consumed for each Segment of Payload Data that is transferred. Transfer of a Payload Data segment (or packet) smaller that SEGMENT_LEN bytes still consumes one whole credit. Therefore, the final PDT for a segment consumes a whole credit even if it is shorter than SEGMENT_LEN*8 bytes in length.

Credits are granted to Pools and are consumed for each PDT Segment transferred from any channel within the Pool. The source of a SPI-S interface shall only schedule transfers for ports whose associated Pool have sufficient credits to support the entire transfer.

6.9 Pool Status Format

The Credit Pools are initialized to Satisfied by the source until updated. The link enters the Pool Status State upon reception of a Pool Status Control Word (Payload = 0 and, with Address[3:0] = 0010). Once the link enters the Credit Pool State, subsequent Data transfers are always Credit Pool Status until a new Control Word changes the link state. Quad Data Words (32-bits) may indicate the 2-bit status for up to 16 Credit Pools, subsequent Octal Data Words may convey pool status of up 32 to subsequent pools. The Pool Status information within the Data Word shall increment sequentially up from the initial Pool indicated by Address[14:4] (as illustrated in Table 6.4). This address points to the Credit Pool associated with the first two bits of the data (Pool Status). Each subsequent two bits of data are associated with the next (incremental) Credit Pool. The addressed Credit Pool always increments from the initial Pool and does not wrap around if the indexed Pool extends beyond the provisioned space. Satisfied status shall be transferred for any un-provisioned Pool that is encompassed by the positional Pool addressing. The values for the 2-bit status are the same as defined for the OIF SPI-4.2 and SPI-5 specifications and are defined in Table 5.

Table 5: Pool Status Report Format

STAT[1]	STAT[0]	Description
1	1	RESERVED
1	0	SATISFIED The SATISFIED status indicates that the FIFO of the corresponding sink Pool is almost full. When the SATISFIED status is received, no further credits are granted. Credits granted previously by HUNGRY or STARVING status reports remain available.
0	1	HUNGRY The HUNGRY status indicates that the FIFO of the corresponding sink Pool is partially empty. When a HUNGRY status is received, the amount of credits at the source Pool is increased to MAXBURST2 blocks if the current value is less than MAXBURST2. If the amount of credit remaining is currently greater than MAXBURST2, due to a previous STARVING status report, the credit count is left unchanged.
0	0	STARVING The STARVING status indicates that the FIFO of the corresponding sink Pool is almost empty. When a STARVING status is received, the amount of credits at the source Pool is set to MAXBURST1 blocks.

The minimum Pools Status granularity is a quad data word (32-bits) indicating the 2-bit status for 16 credit pools. The pointer to the first Credit Pool to be updated is provided in the Control Word address field (ADDRESS[14:4]) as illustrated in Figure 6-9. Bits [3:0] of the address field shall be set to 0010 to indicate that the Signaling context is Credit Pool Status, but are assumed (masked to) 0000 to generate a 15 bit Credit pool pointer (ADDRESS[14:4] & 0000). When AFTER=1, the associated Quad Data word contains the Pool Status for 16 sequential ports starting with the Credit Pool addressed by the address field. Octal Data Words of Credit Pool Status contain Pool Status for 32 sequential ports. The status does not wrap back to Credit Pool zero if the sequential access would exceed the number of provisioned Credit Pools. If sequential addressing from the initial addressed Credit Pool exceeds the provisioned Pools, the status for these Pools shall always be assigned a satisfied (10) state. The positional association always increments by one, for each 2 bits of Status; it does not skip any un-provision Credit Pool.

