OIF OPTICAL INTERNETWORKING FORUM

IA Title: Serdes Framer Interface Level 5 Phase 2 (SFI-5.2): Implementation Agreement for 40Gb/s Interface for Physical Layer Devices.

IA # OIF-SFI5-02.0

October 2, 2006

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

OIF OPTICAL
INTERNETWORKING
FORUM

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

TITLE: **Serdes Framer Interface Level 5 Phase 2 (SFI-5.2): Implementation Agreement for 40Gb/s Interface for Physical Layer Devices.**

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

Klaus-Holger Otto   David R. Stauffer, Ph. D.
Lucent Technologies   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@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 an interface between Serdes and Framer devices within the Physical Layer (Serdes Framer Interface or SFI). SFI-5.2 is an interface, based on 4 data lines plus deskew channel, for aggregate data bandwidths of OC-768, STM-256, OTN OTU3, as well as other applications at the 40 Gbps data rate.

1   Table of Contents

2   List of Figures

3   List of Tables

4   Document Revision History

Revision	Date	Description
OIF-SFI5-02.0	October 2006	IA Text derived from Draft IA text of document oif2005.305.03 accepted by OIF principal ballot #45

5 Introduction

The typical line interface of a communications system with 40Gb/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 will be electrical where 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. Support up to 44.4Gb/s bi-directional Aggregate data throughput such as SONET OC-768 [2], SDH STM-256 [3], OTN OTU3 [4] and other systems operating at a payload data rate in the 40Gb/s range, and including up to 11% FEC overhead.
2. 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.
3. The interface is intended to handle any format of data, which has been randomized, e.g. by a scrambler.
4. Capable of driving at least 8" of FR4 interconnect with one connector
5. Interface should be capable of operating indefinitely, without losing sync.
6. Support both in-band and out-of-band Forward Error Correction (FEC). A flexible clocking arrangement accounts for the additional FEC overhead.
7. The interface is independent of the type of optics – serial, CWDM or parallel, SMF or MMF.
8. 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.
9. Support of simple and robust clock and data recovery of the signals comprising the interface bus.
10. The interface includes independent status monitoring and supports received loss of signal detection.
11. Receive timing reference operates continuously, independent of link status.
12. The receive side of the interface will tolerate being driven while powered down without damage.
13. The interface includes data path link verification for component qualification and system integration test purposes.

5.1 SFI-5 Phase 2 System Reference Model

The following is a general synopsis of the SFI Level 5 Phase 2 interface. For reference, a general block diagram is shown in Figure 1. SFI Level 5 Phase 2 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-5.2 bus in Figure 1 are independent and may operate at different frequencies.
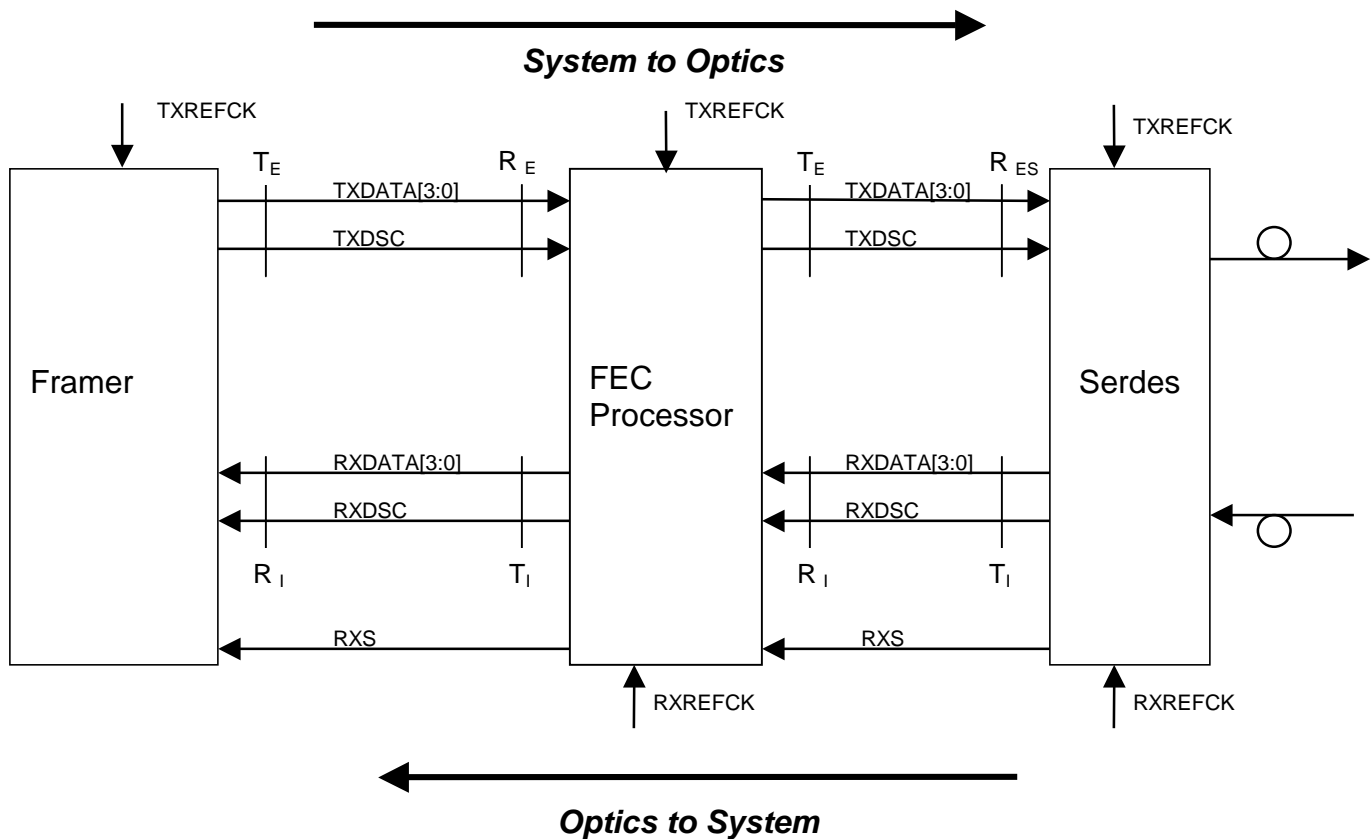
Figure 1: System Reference Model

