OIF OPTICAL INTERNETWORKING FORUM

IA Title: Scalable Serdes Framer Interface (SFI-S): Implementation Agreement for Interfaces beyond 40G for Physical Layer Devices.

IA # OIF-SFI-S-01.0

November 2008

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

OIF-SFI-S-01.0

OIF OPTICAL INTERNETWORKING FORUM

Working Group: **Physical and Link Layer (PLL) Working Group**

TITLE: **Scalable Serdes Framer Interface (SFI-S): Implementation Agreement for Interfaces beyond 40G for Physical Layer Devices.**

SOURCE: **TECHNICAL EDITOR**   **WORKING GROUP CHAIR**

Klaus-Holger Otto   David R. Stauffer, Ph. D.
Alcatel-Lucent   IBM Corporation
Thurn-und-Taxis-Str. 10   1000 River Road, MC 862J
90411 Nuremberg, Germany   Essex Jct., VT 05452
Phone: +49 911 526 3594   Phone: (802) 769-6914
Email: khotto@alcatel-lucent.com   Email: dstauffe@us.ibm.com

WORKING GROUP VICE CHAIR

Karl Gass
Sandia National Laboratories
P. O. Box 5800 MS-0874
Alburquerque, NM 87185
Phone: (505) 844-8849
Email: kgass@sandia.gov

ABSTRACT: This Implementation Agreement describes a scalable interface between Serdes and Framer devices within the Physical Layer (Serdes Framer Interface or SFI). SFI-S is an interface, based on 4 - 20 data lines plus deskew channel, for aggregate data bandwidths in the range of 40 – 160 Gbps data rate.

OIF OPTICAL INTERNETWORKING FORUM

The OIF is an international non profit organization with over 80 member companies, including the world's leading carriers and vendors. Being an industry group uniting representatives of the data and optical worlds, OIF's 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. Working relationships or formal liaisons have been established with IEEE 802.1, IEEE 802.3ba, IETF, IP-MPLS Forum, IPv6 Forum, ITU-T SG13, ITU-T SG15, MEF, ATIS-OPTXS,ATIS-TMOC, TMF and the XFP MSA Group.

For additional information contact:
The Optical Internetworking Forum, 48377 Fremont Blvd.,
Suite 117, Fremont, CA 94538
510-492-4040 ❖ info@oiforum.com
www.oiforum.com

1   <u>Table of Contents</u>

2   List of Figures

3   List of Tables

4   <u>Document Revision History</u>

Working Group: **Physical and Link Layer (PLL) Working Group**

SOURCE:	**Editor's Name**	**Working Group Chair**
	Klaus-Holger Otto	David R. Stauffer, Ph. D.
	Alcatel-Lucent	IBM Corporation
	Thurn-und-Taxis-Str. 10	1000 River Road, MC 862J
	90411 Nuremberg, Germany	Essex Jct., VT 05452
	Phone: +49 911 526 3594	Phone: (802) 769-6914
	Email: khotto@alcatel-lucent.com	Email: dstauffe@us.ibm.com
	Working Group Vice Chair	
	Karl Gass	
	Sandia National Laboratories	
	P. O. Box 5800 MS-0874	
	Alburquerque, NM 87185	
	Phone: (505) 844-8849	
	Email: kgass@sandia.gov	

DATE: **November 2008**

Version 01.0 Finalized document version of OIF2007.304.03 according to
                        Principal Member Ballot #57

5   Introduction

The typical line interface of a communications system with 80 - 160Gb/s optical links is expected to consist of three separate devices; an optical module containing a Serdes component, a forward error correction (FEC) processor and a Framer. The interconnection between these devices is electrical and the maximum data rate per signal is less than the optical data rate. Thus, a multi-bit bus is required. It is desirable for the protocol of this bus to comply with the following objectives and requirements:

1.   The scalable interface shall support aggregation of 4 – 20 lanes with optimization points to be determined (e.g. 4, 5, 8, 10, 11, 12, 16, 20)
2.   The interface shall utilize CEI SR electrical specifications
3.   The implementation shall provide for scaling to future CEI-28 SR line rates
4.   The interface can be used for point-to-point connection between Framer and FEC processor, Framer and Serdes, FEC processor and Serdes. Requirements for electrical parameters and control are the same between the different components.
5.   The interface is intended to handle any format of data, which has been randomized, e.g. by a scrambler.
6.   Capable of driving at least 8" of FR4 interconnect with one connector
7.   Interface should be capable of operating indefinitely, without losing sync.
8.   Support both in-band and out-of-band Forward Error Correction (FEC). A flexible clocking arrangement accounts for the additional FEC overhead.
9.   The interface is independent of the type of optics – serial, CWDM or parallel, SMF or MMF.
10.   Support of skew compensation over the signals comprising the interface bus. The Deskew algorithm and the skew optimized transmission used should minimize the complexity of the Serdes component.
11.   Support of simple and robust clock and data recovery of the signals comprising the interface bus.
12.   The interface includes independent status monitoring and supports received loss of signal detection.
13.   Receive timing reference operates continuously, independent of link status.
14.   The receive side of the interface will tolerate being driven while powered down without damage.
15.   The interface includes data path link verification for component qualification and system integration test purposes.