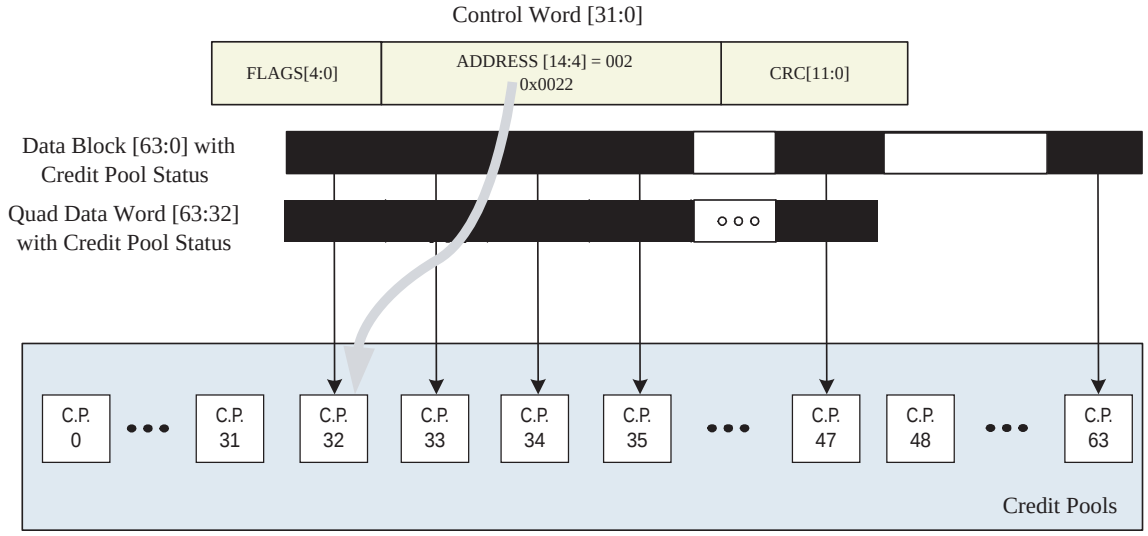


Figure 6-9: Pool Status Addressing

Note: When the positional association of Credit status accesses an un-provisioned Credit Pool, the Status should be set to satisfied by the SPI-S source. SPI-S sinks shall assume a satisfied state for any Credit Pool that has not received Status.

Because the Transmit and Receive interfaces operate independently, their Credit Pools may be provisioned differently.

7 SPI-S Framing Modes

7.1 Mode 1: 64B/66B

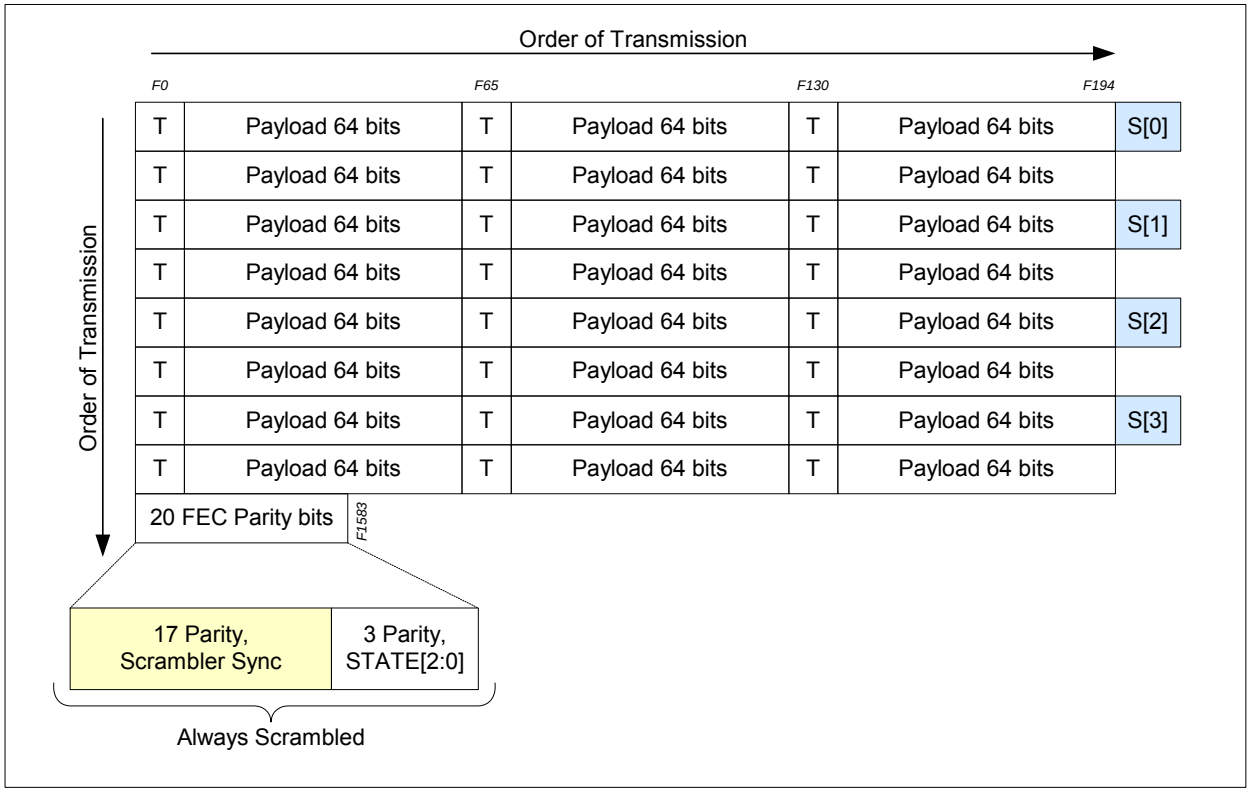
The SPI-S requires support for operation utilizing IEEE 64B/66B coding and scrambling, as defined in 802.3ae-2002 (refer to clause 49.1.4.1 - PCS introduction and to clause 49.2 - Physical Coding Sub layer).