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-5 Phase 2 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-5.2 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 SONET/SDH framer or the Serdes component directly to the SONET/SDH framer.
2.   4-bit wide data bus with each channel operating at up to 11.1 Gbps.
3.   Electrical parameters are defined in the common electrical specification OIF-CEI-02.0 [1].
4.   Maximum bus bandwidth of 44.4 Gbps is sufficient to support SONET OC-768 [2], SDH STM-256 [3], OTN OTU3 [4] and other systems operating at a payload data rate in the 40Gb/s range, with up to 11% FEC overhead.
5.   Deskew Channel included in both RX and TX directions contains out-of-band data samples to enable Deskew algorithm. Deskew Channel is a 5th data channel, running continuously at full data rate. Deskew algorithm will be operating continuously to monitor skew tracking.
6.   By using an alternating odd/even parity frame format on the deskew channel a minimum toggling rate of 1 every 10 bits is guaranteed. This can be used to recover the SFI-5 receive clock on the deskew channel and use DLLs instead of CDRs on the data channels.
7.   A selectable inversion of the data bits during the odd half of the deskew frame (5 clock cycles) shall be implemented to overcome run length issues of the SONET/SDH frame word
8.   Minimize power consumption and the number of I/O signals to simplify PC board manufacture.
9.   Maximize commonality of receive and transmit directions and of instantiations in different applications of the bus.

6   Signal definition

6.1   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 [3:0]	Optics to System	The **Receive Data** (RXDATA [3:0]) signals carry the data in the optics to system direction. Serial optical data is striped onto RXDATA [3:0] in a round-robin fashion. RXDATA [3] contains the first bit received while RXDATA [0] the last bit received. When RXDATA [3:0] is generated by a Serdes component, the 4-bit word on RXDATA [3:0] has arbitrary alignment to the quartets on the receive optical data stream. When RXDATA [3:0] is generated by an FEC processor and the 4-bit words are nibble aligned, RXDATA [3] contains the first bit of the nibble received while RXDATA [0] contains the last bit of the nibble received. Each RXDATA [#] signal is a 9.95 Gbps to 11.1 Gbps stream. It contains every 4th bit of the data stream. Optionally RXDATA[#] can be inverted during every odd half of the deskew frame for 5 data samples.

Signal Name	Direction	Function
RXDSC	Optics to System	The **Receive Deskew Channel** (RXDSC) signal provides reference data to enable skew measurements of the Receive data bus (RXDATA [3:0]). RXDSC contains 10 bit long reference frames consisting of 4 single bit samples of each signal on the Receive data bus (RXDATA [3:0]) followed by the odd parity bit of the 4 samples then 4 further single bit samples of each signal on the Receive data bus followed by the even parity bit of the 4 samples. Samples are taken from the Receive data bus in a round-robin fashion, starting with RXDATA [3] and ending with RXDATA [0]. RXDSC is a 9.95 Gbps to 11.1 Gbps stream. In case of the optional data inversion (see clause 9.2) is selected, the data samples are taken after inversion
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 with TXREFCK.