5.1   SFI-S System Reference Model

The following is a general synopsis of the SFI-S interface. For reference, a general block diagram is shown in Figure 1. SFI-S is the interface between the Serdes component, the forward-error-correction (FEC) processor, and the Framer. It is designed to meet requirements of this particular application, although it may also be used in other applications. "Receive" and "Transmit" refer to data flow and associated control/status information from the optics-to-system direction, and from the system-to-optics direction, respectively. Note that the two instantiations of the SFI-S bus in Figure 1 are independent and may operate at different frequencies. The variable n represents the scalability of this interface and can be selected in the range from 4 to 20 lanes.
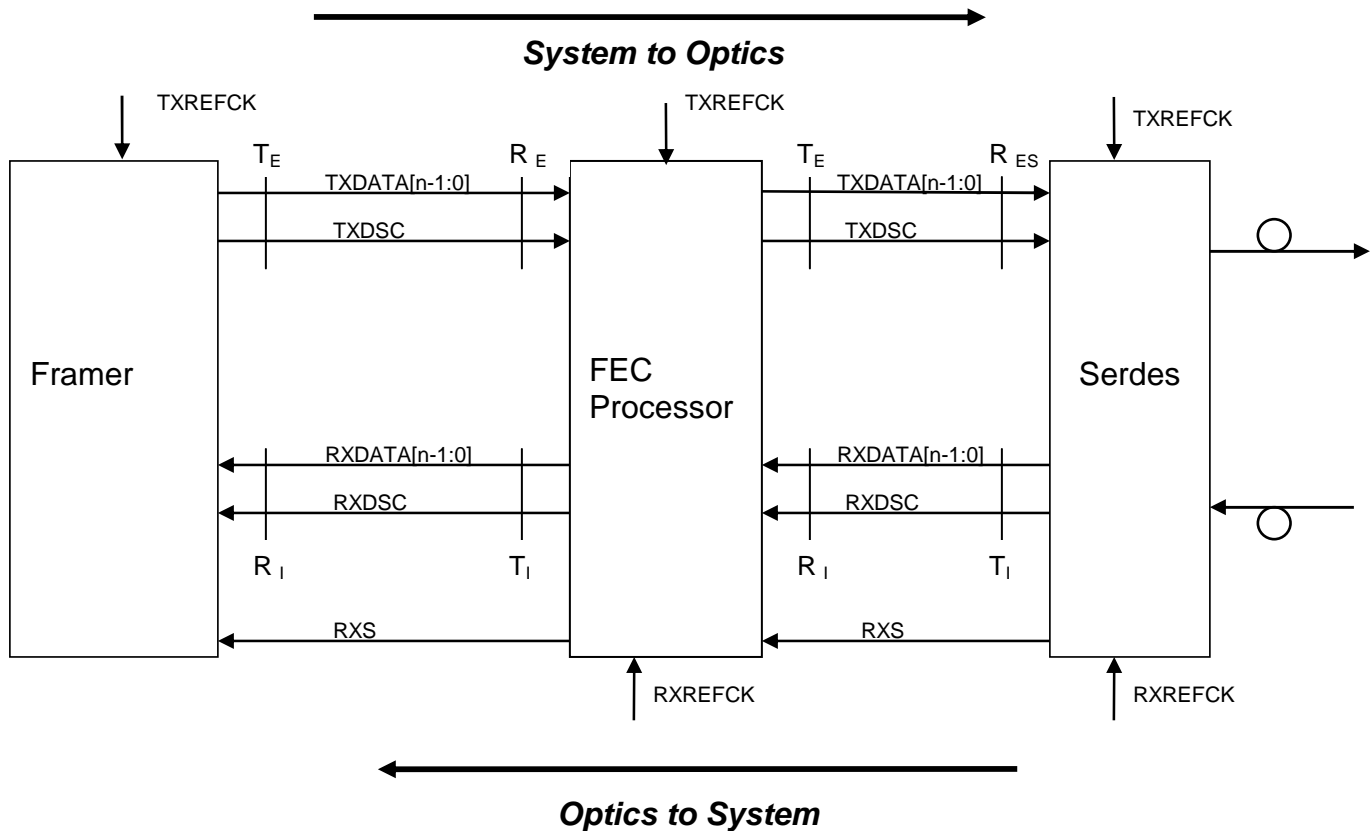
Figure 1: System Reference Model

The points T_I, R_I, T_E and R_E / R_{ES} are the reference points as defined in the common electrical specification OIF-CEI-02.0 11G+ SR [1].

5.2   SFI-S General Description

On both receive and transmit interfaces, control and status information is sent separately from the corresponding data path. To accommodate the expected minute frequency offsets between the data stream in the two directions, the Receive and Transmit interfaces operate independently.

The SFI-S bus has the following general characteristics:
1.   Point-to-point connections applicable for connection of the Serdes component to the FEC processor, the FEC processor to the Framer or the Serdes component directly to the Framer.
2.   n-bit wide (n = 4 to 20) data bus with each channel operating at CEI defined data rates. The Deskew method can be scaled to future CEI-28 SR speeds.
3.   Electrical parameters are defined in the common electrical specification OIF-CEI-02.0 [1]. In principle it will be possible to reuse this interface with future electrical specifications like CEI-28 SR. In case of reuse with CEI-28 SR it should be mentioned, that the jitter, skew and wander budget of the interface may need to be revisited.
4.   Deskew Channel included in both RX and TX directions contains out-of-band data samples to enable Deskew algorithm. Deskew Channel is one additional data channel (n + 1), running continuously at full data rate. Deskew algorithm will be operating continuously to monitor skew tracking.
5.   By using an alternating odd/even parity frame format on the deskew channel a minimum toggling rate of 1 every 18 bits is guaranteed. This can be used to recover the SFI-S receive clock on the deskew channel only and use DLLs instead of CDRs on the data channels.
6.   Minimize power consumption and the number of I/O signals to simplify PC board manufacturing.
7.   Maximize commonality of receive and transmit directions and of instantiations in different applications of the bus.

6   Generic SFI-S Deskew Channel Reference Frame Definition

To support a flexible width of 4 to 20 data lanes of the parallel SFI-S interface with a small number of physical building blocks for an ASIC design, a modular approach for generating the Deskew Reference Frame is chosen.

The basic building block for the Deskew Reference Frame is called a Reference Frame Element. The two different Reference Frame Elements are defined like this:

- The Even Parity element consists of 4 data lane bit samples, followed by the even parity (Peven) of the 4 bits. Its length is therefore 5 bits.
- The Odd Parity element consists of 4 data lane bit samples, followed by the odd parity (Podd) of the 4 bits. Its length is therefore 5 bits.
  To optionally randomize the data stream to improve the transition density, it shall be provisionable to use inverted data during Odd Parity elements for transmission and deskew generation.

To build the Reference Frame to support a flexible number n (from 4 to 20) of data lanes the following rules, how to concatenate frame elements and how to fill the bit samples apply:

1.   A reference frame shall always start with an even parity element.
2.   A reference frame shall always end with an odd parity element to achieve a reasonable toggle rate on the deskew channel for CDR usage. The shortest possible Deskew frame consists of one even parity element followed by one odd parity element.
3.   The number of required frame elements is determined by the number of data lanes n divided by the number of data bit samples (4) per element. For n = 4 to 8 lanes this results in 2 frame elements, n = 9 to 12 results in 3 frame elements, n = 13 to 16 results in 4 frame elements and n = 17 to 20 results in 5 frame elements being required.
4.   If more then two frame elements are required, a frame start shall be marked by two consecutive even parity elements. Remaining frame elements shall alternate odd and even parity elements as often as possible within the rules as defined above. This defines a 3 element reference frame as **even, even, odd**, a 4 element reference frame as **even, even, odd, odd** and a 5 element reference frame as **even, even, odd, even, odd**.
5.   The bit samples of the reference frame shall be filled, starting at the highest number data lane RX/TXDATA(n-1), which is filled into the first sample position of the even parity element at the beginning of the frame. RX/TXDATA(n-2) is filled into the second sample position and so on, until RX/TXDATA(0) is filled in. If the RX/TXDATA(0) sample does not land in the last sample position of the frame ending odd element, the remaining bit samples are filled again, starting from RX/TXDATA(n-1). As for

validation of the lane to lane skew a single bit sample per lane each frame
is sufficient, the second samples of the remaining locations in the frame
can be ignored by the deskewing algorithm.