The 802.3ae (64B/66B) framing is either a 01 or a 10 followed by 64-bit Payload. For SPI-S applications, the 01 transition denotes a TAG bit value of one, the 10 transition denotes a TAG bit value of zero.

7.2 Mode 2: CEI-P

The SPI-S allows support for an optional mode of operation utilizing OIF CEI-P framing, scrambling, and FEC. This mode is recommended for use with interfaces requiring the use of Forward Error Correction (FEC). The SPI-S data flow of a TAG bit and 64 data bits of associated information is mapped into a CEI-P frame using consecutive T and 64 payload bit fields as shown in Figure 7-1.

Figure 7-1: CEI-P Frame Format



Each CEI-P frame carries 24 Tag bits and associated 64 bits of information. The S[0:3] bits may be used for low latency flow control applications.

8 Interface Parameters

The SPI-S interface is configured by five types of parameters. The first configures the data path for the Logical Link; the second the addressing of Ports, Pools and Classes; the third the Payload Data Transfer segmentation; the fourth the credits granted by Pool Status Reports, and the fifth the Burst Admission Procedure.

The complete list of interface parameters is listed below in Table 6. The column labeled “P” denotes whether the associated parameter is configurable within devices that are connected to the interface. The column labeled “S” defines the scope (per link/per pool) of the parameter. A device is deemed compliant if it supports Source interface configuration on parameters labeled “Y/O.” The column labeled “CP” describes whether the parameter is configurable on a per-Pool basis or only configurable on a per-logical link basis. A device is deemed compliant if it supports either per-Pool or per-Link configuration on parameters labeled “P/L.”

In the event of discrepancy between Table 6 and other portions of this document the contents of Table 6 takes precedence.

Table 6: Interface Parameters

Parameter	Definition	P	S	Unit
FRAMING_MODE	Indicates whether the interface is operating in 64B/66B mode or CEI-P mode. This parameter shall be provisionable for interfaces that support CEI-P mode.	O	L	
MAX_LANES	Maximum supported number of Physical Lanes for the Logical Link	N	L	Lanes
PHY_LANES	Number of provisioned Physical Lanes for the Logical Link	O	L	Lanes
SEGMENT_LEN	Nominal length of a Data Segment. All Segments have identical length except the last segment of a Packet which may be shorter and have an odd or even length. $8 \leq$ SEGMENT_LEN ≤ 62 AND SEGMENT_LEN modulo 2 = 0	O	L	An even number of : Octal Data Words
MAX_BURST	Maximum length of a Payload Data burst (between Control Words)	Y/O	L	Segments
SEG_INTERLEAVE	Indicates that segments of different packets may be interleaved on the Link. This parameter shall be provisionable for interfaces that support segment interleaved operation.	O	L	Boolean
TOP_PADDR	Highest Port Address supported. The range of supported Port Addresses is 0 to TOP_PADDR.	N	L	Ports
TOP_POOL	Highest Credit Pool supported. The range of supported Credit Pools is 0 to TOP_POOL.	N	L	Ports