Signal Name	Direction	Function
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 SONET/SDH framer, or from the Serdes component directly to the SONET/SDH 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-5.2 Receive Signal Summary

6.2 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 [3:0]	System to Optics	The **Transmit Data** (TXDATA [3:0]) signals carry data in the system to optics direction. Data on TXDATA [3:0] are placed on the transmit optical stream in a round-robin fashion. TXDATA [3] contains the first bit transmitted while TXDATA [0] the last bit transmitted. When TXDATA [3:0] is quartet aligned, TXDATA [3] contains the first bit of the nibble, while TXDATA [0] contains the last bit of the nibble transmitted. Each TXDATA [#] signal is a 9.95 Gbps to 11.1 Gbps stream. It contains every 4th bit of the transmitted data stream. TXDATA [3:0] is frequency locked to TXREFCK with specified maximum skew between all channels. Optionally TXDATA[#] can be inverted during every odd half 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 [3:0]) between FEC processor and Framer. TXDSC is a 9.95 Gbps to 11.1 Gbps stream. TXDSC contains 10 bit long reference frames consisting of 4 single bit samples of each signal on the Receive data bus (TXDATA [3:0]) followed by the odd parity bit of the 4 samples then 4 further single bit samples of each signal on the Receive data bus followed by the even parity bit of the 4 samples. Samples are taken from the Transmit data bus in a round-robin fashion, starting with TXDATA [3] and ending with TXDATA [0]. TXDSC is a 9.95 Gbps to 11.1 Gbps stream. In case of the optional data inversion (see clause 9.2) 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. Static phase offset between TXDSC and TXDATA is unspecified. It is mandatory for one device in the Transmit chain to have an external clock source connected to TXREFCK.

Table 2: SFI-5.2 Transmit Signal Summary

7   Logical Reference Models

7.1   Serdes component Receive interface

Figure 2 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.
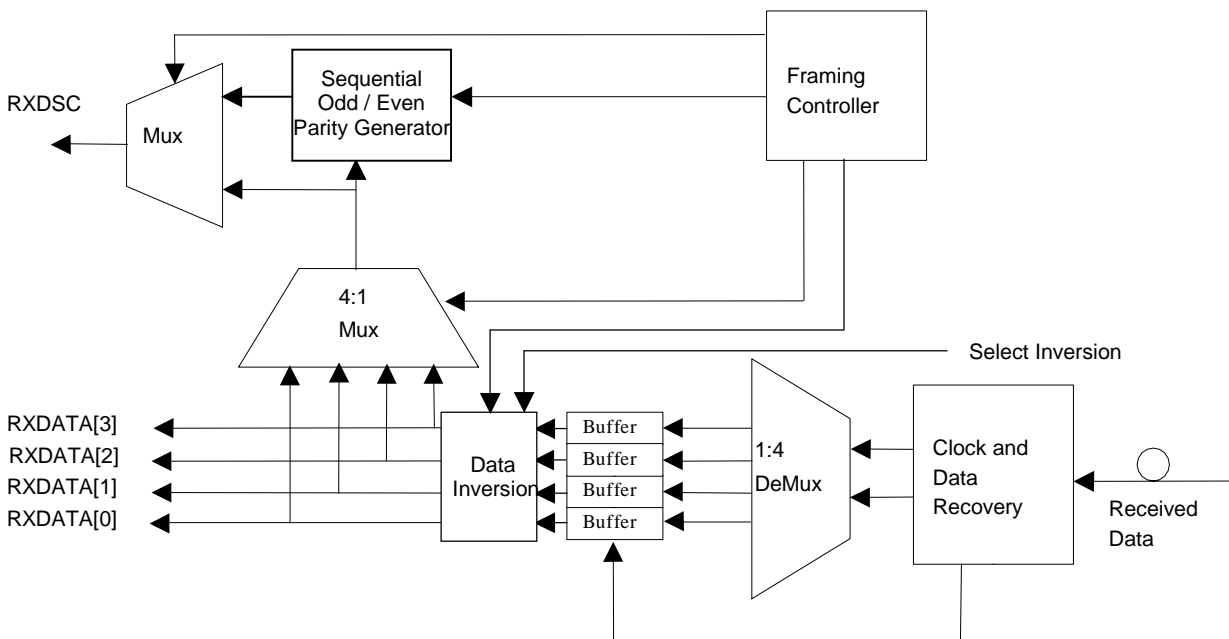
Figure 2: Model of Serdes component - Receive I/F Source

The data in the optical stream from each fiber is striped across the 4 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 [3] 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 half of the deskew frame, as described in clause 9.2

RXDSC replicates the data bits sent on each signal of the Receive Data bus RXDATA [3:0] cyclically. First, each RXDATA [#] signal is sampled, in turn, for 1 bit time. Bit sampling begins with RXDATA [3] and ends with RXDATA [0]. Then it is followed by the corresponding odd parity bit over the 4 data channel samples, generated by the parity generator block. After this each RXDATA [#]

signal is sampled again, in turn, for 1 bit time. Bit sampling begins as before with RXDATA [3] and ends with RXDATA [0]. Then it is followed by the corresponding even parity bit over the 4 data channel samples, generated by the parity generator block. After all 10 bits have been sent, a new reference frame is initiated on RXDSC and then continuously generated. The generation process is illustrated in Figure 3.