The result of these rules is shown in Figure 2 for an example with n = 10 data
lanes. The number of required frame elements to handle 10 lanes is 3; therefore
an even, even, odd parity frame is selected. Sampling starts at DATA[9], and fills
half of the final odd parity element with Data[1] and Data[0]. The grey shaded
samples of Data[9] and Data[8] mark the redundant samples to fill the frame,
which could be ignored by the deskewing algorithm at the sink side of the link.
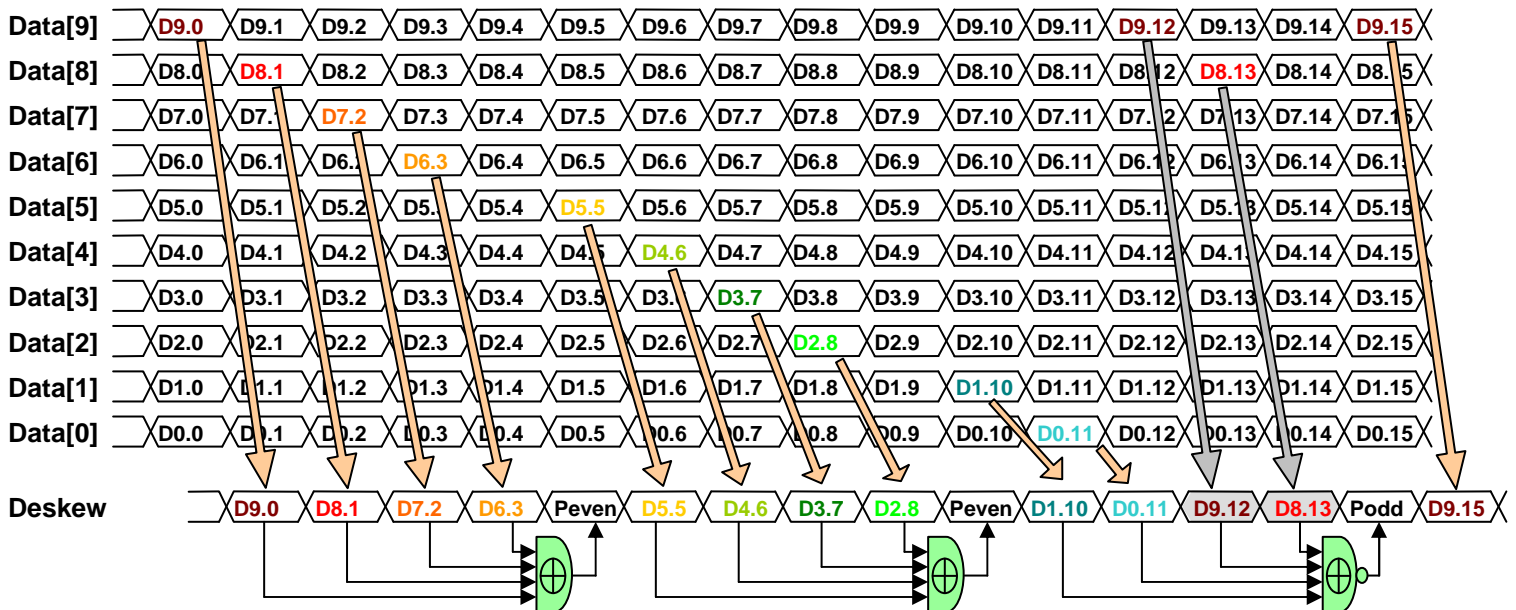
Figure 2: Example Reference Frame Generation for n = 10

The minimum toggle rate for this example reference frame is one toggle in 18
bits, due to the nature of the selected reference frame format.

Although the scope of this IA was limited to 4 to 20 lanes support, this deskew
channel frame concept is in principle scalable beyond this point. For example for
21 to 24 lanes the reference frame would be defined by above rules to consist of
the 6 frame elements: **even, even, odd, even, odd, odd**.

7   Signal definition

7.1   Receive signals

The Receive signals are related to the transport of data from the optics towards the system. They are applicable to conducting data from the Serdes component to the FEC processor, from the FEC processor to the Framer, or from the Serdes Component directly to the Framer. All Receive signals, unless otherwise specified below, are differential CML as defined in the Common Electrical Implementation agreement OIF-CEI-02.0 [1].

Signal Name	Direction	Function
RXDATA [n-1:0]	Optics to System	The **Receive Data** (RXDATA [n-1:0]) signals carry the data in the optics to system direction. Serial optical data is bit striped onto RXDATA [n-1:0] in a round-robin fashion. RXDATA [n-1] contains the first bit received while RXDATA [0] the last bit received. Each RXDATA [#] signal contains every n^{th} bit of the data stream. Optionally RXDATA[#] can be inverted during (see section 10) every odd parity element of the deskew frame for 5 data samples.
RXDSC	Optics to System	The **Receive Deskew Channel** (RXDSC) signal provides reference data to enable skew measurements of the Receive data bus (RXDATA [n-1:0]). Depending on the number of data lanes n RXDSC contains 2 to 5 5-bit long reference frame elements. A two elements frame is composed of alternating even and odd elements. A three elements frame consists of even, even and odd elements, where even, even marks the frame start. Consequently the four element frame is even, even, odd, odd. Samples are taken from the Receive data bus in a round-robin fashion, starting with RXDATA [n-1] and ending with RXDATA [0]. If the frame holds more samples then lanes are available, the remaining samples are taken again from the upper number lanes starting with RXDATA [n-1]. In case of the optional data inversion (see section 10) is selected, the data samples are taken after inversion

Signal Name	Direction	Function
RXREFCK	Optics to System	The **Receive Reference Clock** (RXREFCK) signal provides reference timing to the Receive interface. RXREFCK is nominally a 50% duty cycle clock with a frequency that is $1/16^{th}$ of the data bit rate of RXDATA and RXDSC. In the FEC processor, RXREFCK is the frequency reference to the source of RXDATA, RXDSC. Jitter characteristics of RXREFCK do not directly concern interoperability and are beyond the scope of this implementation agreement. Implementation of RXREFCK is mandatory for a Serdes and FEC component, and not necessary for a Framer component. In some implementations, RXREFCK could share a common pin pair with TXREFCK.
RXS	Optics to System	The **Receive Status** (RXS) signal carries status from the Serdes component to the FEC processor, from the FEC processor to the Framer, or from the Serdes component directly to the Framer. The encoding of RXS is: RXS = 'b0: Idle, RXS = 'b1: Receive alarm, Receive alarm shall indicate that RXDATA is not derived from the optical receive signal. The electrical I/O type shall be LVCMOS. It is mandatory for the source device of the receive interface to generate RXS. It is optional for the sink device in the receive interface to use RXS. RXS is an asynchronous signal.

Table 1: SFI-S Receive Signal Summary

7.2 Transmit Signals

The Transmit signals are related to the transport of data from the system towards the optics. They are applicable to conducting data from the Framer to the FEC processor, from the FEC processor to the Serdes Component, or from the Framer directly to the Serdes component. All Transmit signals, unless otherwise specified below, are differential CML as defined in the Common Electrical Implementation Agreement OIF-CEI-02.0 [1].

Signal Name	Direction	Function
TXDATA [n-1:0]	System to Optics	The **Transmit Data** (TXDATA [n-1:0]) signals carry data in the system to optics direction. Data on TXDATA [n-1:0] are placed on the transmit optical stream in a round-robin fashion. TXDATA [n-1] contains the first bit transmitted while TXDATA [0] the last bit transmitted. Each TXDATA [#] signal contains every n^{th} bit of the transmitted data stream. TXDATA [n-1:0] is frequency locked to TXREFCK with specified maximum skew between all channels. Optionally TXDATA[#] can be inverted (see section 10) during every odd parity element of the deskew frame for 5 data samples.