Parameter	Definition	P	S	Unit
POOL_LEN	Number of bits in the port address used to uniquely decode for the associated Credit Pool. $0 \leq POOL\_LEN \leq 15$	O	L	Bits
TOP_CLASS	Highest Class supported. The range of supported Classes is 0 to TOP_CLASS.	N	L	Ports
CLASS_LEN	Number of bits in the port address used to uniquely decode for the associated Credit Pool. $0 \leq CLASS\_LEN \leq POOL\_LEN$	O	L	Bits
PS_PRESENT	Indicates presence of associated Pool Status Report path (Sink to Source) for this Payload Data path (Source to Sink)	Y	L	Boolean
MAX_STARVING	Number of credits granted when Pool Status Report indicates Starving	Y	P/L	Segments
MAX_HUNGRY	Number of credits granted when Pool Status Report indicates Hungry. $MAX\_HUNGRY \leq MAX\_STARVING$	Y	P/L	Segments
TOKEN_GRAN	The quantum of tokens that the Source is required to consume atomically.			Quad Data Words
TOKEN_MAX	Maximum number of tokens for 32 byte blocks that the Source is allowed to accumulate. $TOKEN\_MAX \geq minimum($ $(2 * PHY\_LANES),$ $(2 * SEGMENT\_LEN + 1))$	Y/O	L	Quad Data Words
TOKEN_GEN	Number of Block Cycles that must expire in order to accumulate additional tokens.	Y/O	L	Block Cycles
TOKEN_ADD	Number of tokens that the Source may accumulate per Block Cycle	Y/O	L	Quad Data Words

Notes:
P : Provisionable Parameter:
Yes [Y], No (Fixed within device) [N], Optional [O],
Yes (Source) Optional (Sink) [Y/O]
S : Scope of Parameter instance:
per-Pool [P], per-Link [L], either [P/L]