Alternating Odd/Even Parity
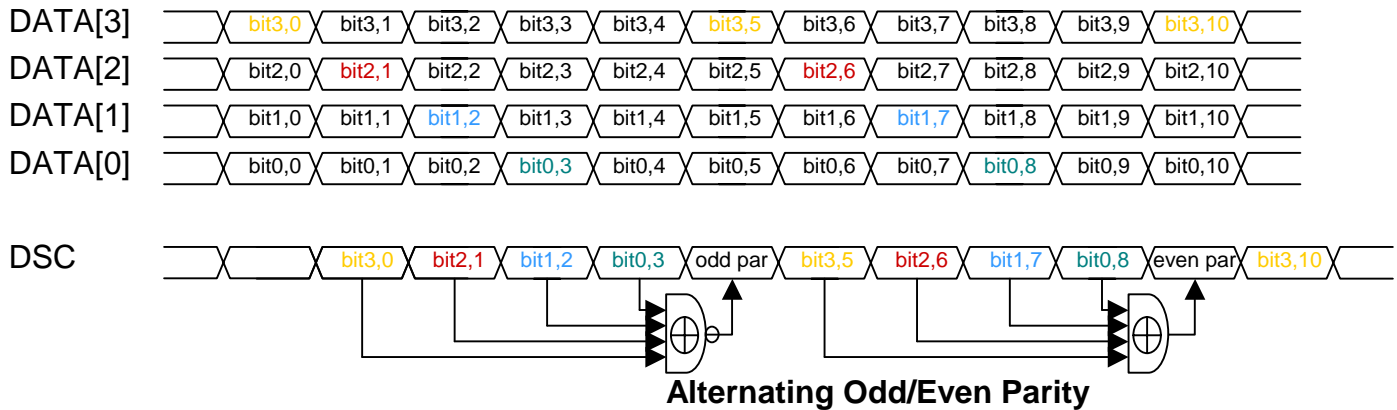
Figure 3: Timing Diagram of Deskew Channel Generation

7.2   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.
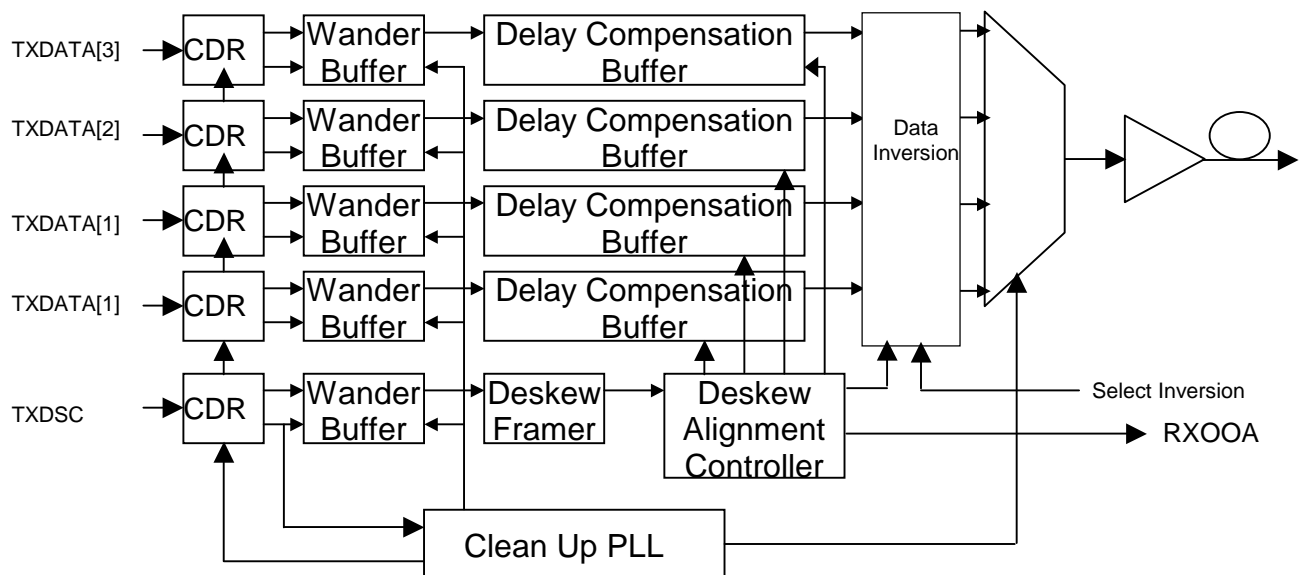
Figure 4: Model of Serdes component - Transmit I/F Sink