Signal Name	Direction	Function
TXDSC	System to Optics	The **Transmit Deskew Channel** (TXDSC) provides reference data to enable skew measurements of the Transmit data bus (TXDATA [n-1:0]) between FEC processor and Framer. Depending on the number of data lanes n TXDSC contains 2 to 5 5-bit long reference frame elements. A two elements frame is composed of alternating even and odd elements. A three elements frame consists of even, even and odd elements, where even, even marks the frame start. Consequently the four element frame is even, even, odd, odd. Samples are taken from the Transmit data bus in a round-robin fashion, starting with TXDATA [n-1] and ending with TXDATA [0]. If the frame holds more samples then lanes are available, the remaining samples are taken again from the upper number lanes starting with TXDATA [n-1]. In case of the optional data inversion (see section 10) is selected, the data samples are taken after inversion
TXREFCK	System to Optics	The **Transmit Frequency Reference** (TXREFCK) signal provides the frequency reference for the Transmit data path devices (Framer, FEC, and Serdes). TXREFCK is nominally a 50% duty cycle clock with a frequency that is $1/16^{th}$ of the data bit rate of TXDATA and TXDSC. TXDATA and TXDSC shall be frequency locked to TXREFCK. It is mandatory for one device in the Transmit chain to have an external clock source connected to TXREFCK.

Table 2: SFI-S Transmit Signal Summary

8   Logical Reference Models

8.1   Serdes component Receive interface

Figure 3 below shows a logical model of the Receive data interface for the first channel or fiber in a Serdes component. The intent is to show the source of data in the Receive Deskew Channel (RXDSC). No limit is placed on device implementation, and this model is not intended to represent an actual design.
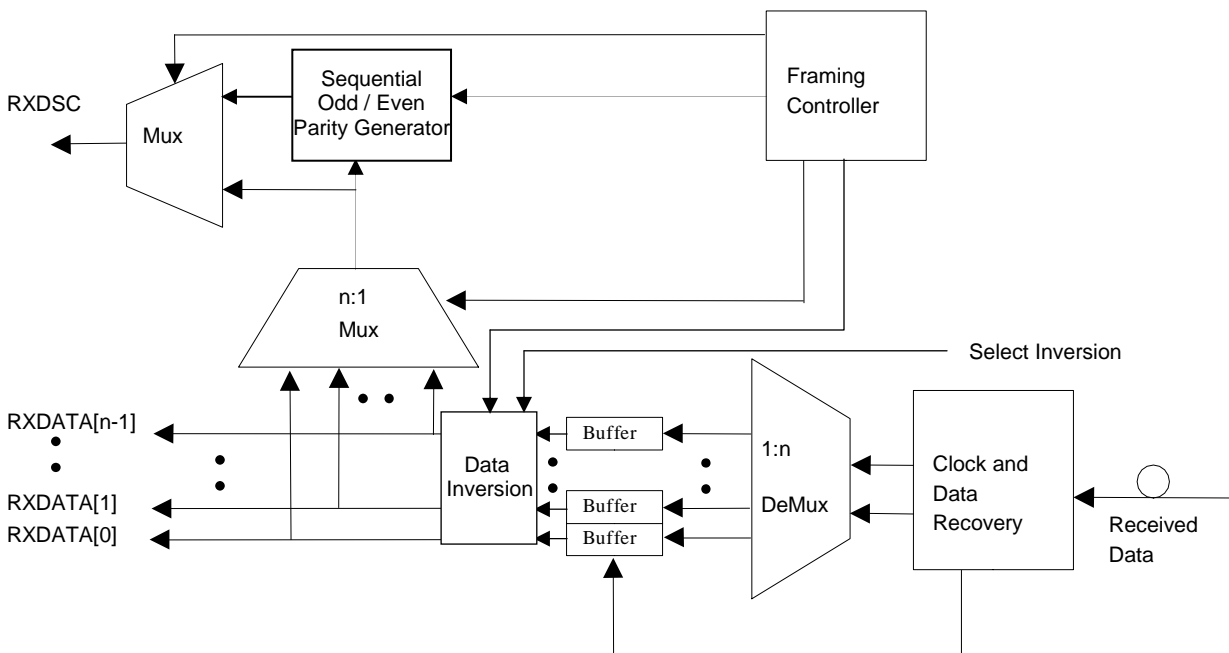
Figure 3: Model of Serdes component - Receive I/F Source

The data in the optical stream from each fiber is striped across the n bit lanes of the Receive Data bus (RXDATA [#]) in a round-robin fashion. The first bit received from the fiber is written into the re-timing buffer associated with RXDATA [n-1] and the last bit into that associated with RXDATA [0]. The re-timing buffers act as a set of FIFOs to bridge between the optical timing domain and the Receive interface-timing domain. The re-timing buffers absorb the timing offset between the bit lanes.

The data inversion can be optionally selected to invert 5 consecutive data samples during the odd parity elements of the deskew frame, as described in section 10.

RXDSC replicates the data bits sent on each signal of the Receive Data bus RXDATA [n-1:0] cyclically. First, each RXDATA [#] signal is sampled, in turn, for

1 bit time. Bit sampling begins with RXDATA [n-1] and ends with RXDATA [0]. Every 4 samples it is followed by the corresponding even or odd parity bit over the 4 data channel samples, generated by the parity generator block. The sequence of even and odd parity frame elements depends on the required frame composition, which differs for each dedicated number n of aggregated interface lanes. The detailed rules for the frame generation can be found in section 6. After a complete frame has been sent, a new reference frame is initiated on RXDSC and then continuously generated. The generation process for an example with n=10 is illustrated in Figure 2.

8.2 Serdes Component Transmit interface

Figure 4 below shows a logical model of the Transmit interface to the Serdes Component. The intent is to show operation of the Deskew algorithm, controlled by the Transmit Deskew Channel TXDSC. No limit is placed on device implementation, and this model is not intended to represent an actual design.
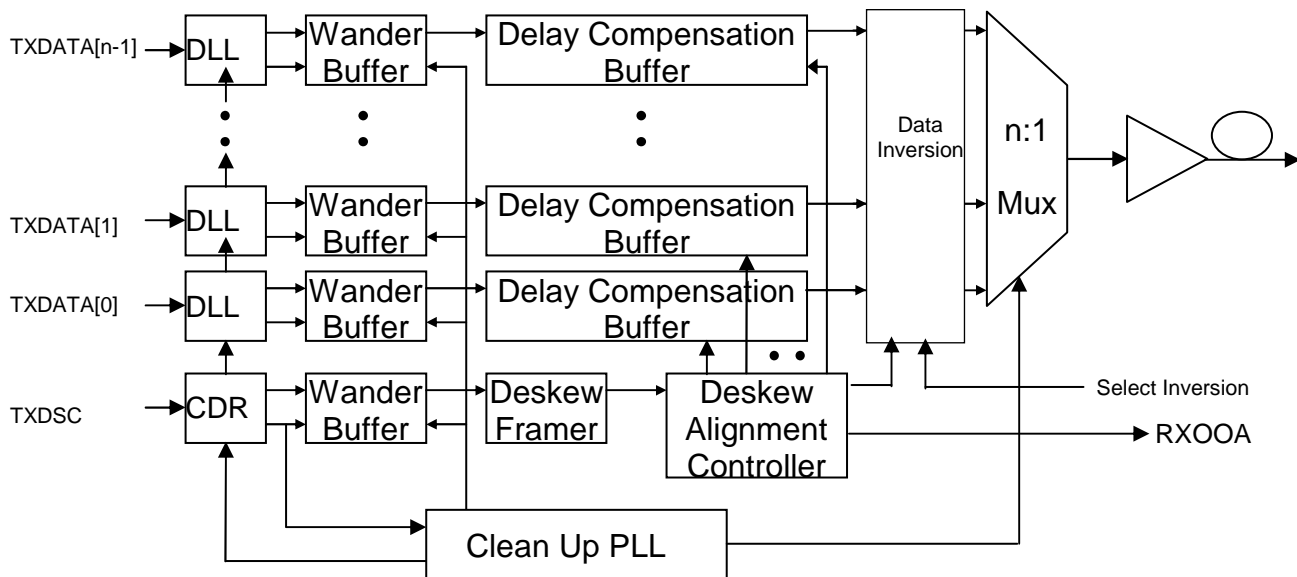
**Figure 4: Model of Serdes component - Transmit I/F Sink**

The Clock and Data Recovery (CDR) block tracks the center of the "eye" of the Transmit Deskew Channel TXDSC signal and recovers the clock for the complete SFI-S interface. Delay locked loops (DLL) are sufficient to track the center of the "eye" of the Transmit Data bus TXDATA [#] signals. The wander buffers bridge between the system timing domain and the clean (low jitter) line-timing domain. The delay compensation buffers absorb timing offsets between the bit lanes.