9 ANNEX A: Link State Transition Matrix (NORMATIVE)

9.1 State Transitions from PDS

Table 7: State Transitions from Payload Data State

Previous Link State	CONTROL WORD			New Link State							Reverse Link PDR SIGNALING
				PAYLOAD = 1			PAYLOAD = 0				
	FLAG[4:0] (hex)	EOTS	SOP/PDR	No Address Change	Address Change			Pool Status	Sync	Idle	
					Packet Interleave	Segment Interleave	ADDR[3:0] = 0100	ADDR[3:0] = 0010	ADDR[3:0] = xxx1	ADDR[3:0] = 0000	
PDS	00	EOS	NOT PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
PDS	01	EOS	PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	PDR
PDS	02	EOS	NOT SOP	Continue Current PDT	Not Allowed-(3)	Resume Paused PDT					NO CHANGE
PDS	03	EOS	SOP	Not Allowed-(1)	Not Allowed-(3)	Begin New PDT					NO CHANGE
PDS	04	SUS	NOT PDR				Not Allowed-(4)	Pool Status (Sus: PDT)	Not Allowed-(7)	Suspend PDT	NOT PDR
PDS	05	SUS	PDR				Not Allowed-(4)	Pool Status (Sus: PDT)	Not Allowed-(7)	Suspend PDT	PDR
PDS	06	SUS	NOT PDR	Continue Current PDT	Not Allowed-(4)	Not Allowed-(4)					NOT PDR
PDS	07	SUS	PDR	Continue Current PDT	Not Allowed-(4)	Not Allowed-(4)					PDR
PDS	08	EOP	NOT PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
PDS	09	EOP	PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	PDR
PDS	0A	EOP	NOT SOP	Not Allowed-(2)	Not Allowed-(2)	Resume Paused PDT					NO CHANGE
PDS	0B	EOP	SOP	Begin New PDT	Begin New PDT	Begin New PDT					NO CHANGE
PDS	0C	EPAD	NOT PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
PDS	0D	EPAD	PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	PDR
PDS	0E	EPAD	NOT SOP	Not Allowed-(2)	Not Allowed-(2)	Resume Paused PDT					NO CHANGE
PDS	0F	EPAD	SOP	Begin New PDT	Begin New PDT	Begin New PDT					NO CHANGE
PDS	10	EOS	NOT PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
PDS	11	EOS	PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	PDR
PDS	12	EOS	NOT SOP	Continue Current PDT	Not Allowed-(3)	Resume Paused PDT					NO CHANGE
PDS	13	EOS	SOP	Not Allowed-(1)	Not Allowed-(3)	Begin New PDT					NO CHANGE
PDS	14	ABORT	NOT PDR				New SDT-(5,6)	Pool Status-(6)	Not Allowed-(7)	IDLE-(6)	NOT PDR
PDS	15	ABORT/SYNC	PDR/NA				New SDT-(5,6)	Pool Status-(6)	Sync (NO State Change)	IDLE-(6)	PDR (except SYNC)
PDS	16	ABORT	NOT SOP	Not Allowed-(2)	Not Allowed-(2)	Resume Paused PDT-(6)					NO CHANGE
PDS	17	ABORT	SOP	Begin New PDT-(6)	Begin New PDT-(6)	Begin New PDT-(6)					NO CHANGE
PDS	18	EOP	NOT PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
PDS	19	EOP	PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	PDR
PDS	1A	EOP	NOT SOP	Not Allowed-(2)	Not Allowed-(2)	Resume Paused PDT					NO CHANGE
PDS	1B	EOP	SOP	Begin New PDT	Begin New PDT	Begin New PDT					NO CHANGE
PDS	1C	EPAD	NOT PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
PDS	1D	EPAD	PDR				New SDT-(5)	Pool Status	Not Allowed-(7)	IDLE	PDR
PDS	1E	EPAD	NOT SOP	Not Allowed-(2)	Not Allowed-(2)	Resume Paused PDT					NO CHANGE
PDS	1F	EPAD	SOP	Begin New PDT	Begin New PDT	Begin New PDT					NO CHANGE

NOTES:

1	SOP for a port address that is paused
2	Previoius Port is now inactive can't enter active state without SOP
3	Packet Interleave mode can only have one active or paused port
4	Changing port address not allowed with Suspends
5	Implied SOP for Signal Data Transfer (enter SDT state)
6	Current PDT is terminated by ABORT
7	Payload = 0 @ this address only valid with SYNC (FLAG = 15h)