The Clock and Data Recovery (CDR) blocks track the center of the "eye" of the Transmit Data bus TXDATA [#] and the Transmit Deskew Channel TXDSC signal. 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 [3:0]. Each TXDATA bit is then compared with the sampled data bit in the Deskew Channel in turn according to Table 3, transmission is from top to bottom, see also Figure 3.

The Out-of-alignment alarm (RXOOA) is set, if a match has not been found on any of the 4 data channels. When a data match has been found on all 4 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 [3:0]. Any mismatch errors can be detected, which would represent errors generated over the SFI-5.2 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 half of the deskew frame, as described in clause 9.2

7.3   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.
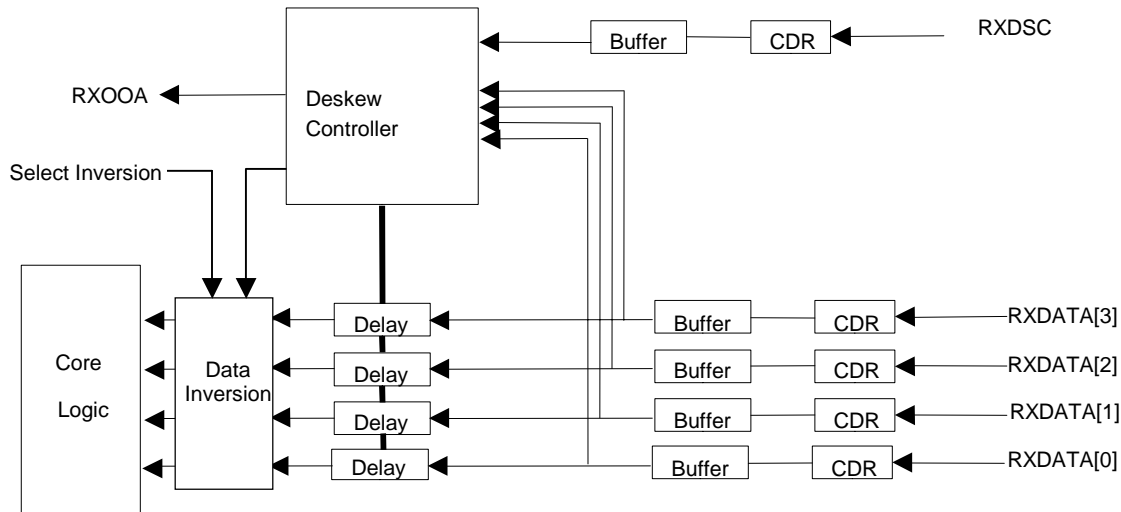
Figure 5: Model of FEC Processor / SONET Framer - Receive I/F Sink

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 [3:0]. Each RXDATA channel is then compared with the sampled data in the Deskew Channel in turn according to Table 3, transmission is from top to bottom, see also Figure 1.

The Out-of-alignment alarm (RXOOA) is set if a match has not been found on any of the 4 data channels. When a data match has been found on all 4 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 [3:0]. Any mismatch errors can be detected, which would represent errors generated over

the SFI-5.2 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 half of the deskew frame, as described in clause 9.2