The function of the Deskew Framer block is to recognize the alternating odd/even parity framing bits on the Transmit Deskew Channel TXDSC. These identify the position of the reference data bits, which are replicated from each of TXDATA [n-1:0]. Each TXDATA bit is then compared with the sampled data bit in the Deskew Channel in turn according to section 6, see also Figure 2.

The Out-of-alignment alarm (RXOOA) is set, if a match has not been found on any of the n data channels. When a data match has been found on all n data channels and stable skew data derived, then the De-skew algorithm is considered as locked and RXOOA is cleared. Skew measurements are monitored continuously; RXOOA alarm remains cleared as consistent skew data is generated.

The function of the Deskew Alignment Controller is to continuously compare data on the Deskew Channel TXDSC with the respective Data Channels TXDATA [n-1:0]. Any mismatch errors can be detected, which would represent errors generated over the SFI-S interface. These errors should be reported as part of the minimum test requirements for the interface.

The data inversion can be optionally selected to re-invert the 5 consecutive data samples during the odd parity elements of the deskew frame, as described in section 10.

8.3   Receive Interface of FEC Processor / Framer

Figure 5 below shows a model of the Receive interface in the FEC processor or in the Framer. The intent is to show operation of the Deskew algorithm, controlled by the Receive Deskew Channel RXDSC. No limit is placed on device implementation, and this model is not intended to represent an actual design.

Figure 5: Model of FEC Processor / SONET Framer - Receive I/F Sink

The Clock and Data Recovery (CDR) blocks track the center of the "eye" of the Receive Data bus RXDATA [#] and the Receive Deskew Channel RXDSC signal. The re-timing buffers bridge between the Receive interface timing domain and the system timing domain. The re-timing buffers absorb timing offsets between the bit lanes.

The function of the Deskew Controller block is to recognize the alternating odd/even parity framing bits on the Receive Deskew Channel RXDSC. These identify the position of the reference data bits, which are replicated from each of RXDATA [n-1:0]. Each RXDATA channel is then compared with the sampled data in the Deskew Channel in turn according to section 6, see also Figure 2.

The Out-of-alignment alarm (RXOOA) is set if a match has not been found on any of the n data channels. When a data match has been found on all n data channels and stable skew data derived, then the De-skew algorithm is considered as locked and RXOOA is cleared. Skew measurements are monitored continuously; RXOOA alarm remains cleared as consistent skew data is generated.

The function of the Deskew Controller is to continuously compare data on the Deskew Channel RXDSC with the respective Data Channels RXDATA [n-1:0]. Any mismatch errors can be detected, which would represent errors generated over the SFI-S interface. These errors should be reported as part of the minimum test requirements for the interface.

The data inversion can be optionally selected to re-invert the 5 consecutive data samples during the odd parity elements of the deskew frame, as described in section 10.

8.4   Transmit interface in FEC Processor / Framer

Figure 6 below shows a model of the Transmit interface in the FEC processor or in the Framer. The intent is to show the source of data in the Transmit Deskew Channel. No limit is placed on device implementation, and this model is not intended to define an actual design.

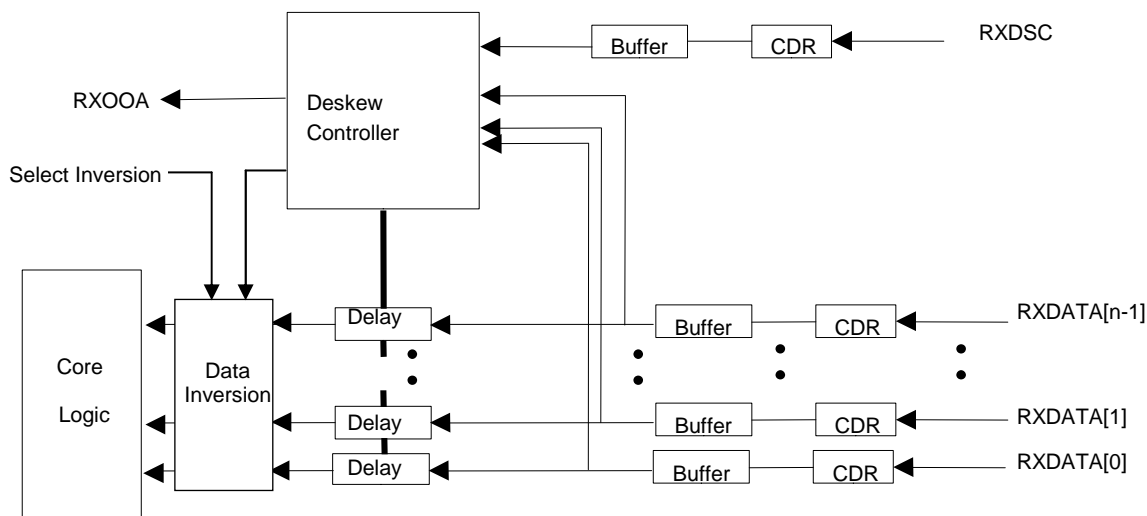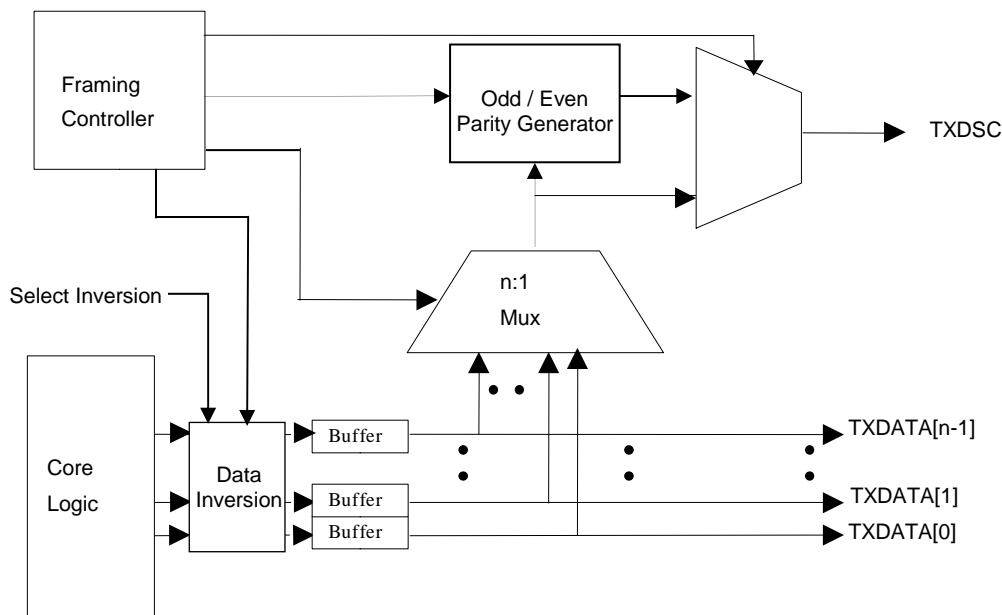**Figure 6: Model of FEC Processor / SONET Framer - Transmit I/F Source**