9.2 State Transitions from SDS

Table 8: State Transitions from Signaling Data State

Previous	CONTROL WORD				New Link State							
Link					PAYLOAD = 1			PAYLOAD = 0				Reverse Link
State	FLAG[4:0]	SOP-PDR	EOTS	SOP-PDR	No Address Change	Address Change (since previous Payload=1)		Signaling Data	Pool Status	Sync	Idle	PDR
	hex					Packet Interleave	Segment Interleave	ADDR[3:0] = 0100	ADDR[3:0] = 0010	ADDR[3:0] = xxx1	ADDR[3:0] = 0000	Signaling
SDS	00	0	EOS	NOT PDR				Not Allowed-(1)	Not Allowed-(1)	Not Allowed-(1,7)	Not Allowed-(1)	NOT PDR
SDS	01	1	EOS	PDR				Not Allowed-(1)	Not Allowed-(1)	Not Allowed-(1,7)	Not Allowed-(1)	PDR
SDS	02	0	EOS	NOT SOP	Not Allowed-(1)	Not Allowed-(1)	Not Allowed-(1)					NO CHANGE
SDS	03	1	EOS	SOP	Not Allowed-(1)	Not Allowed-(1)	Not Allowed-(1)					NO CHANGE
SDS	04	0	SUS	NOT PDR				Continue Current SDT	Pool Status (Sus:SDT)	Not Allowed-(7)	Suspend SDT	NOT PDR
SDS	05	1	SUS	PDR				Continue Current SDT	Pool Status (Sus:SDT)	Not Allowed-(7)	Suspend SDT	PDR
SDS	06	0	SUS	NOT PDR	Not Allowed-(2)	Not Allowed-(2)	Not Allowed-(2)					NOT PDR
SDS	07	1	SUS	PDR	Not Allowed-(2)	Not Allowed-(2)	Not Allowed-(2)					PDR
SDS	08	0	EOP	NOT PDR				New SDT (5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
SDS	09	1	EOP	PDR				New SDT (5)	Pool Status	Not Allowed-(7)	IDLE	PDR
SDS	0A	0	EOP	NOT SOP	Resume Paused PDT	Resume Paused PDT-(3)	Resume Paused PDT					NO CHANGE
SDS	0B	1	EOP	SOP	Begin New PDT	Begin New PDT-(4)	Begin New PDT					NO CHANGE
SDS	0C	0	EPAD	NOT PDR				New SDT (5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
SDS	0D	1	EPAD	PDR				New SDT (5)	Pool Status	Not Allowed-(7)	IDLE	PDR
SDS	0E	0	EPAD	NOT SOP	Resume Paused PDT	Resume Paused PDT-(3)	Resume Paused PDT					NO CHANGE
SDS	0F	1	EPAD	SOP	Begin New PDT	Begin New PDT-(4)	Begin New PDT					NO CHANGE
SDS	10	0	EOS	NOT PDR				Not Allowed	Not Allowed	Not Allowed-(7)	Not Allowed	NOT PDR
SDS	11	1	EOS	PDR				Not Allowed	Not Allowed	Not Allowed-(7)	Not Allowed	PDR
SDS	12	0	EOS	NOT SOP	Not Allowed-(1)	Not Allowed-(1)	Not Allowed-(1)					NO CHANGE
SDS	13	1	EOS	SOP	Not Allowed-(1)	Not Allowed-(1)	Not Allowed-(1)					NO CHANGE
SDS	14	0	ABORT	NOT PDR				New SDT-(5,6)	Pool Status-(6)	Not Allowed-(7)	IDLE-(6)	NOT PDR
SDS	15	1	ABORT/SYNC	PDR/NA				New SDT-(5,6)	Pool Status-(6)	SYNC (NO State Change)	IDLE-(6)	PDR
SDS	16	0	ABORT	NOT SOP	Resume Paused PDT-(6)	Resume Paused PDT-(3,6)	Resume Paused PDT-(6)					NO CHANGE
SDS	17	1	ABORT	SOP	Begin New PDT-(6)	Begin New PDT-(4,6)	Begin New PDT-(6)					NO CHANGE
SDS	18	0	EOP	NOT PDR				New SDT (5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
SDS	19	1	EOP	PDR				New SDT (5)	Pool Status	Not Allowed-(7)	IDLE	PDR
SDS	1A	0	EOP	NOT SOP	Resume Paused PDT	Resume Paused PDT-(3)	Resume Paused PDT					NO CHANGE
SDS	1B	1	EOP	SOP	Begin New PDT	Begin New PDT-(4)	Begin New PDT					NO CHANGE
SDS	1C	0	EPAD	NOT PDR				New SDT (5)	Pool Status	Not Allowed-(7)	IDLE	NOT PDR
SDS	1D	1	EPAD	PDR				New SDT (5)	Pool Status	Not Allowed-(7)	IDLE	PDR
SDS	1E	0	EPAD	NOT SOP	Resume Paused PDT	Resume Paused PDT-(3)	Resume Paused PDT					NO CHANGE
SDS	1F	1	EPAD	SOP	Begin New PDT	Begin New PDT-(4)	Begin New PDT					NO CHANGE

NOTES:

1	Signal Data Transfers may not be segmented
2	Would be a change of transfer type during SUSPEND
3	Payload address must be the same as (single) paused PDT
4	Only ALLOWED if there is no paused PDT for this port
5	Implied SOP
6	Current SDT is terminated by ABORT
7	Payload = 0 @ this address only valid with SYNC (FLAG = 15h)