7.4   Transmit interface in FEC Processor / Framer

Figure 6 below shows a model of the Transmit interface in the FEC processor or in the SONET/SDH 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.
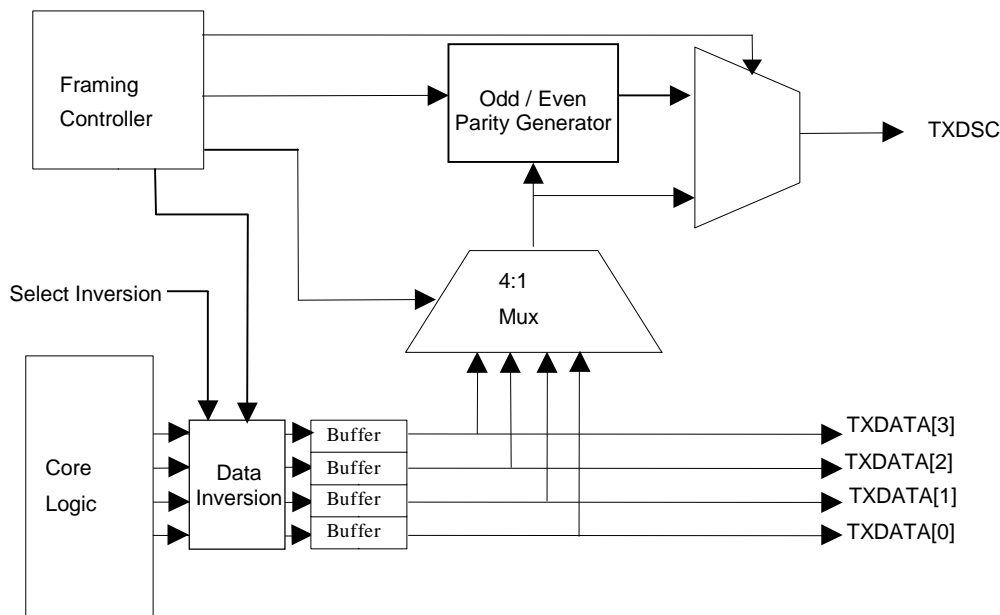
Figure 6: Model of FEC Processor / SONET Framer - Transmit I/F Source

Data from the core logic of the FEC Processor or Framer is striped across the 4 bit lanes of the first Transmit Data bus TXDATA [3:0] channel in a round-robin fashion. The first bit received is written into the re-timing buffer associated with TXDATA [3] 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 half of the deskew frame, as described in clause 9.2

TXDSC replicates the data bits sent on each signal of the Transmit Data bus TXDATA [3:0] cyclically. First, each TXDATA [#] signal is sampled, in turn, for 1

bit time. Bit sampling begins with TXDATA [3] and ends with TXDATA [0]. Then it is followed by the corresponding odd parity bit over the 4 data channel samples, generated by the parity generator block. After this each TXDATA [#] signal is sampled again, in turn, for 1 bit time. Bit sampling begins as before with TXDATA [3] and ends with TXDATA [0]. Then it is followed by the corresponding even parity bit over the 4 data channel samples, generated by the parity generator block. After all 10 bits have been sent, a new reference frame is initiated on TXDSC and then continuously generated. The generation process is illustrated in Figure 3.

8   SFI-5.2 Bit-lane Deskew

In operation the data signals may encounter different delays in transit from the SFI-5.2 source device to the sink device. The maximum differential skew introduced by the connection system is specified in clause 8.2. The earliest arriving signal may lead the latest arriving by n bits. Relative to the earliest, each of the remaining signals is coincident, or is up to n unit intervals late. The search space for determining the relative delays of 5 signals on SFI 5.2 is (n+1) * 5 combinations. To make the problem tractable, a reference signal is used in SFI-5.2, 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-5.2 source and sink devices at either end of the Receive and Transmit interfaces. In the source device, data is sampled from each of the 4 data channels sequentially, and copied onto the Deskew Channel. The Deskew Channel is then sent with the 4 data channels to the sink device over the SFI-5.2 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-5.2 interface, and continuously track skew.

Table 3 shows a reference frame on the Receive Deskew Channel (RXDSC) or the Transmit Deskew Channel (TXDSC). RX/TXDSC replicates the data bits sent on each signal of the Transmit Data bus RX/TXDATA [3:0] cyclically. First, each RX/TXDATA [#] signal is sampled, in turn, for 1 bit time. Bit sampling begins with

RX/TXDATA [3] and ends with RX/TXDATA [0]. Then it is followed by the corresponding odd parity bit over the 4 data channel samples, generated by the parity generator block. After this each RX/TXDATA [#] signal is sampled again, in turn, for 1 bit time. Bit sampling begins as before with RX/TXDATA [3] and ends with RX/TXDATA [0]. Then it is followed by the corresponding even parity bit over the 4 data channel samples, generated by the parity generator block. After all 10 bits have been sent, a new reference frame is initiated on RX/TXDSC and then continuously generated. In Table 3, transmission is from top to bottom, see also Figure 3.