Data from the core logic of the FEC Processor or Framer is striped across the n bit lanes of the first Transmit Data bus TXDATA [n-1:0] channel in a round-robin fashion. The first bit received is written into the re-timing buffer associated with TXDATA [n-1] and the last into that associated with TXDATA [0]. The re-timing

buffers act as a set of FIFOs to bridge between the core logic-timing domain and the Transmit interface-timing domain.

The data inversion can be optionally selected to invert 5 consecutive data samples during the odd parity elements of the deskew frame, as described in section 10.

TXDSC replicates the data bits sent on each signal of the Receive Data bus TXDATA [n-1:0] cyclically. First, each TXDATA [#] signal is sampled, in turn, for 1 bit time. Bit sampling begins with TXDATA [n-1] and ends with TXDATA [0]. Every 4 samples it is followed by the corresponding even or odd parity bit over the 4 data channel samples, generated by the parity generator block. The sequence of even and odd parity frame elements depends on the required frame composition, which differs for each dedicated number n of aggregated interface lanes. The detailed rules for the frame generation can be found in section 6. After a complete frame has been sent, a new reference frame is initiated on TXDSC and then continuously generated. The generation process for an example with n=10 is illustrated in Figure 2.

9 SFI-S Bit-lane Deskew

In operation the data signals may encounter different delays in transit from the SFI-S source device to the sink device. The maximum differential skew introduced by the connection system is specified in section 9.2. The earliest arriving signal may lead the latest arriving by j bits. Relative to the earliest, each of the remaining signals is coincident, or is up to j unit intervals late. The search space for determining the relative delays of 5 signals on SFI-S is (j+1) * 5 combinations. To make the problem tractable, a reference signal is used in SFI-S, known as the Deskew Channel, to allow each bit-lane to independently measure its own delay relative to the reference signal. Having de-coupled the measurement to individual signals, the search space becomes trivially small.

The bit-lane Deskew function is shared between the SFI-S source and sink devices at either end of the Receive and Transmit interfaces. In the source device, data is sampled from each of the n data channels sequentially, and copied onto the Deskew Channel. The Deskew Channel is then sent with the n data channels to the sink device over the SFI-S interface.

Data input to the sink device, will be skewed by the different delays in each of the data channels. It is the function of the Deskew algorithm operating in the sink device to measure the amount of skew on each data channel, and then to use this skew information to compensate for the amount of skew. The bit lane Deskew algorithm will make an initial skew measurement on initial power up or connection. Subsequent to skew measurement, external conditions may vary causing the skews to change. It is a requirement that devices that are compliant to this implementation agreement be able to track skew changes of the minimum value to the maximum value without introducing errors. The Deskew algorithm will operate continuously during normal operation of the SFI-S interface, and continuously track skew.

9.1 Deskew of Receive interface

At the receive interface, a reference frame is generated in the source device, like described in section 6. The data bit samples are taken from the each of the Receive data bus RXDATA as defined in section 6, see also Figure 2.

Reference frames are sent continuously over the receive interface to enable the Deskew algorithm in the sink device and continuous monitoring of skew.

The sink device monitors the Receive Deskew Channel (RXDSC) for the reference data. It shall adjust the delay of each RXDATA [#] on a unit interval by unit interval basis, such that the delay from source device to the output of the re-timing buffers at the sink device is identical for all data signals. When the skew of all data channels is compensated and locked, the RXOOA alarm is cleared. Until

the Deskew Algorithm finds lock on all n data channels, the Receive interface is in an out-of-alignment state, and the RXOOA alarm is set. The Deskew algorithm will continue to operate after the RXOOA is cleared. Under normal circumstances, the interface will operate continuously with skew being monitored, and the RXOOA alarm remains off. If failure of the Deskew algorithm occurs, and any of the data channels fall out of lock, then the RXOOA alarm will be set. The RXOOA alarm monitors de-skew out-of-alignment of the channel.

9.2 Interface Skew Budget

In OIF-CEI-02.0 [1] the RX and TX skew between channels (at the package pins) is defined to be less than 500ps or for example 5.50 UI @ 11.2 Gbps. These values are mainly driven by the desire to simplify the CMOS high-speed interface design. In bipolar technologies, as used for Serdes design, these values could be defined to be much smaller. This knowledge can be used to make the deskew buffers in a Serdes design significantly smaller (and therefore less power hungry).

To optimize the deskew concept for the Serdes design PCB traces and connector from FEC processor/Framer towards Serdes (Point R_{es} in Figure 1) shall be designed with a tight skew budget. This skew budget mainly defines the required size of the Serdes internal deskew alignment buffer. The skew budget for this interface is shown in Table 3. Please note, that the absolute values (in ps) reflect the actual requirement, and the relative values (in UI) are provided as an example for reference only.

Skew Contributor	Skew Budget @ 6.25 Gbps	Skew Budget @ 11.2 Gbps	Skew Budget [ps]
CEI TX Interface @ package pins (Te)	3.13 UI	5.50 UI	500
PCB and Connector	0.3 UI	0.6 UI	50
Subtotal @ package pins (Res)	3.43 UI	6.10 UI	550
Serdes Input	0.3 UI	0.6 UI	50
Total	3.75 UI	6.70 UI	600

Table 3: Channel to Channel Skew Budget for FEC processor/Framer to Serdes Link

For all other applications of the interface, where CMOS devices receive the deskew channel, the size of the deskew alignment buffer is only a minor issue. Therefore for those applications the PCB and connector skew budget can be

defined to be very relaxed, as the possibility for skew compensation of this deskew method is virtually infinite. The relaxed skew budget for these applications is shown in Table 4.

Skew Contributor	Skew Budget @ 6.25 Gbps	Skew Budget @ 11.2 Gbps	Skew Budget [ps]
CEI TX Interface @ package pins (Ti/Te)	3,13 UI	5.50 UI	500
PCB and Connector	3.13 UI	5.50 UI	500
Subtotal @ package pins (Ri/Re)	6.26 UI	11.00 UI	1000
CEI RX Interface incl. package	3.13 UI	5.50 UI	500
Total	9.39 UI	16.50 UI	1500

Table 4: Relaxed Channel to Channel Skew Budget for all other interface applications

9.3   Complete Jitter/Wander/Skew budget

A complete Jitter/Wander/Skew budget for 11.2 Gbps is shown in Table 5. For
the location of the listed system points, please refer to Figure 1. In the tables
below jitter is as defined in OIF-CEI [1] in UI, wander is defined normative in UI
(below the jitter cut-off frequency), where as for skew the time in ps is normative
as defined in section 9.2.