9.3 State Transitions from PSTAT or IDLE

Table 9: State Transitions from PSTAT or IDLE

Previous	CONTROL WORD			New Link State							Reverse Link
Link				PAYLOAD = 1			PAYLOAD = 0				
State	FLAG[4:0]	EOTS	SOP-PDR	No Address Change	Address Change (since previous Payload=1)		Signaling Data	Pool Status	Sync	Idle	FLOW CONTROL
	(hex)				Packet Interleave	Segment Interleave	ADDR[3:0] = 0100	ADDR[3:0] = 0010	ADDR[3:0] = xxx1	ADDR[3:0] = 0000	
Idle / Pstat	00	EOS	NOT PDR				New SDT-(4)	Pool Status	Not Allowed-(6)	IDLE (not Suspend)	NOT PDR
Idle / Pstat	01	EOS	PDR				New SDT-(4)	Pool Status	Not Allowed-(6)	IDLE (not Suspend)	PDR
Idle / Pstat	02	EOS	NOT SOP	Resume Paused PDT	Resume Paused PDT-(1)	Resume Paused PDT					NO CHANGE
Idle / Pstat	03	EOS	SOP	Begin New PDT	Begin New PDT-(2)	Begin New PDT					NO CHANGE
Idle / Pstat	04	SUS	NOT PDR				Continue Suspended SDT	Not Allowed-(5)	Not Allowed-(6)	Extend PDT or SDT Suspend	NOT PDR
Idle / Pstat	05	SUS	PDR				Continue Suspended SDT	Not Allowed-(5)	Not Allowed-(6)	Extend PDT or SDT Suspend	PDR
Idle / Pstat	06	SUS	NOT PDR	Continue Suspended PDT-(1)	Not Allowed-(1)	Not Allowed-(1)					NOT PDR
Idle / Pstat	07	SUS	PDR	Continue Suspended PDT-(1)	Not Allowed-(1)	Not Allowed-(1)					PDR
Idle / Pstat	08	EOP	NOT PDR				Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3,6)	Not Allowed-(3)	NOT PDR
Idle / Pstat	09	EOP	PDR				Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3,6)	Not Allowed-(3)	PDR
Idle / Pstat	0A	EOP	NOT SOP	Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3)					NO CHANGE
Idle / Pstat	0B	EOP	SOP	Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3)					NO CHANGE
Idle / Pstat	0C	EPAD	NOT PDR				Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3,6)	Not Allowed-(3)	NOT PDR
Idle / Pstat	0D	EPAD	PDR				Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3,6)	Not Allowed-(3)	PDR
Idle / Pstat	0E	EPAD	NOT SOP	Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3)					NO CHANGE
Idle / Pstat	0F	EPAD	SOP	Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3)					NO CHANGE
Idle / Pstat	10	EOS	NOT PDR				New SDT-(4)	Pool Status	Not Allowed-(6)	IDLE (not Suspend)	NOT PDR
Idle / Pstat	11	EOS	PDR				New SDT-(4)	Pool Status	Not Allowed-(6)	IDLE (not Suspend)	PDR
Idle / Pstat	12	EOS	NOT SOP	Resume Paused PDT	Resume Paused PDT-(1)	Resume Paused PDT					NO CHANGE
Idle / Pstat	13	EOS	SOP	Begin New PDT	Begin New PDT-(2)	Begin New PDT					NO CHANGE
Idle / Pstat	14	ABORT	NOT PDR				Continue Suspended SDT	Not Allowed-(5)	Not Allowed-(7)	Extend PDT or SDT Suspend	NOT PDR
Idle / Pstat	15	ABORT/SYNC	PDR/NA				Continue Suspended SDT	Not Allowed-(5)	Sync (NO State Change)	Extend PDT or SDT Suspend	PDR
Idle / Pstat	16	ABORT	NOT SOP	Continue Suspended PDT-(1)	Not Allowed-(1)	Not Allowed-(1)					NO CHANGE
Idle / Pstat	17	ABORT	SOP	Not Allowed-(6)	Not Allowed-(1)	Not Allowed-(1)					NO CHANGE
Idle / Pstat	18	EOP	NOT PDR				Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3,6)	Not Allowed-(3)	NOT PDR
Idle / Pstat	19	EOP	PDR				Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3,6)	Not Allowed-(3)	PDR
Idle / Pstat	1A	EOP	NOT SOP	Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3)					NO CHANGE
Idle / Pstat	1B	EOP	SOP	Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3)					NO CHANGE
Idle / Pstat	1C	EPAD	NOT PDR				Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3,6)	Not Allowed-(3)	NOT PDR
Idle / Pstat	1D	EPAD	PDR				Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3,6)	Not Allowed-(3)	PDR
Idle / Pstat	1E	EPAD	NOT SOP	Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3)					NO CHANGE
Idle / Pstat	1F	EPAD	SOP	Not Allowed-(3)	Not Allowed-(3)	Not Allowed-(3)					NO CHANGE