Bit time	Value	Comments
1	R/TXDATA [3] Bit 0	Single bit from RXDATA [3] or TXDATA [3]
2	R/TXDATA [2] Bit 1	Single bit from RXDATA [2] or TXDATA [2]
3	R/TXDATA [1] Bit 2	Single bit from RXDATA [1] or TXDATA [1]
4	R/TXDATA [0] Bit 3	Single bit from RXDATA [0] or TXDATA [0]
5	ODD Parity	Odd parity generated over previous 4 deskew channel bits (Bit 1 – 4)
6	R/TXDATA [3] Bit 5	Single bit from RXDATA [3] or TXDATA [3]
7	R/TXDATA [2] Bit 6	Single bit from RXDATA [2] or TXDATA [2]
8	R/TXDATA [1] Bit 7	Single bit from RXDATA [1] or TXDATA [1]
9	R/TXDATA [0] Bit 8	Single bit from RXDATA [0] or TXDATA [0]
10	Even Parity	Even parity generated over previous 4 deskew channel bits (Bit 6 – 9)

Table 3: Reference Frame on RXDSC or TXDSC

8.1 Deskew of Receive interface

At the receive interface, a reference frame is generated in the source device, consisting of 4 data bit samples followed by odd parity and 4 data bit samples followed by even parity. The data bit samples are taken from the each of the Receive data bus RXDATA as defined in Table 3, transmission is from top to bottom, see also Figure 3.

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

Alternating Odd/Even Parity
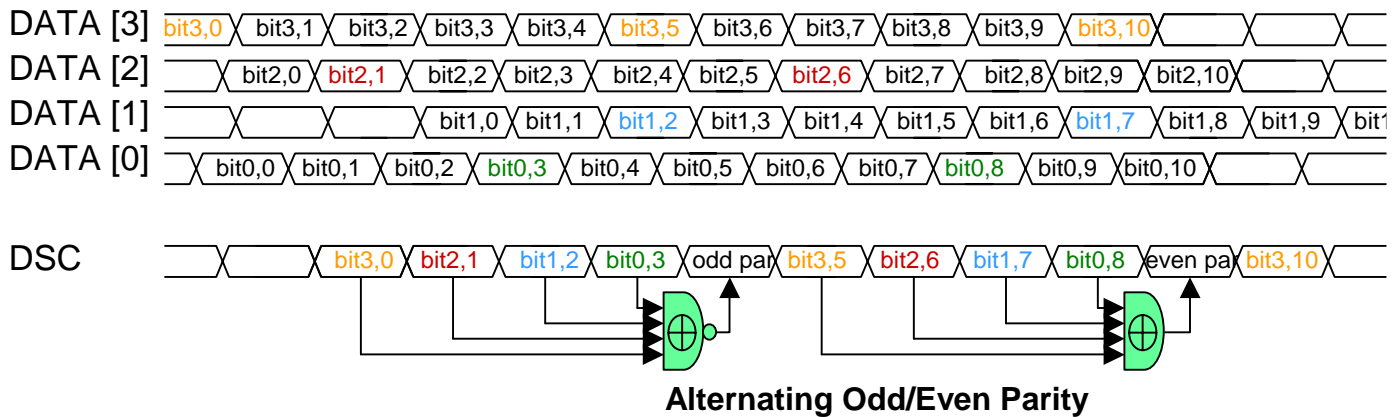
Figure 7: Receive Deskew Channel (RXDSC) Functional Timing

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 4 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.

8.2   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 5.50 UI @ 11.1 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 4.

Skew Contributor	Skew Budget @ 11.1 Gbps	Skew Budget [ps]
CEI TX Interface @ package pins	5.50 UI	500
PCB and Connector	0.6 UI	50
Serdes Input	0.6 UI	50
Total	6.70 UI	600

Table 4: 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 5.

Skew Contributor	Skew Budget @ 11.1 Gbps	Skew Budget [ps]
CEI TX Interface @ package pins	5.50 UI	500
PCB and Connector	5.50 UI	500
CEI RX Interface incl. package	5.50 UI	500
Total	16.50 UI	1500

Table 5: Relaxed Channel to Channel Skew Budget for all other interface applications

A complete Jitter/Wander/Skew budget is shown in 13 Appendix A.

9   OC-768/STM-256 Frame Alignment Word Issue

9.1   Problem description

The problem described in this clause is **not** related to the deskew method described in this document itself, but to the format of one of the target applications for this interface: the SONET/SDH frame format.
Other target applications, like OTU3 frames or PRBS test sequences don't have an issue with bit-striping over 4 lines.

Distributing OC-768/STM-256 signals bit-wise over 4 signal lines shows an issue associated with the 64 unscrambled SONET/SDH frame alignment pattern (A1=0xF6, A2=0x28). It leads to long sequences of '1' or '0' data patterns on individual signal lines.
When distributed over 4 lines, the OC-768/STM-256 frame alignment word (64 unscrambled A1, followed by 64 unscrambled A2) looks like shown in Figure 8.

```
…F6F6F6F6F6F6 | 282828282828 …

…101010101010 | 010101010101 …           o.k.            |           o.k.
…111111111111 | 000000000000 …   128 consecutive '1'  | 128 consecutive '0'
…111111111111 | 101010101010 …   128 consecutive '1'  |           o.k.
…101010101010 | 000000000000 …           o.k.            | 128 consecutive '0'
```