Parameter	Signal	System Points			Units
	Type	Ti/Te	Ri/Re	Res	
Skew (Absolute values in [ps] as defined in section 9.2)	Data	5.50	11.00	6.10	UI peak @ 11.2 Gbps
Correlated Wander	All	5.00	7.00	7.00	UI peak to peak
Uncorrelated Wander	All	0.65	0.75	0.75	UI peak to peak
Total Wander	All	5.65	7.75	7.75	UI peak to peak
Relative Wander	All	1.30	1.50	1.50	UI peak to peak
Skew + (Rel. Wander)/2	All	6.15	11.75	6.85	UI peak to peak
Deterministic Jitter	Data	0.15			UI peak to peak
Total Jitter	Data	0.30	0.65	0.65	UI peak to peak

Table 5: Complete Jitter/Wander/Skew Budget Example for 11.2 Gbps

A second example for a budget at 6.25 Gbps is shown in Table 6.

Parameter	Signal	System Points			Units
	Type	Ti/Te	Ri/Re	Res	
Skew (Absolute values in [ps] as defined in section 9.2)	Data	3.13	6.26	3.43	UI peak @ 6.25 Gbps
Correlated Wander	All	5.00	7.00	7.00	UI peak to peak
Uncorrelated Wander	All	0.65	0.75	0.75	UI peak to peak
Total Wander	All	5.65	7.75	7.75	UI peak to peak
Relative Wander	All	1.30	1.50	1.50	UI peak to peak
Skew + (Rel. Wander)/2	All	3.78	7.01	4.18	UI peak to peak
Deterministic Jitter	Data	0.15			UI peak to peak
Total Jitter	Data	0.30	0.65	0.65	UI peak to peak

Table 6: Complete Jitter/Wander/Skew Budget Example for 6.25 Gbps

10   Inversion of data lanes during odd parity element of deskew frame to resolve potential transition density issues (CID)

To resolve a potential transition density issue on the data lanes, a simple to implement inversion of the 5 data samples per odd parity element shall be used.

The inversion shall be implemented prior to the deskew sampling and on the odd half of the deskew frame. This means the inverted data bits go into the deskew signal and are used to calculate the odd parity.

The deskew frame counter can be used to determine the position for re-inverting the data bits after receiving and de-skewing them on the other side of the SFI-S link.

This process shall look like shown in Figure 7.
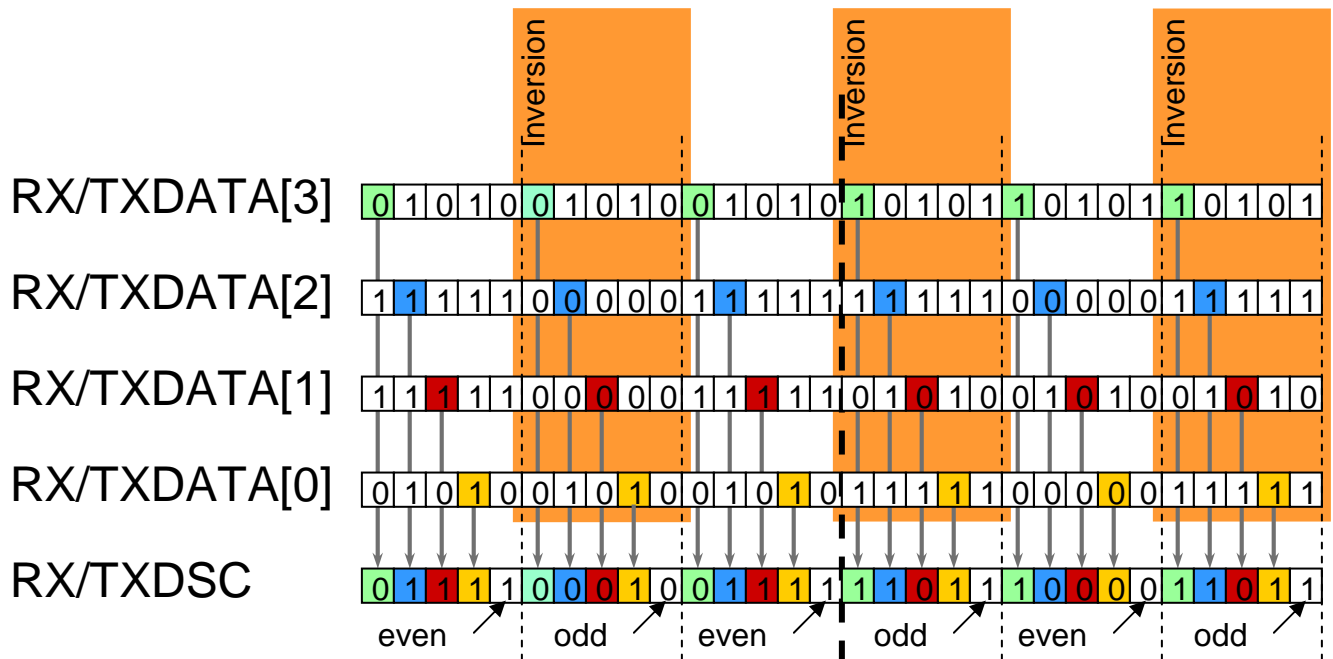
Figure 7: Bit-striped 0xF6/0x28 example with data inversion at odd parity elements of deskew frame

This data bit inversion shall be a selectable feature and shall be used for applications, where a proper transition density on the data lanes is in question. For instance to ease the implementation of long PRBS sequences with current test equipment, this feature can be switched off, as not required for this kind of signal (No CID issue with PRBS signals).

Nevertheless it is also possible to enable the inversion for all other target applications, as it has no negative effect for signals, like OTU-4, 100GbE or PRBS, which don't have a CID problem, when bit-striped over n data lines.

11   Interface test requirements

It is a requirement of the SFI-S architecture that it will be capable of operating indefinitely without inducing errors. A minimum set of test requirements is defined to indicate that the interface is set up correctly and that it monitors itself for continuous reliable operation. The following functions are necessary to provide a minimum set of test functions:

1. Confirm that each of the n data channels is connected correctly in both Receive and Transmit directions. The Deskew Controller in each direction can implement this function. The Deskew Channel (RXDSC and TXDSC) is required to match with each of the data channels in turn. If persistent mismatch occurs with any of the n channels, then the RXOOA or TXOOA alarm is raised relating to the respective Receive or Transmit directions. The minimum requirement is to monitor each alarm. It would be possible to include functionality to indicate which Channel is faulty – this would be within the scope of individual implementations and not mandatory in this implementation agreement.

2. Verify that the Deskew function correctly compensates for Skew over the SFI-S interface. It is the function of the Deskew Algorithm to measure the level of skew on each data channel in both Receive and Transmit directions and to control the delay compensation on each channel. When a stable skew measurement is made on each channel, the interface is considered in lock, and the RXOOA and TXOOA alarms are cleared. These alarms should indicate that the Deskew compensation is working correctly after initial power up and during continuous operation.

3. Error checking during continuous operation. It is possible to use the Deskew Channel in both Receive and Transmit directions to detect gross errors over the interface. The function of the Deskew Controller in the sink side of each interface is to compare the replicated data on the Deskew Channel with the respective data on each data channel. Mismatch errors can then be detected in the data sample. It is a requirement of the interface to continuously monitor for these potential errors. The procedure for reporting errors is not part of the SFI-S implementation agreement, but should be included as part of the management function.