NOTES:

1	Payload address must be the same as paused (or Suspended) PDT
2	Only ALLOWED if there is no paused PDT for this port
3	No currently active PDT or SDT therefore EOP or EPAD are not valid
4	Implied SOP
5	Pool Status simply terminates, never Suspended or Abort
6	SOP for a port address that is paused

10 Appendix A: Glossary

The following acronyms and abbreviations are used in this specification

CEI	Common Electrical I/O
CEI-P	Common Electrical I/O Protocol
EOP	End of Packet
EOS	End of Segment
EPAD	End of Packet with PAD bytes
IEEE	Institute of Electrical and Electronics Engineers
I/O	Input / Output (signal).
LSB	Least Significant Bit
MSB	Most Significant Bit
NPE	Network Processing Element
NPF	Network Processing Forum
NPSI	Network Processing Forum Streaming Interface
OIF	Optical Internetworking Forum
PDR	Payload Data Ready
PDS	Payload Data State
PDT	Payload Data Transfer
PHY	Physical Layer Device
PSTAT	Pool Status State
SDS	Signaling Data State
SDT	Signaling Data Transfer
SONET	Synchronous Optical Network
SOP	Start of Packet
SPI-S	Scalable System Packet Interface

11 Appendix B: List of Companies belonging to OIF when Document was Approved

ADVA Optical Networking
Agere Systems
Agilent Technologies
Alcatel
Altera
AMCC
Ample Communications
Analog Devices
Anritsu
Apogee Photonics, Inc.
AT&T
Atos Origin Integration
Azna
Bay Microsystems
Bookham
Booz-Allen & Hamilton
Broadcom
China Telecom
Ciena Corporation
Cisco Systems
CoreOptics
Cortina Systems
Data Connection
Department of Defense
Deutsche Telekom
Ericsson
Essex Corporation
Finisar Corporation
Flextronics
Force 10 Networks
Foxconn
Freescale Semiconductor
Fujitsu
Furukawa Electric Japan
Hi/fn
Huawei Technologies
IBM Corporation
IDT
Infinera
Intel
IP Infusion
JDSU
KDDI R&D Laboratories
KT Corporation
LSI Logic
Lucent
Mercury Computer Systems, Inc
MergeOptics GmbH
Mintera
MITRE Corporation
Mitsubishi Electric Corporation
Molex
Motorola
NEC
Nortel Networks
NTT Corporation
Opnext
OpVista Inc
Orange World
Paxera Corp
Phyworks Limited
PMC Sierra
Radisys Corp
Redfern Integrated Optics, Inc.
RSoft Design Group, Inc.
Sandia National Laboratories
Santur
Scintera Networks
Siemens
Silicon Logic Engineering
StrataLight Communications
Sun Microsystems, Inc.
SwitchCore AB
Sycamore Networks
Syntune
Tektronix
Telcordia Technologies
Telecom Italia Lab
Tellabs
Texas Instruments
Time Warner Cable
Transwitch Corporation
Tyco Electronics
Verizon
Vitesse Semiconductor
Xilinx
ZTE Corporation