11.1   <u>Optional Embedded Test Features (Informative Part)</u>

To simplify debugging of the transmission quality it is recommended to implement in devices with SFI-S links the capability to generate and analyze PRBS sequences per lane and per complete link. For a more complete set of test patterns refer to CEI-02.0 Implementation Guide [5] and CEI Short Stress Patterns White Paper [6].

12 <u>References</u>

12.1 <u>Normative references</u>

[1] Optical Internetworking Forum OIF-CEI-02.0, " Common Electrical I/O (CEI) – Electrical and Jitter Interoperability agreement for 6G+ bps and 11G+ bps I/O".

12.2 <u>Informative references</u>

[2] ANSI - T1.105-1995, "Synchronous Optical Network (SONET) – Basic Description including Multiplex Structure, Rates, and Formats", 1995.

[3] ITU-T, Recommendation G.707/Y.1322 - "Network Node Interface For The Synchronous Digital Hierarchy", Dec 2003.

[4] ITU-T, Recommendation G.709/Y.1331 – "Interfaces for the Optical Transport Network (OTN)", Mar 2003.

[5] Optical Internetworking Forum CEI-02.0 Implementation Guide, "Implementation Guide for the Common Electrical Interface 2.0 (CEI 2.0)"

[6] Optical Internetworking Forum White Paper, "CEI Short Stress Patterns"

13   Appendix A: Recommended extended static Skew Compensation Capability

It is recommended that SFI-S Receivers are designed to tolerate a minimum of 84 U.I. of static skew at test points RES, RI, and RE. This provides the same skew tolerance as in the informative recommendation for SFI-5.2 receivers.

14   Appendix B: Recommended Pinout Order for PCB Routing Optimization of SFI-S Links

To avoid the need to cross over the n + 1 critical differential interface traces on the PCB, a pinout order, as described below, should be maintained by the SFI-S devices TXDATA_SRC/TXDATA_SNK [n-1:0] and TXDSC_SRC/TXDSC_SNK pins and also at the according transponder connector pinout.
The same pinout order should also be maintained in RX direction for signals RXDATA_SRC/RXDATA_SNK[n-1:0] and RXDSC_SRC/RX_DSC_SNK.
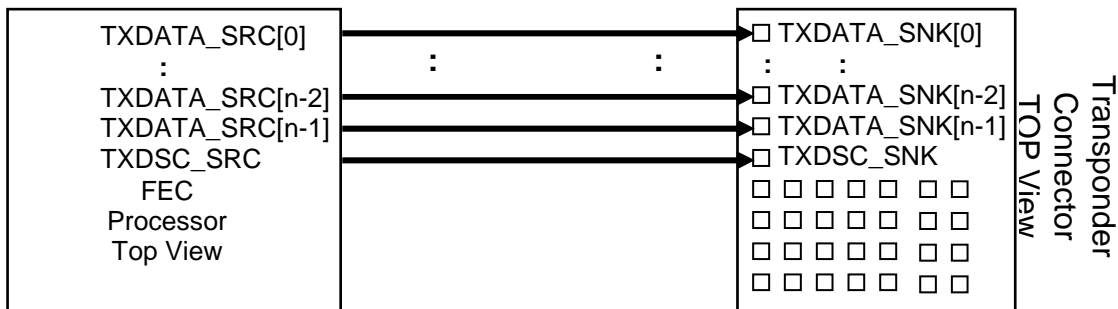

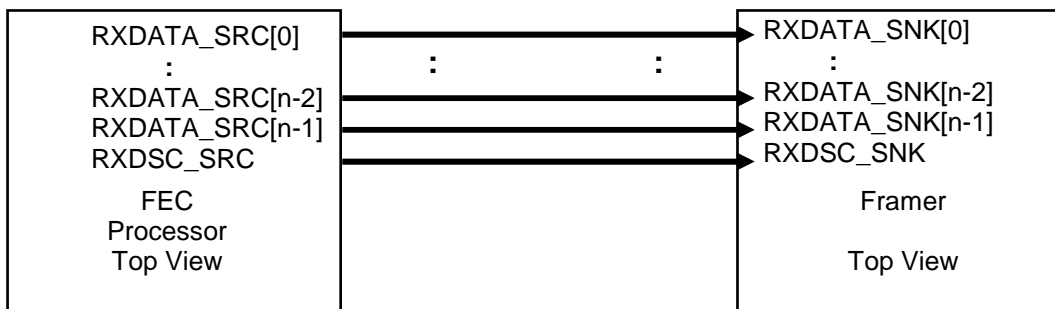**Figure 8: Pinout Order between SFI-S Device and Transponder Connector**

Figure 9: Pinout Order between SFI-S Devices

15   Appendix C: Glossary

CEI	Common Electrical I/O
CDR	Clock Data Recovery
CID	Consecutive Identical Digits
CML	Current Mode Logic
CWDM	Coarse Wavelength Division Multiplexing
DLL	Delay Locked Loop
FEC	Forward Error Correction
FIFO	First In First Out Buffer
LVCMOS	Low Voltage CMOS
MMF	Multi Mode Fiber
OC-N	Optical Carrier of level N
OTN	Optical Transport Network
OTU-N	Optical Transport Unit of level N
PCB	Printed Circuit Board
PLL	Phase Locked Loop
PRBS	Pseudo Random Binary Sequence
RX	Receive
SDH	Synchronous Digital Hierarchy
SERDES	Serializer De-serializer
SFI	SERDES Framer Interface
SMF	Single Mode Fiber
SONET	Synchronous Optical Network
STM-N	Synchronous Transport Module of level N
TX	Transmit

16   Appendix D: Open Issues / current work items

17   Appendix E: List of companies belonging to OIF when document is approved

ADVA Optical Networking	Flextronics	Opnext
Alcatel-Lucent	Force 10 Networks	Optametra
Altera	France Telecom	OpVista
AMCC	Fujitsu	Picometrix
Analog Devices	Furukawa Electric Japan	PMC Sierra
Anritsu	Huawei Technologies	Sandia National Laboratories
AT&T	IBM Corporation	Santur
Avago Technologies Inc.	IDT	Sierra Monolithics
Avalon Microelectronics	Infinera	Silicon Logic Engineering
Avanex Corporation	Inphi	Soapstone Networks
Bookham	IP Infusion	Stratalight
Broadcom	JDSU	Sycamore Networks
China Telecom	Juniper Networks	Syntune
Ciena Corporation	KDDI R&D Laboratories	Tektronix
Cisco Systems	Kotura	Telcordia Technologies
ClariPhy Communications	LSI Corporation	Telecom Italia Lab
CoreOptics	Marben Products	Tellabs
Cortina Systems	Mintera	TeraXion
Data Connection	MITRE Corporation	Texas Instruments
Department of Defense	Mitsubishi Electric Corporation	Time Warner Cable
Deutsche Telekom	Molex	Tyco Electronics
Discovery Semiconductors	NEC	u2t Photonics AG
Emcore	NeoPhotonics	Verizon
Ericsson	Nokia Siemens Networks	Vitesse Semiconductor
Eudyna Devices	Nortel Networks	Yamaichi Electronics Ltd.
Finisar Corporation	NTT Corporation	ZTE Corporation