Figure 8: Consecutive identical digits generated by unscrambled SONET/SDH frame alignment word

So for some of the 4 signal lines the resulting frame word pattern would be more challenging for the clock data recovery (CDR) at the receiver, than the usually specified 72 CID for SONET/SDH. Also permutations of the frame alignment pattern, generated by a bit shift, result again in 128 CIDs on different data lines. The actual run length is also dependent on the values used for the bytes before the first A1 and after the last A2 byte. Simulations with all 0, all 1 or 0xF6/0x28 showed a maximum of 131 CIDs. Unfortunately the standards leave the usage for the bytes before first A1 and after last A2 open for future standardization. This indicates that 131 CIDs may even be exceeded by unfortunate usage of these bytes.
One work-around for this problem is provided by the proposed deskew channel format. As it has a guaranteed toggle rate of 1 per 10 cycles, it can be used to recover the clock on the deskew channel and only adjust the phase for the data lines.
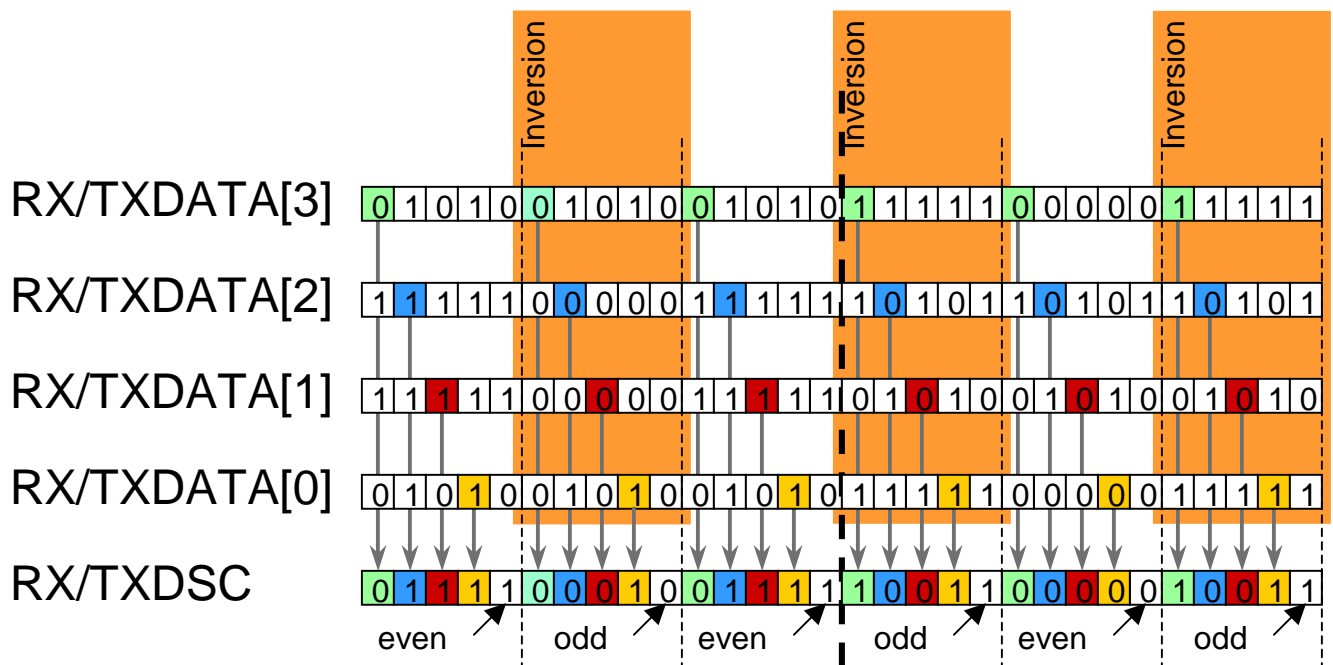
9.2 Inversion of 5 nibbles (odd half of deskew frame) to resolve the CID issue

To resolve the CID issue, a simple to implement **inversion** of 5 data nibbles per deskew frame shall be used when bit-striping an OC-768/STM-256 signal over the 4 data lines. The advantage of inverting half of the deskew frame is that, in addition to resolving the CID issue, this improves the dc-balance of the frame word part of the OC-768/STM-256 signals.

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-5.2 link.

This process shall look like shown in Figure 9.

Figure 9: Bit-striped A1/A2 with data inversion at odd half of deskew frame

This data bit inversion shall be a selectable feature and shall be used for OC-768/STM-256 applications. 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-3 or PRBS, which don't have a CID problem, when bit-striped over 4 data lines.

10   Pinout Order for PCB Routing Optimization of SFI-5 Phase 2 Links

To avoid the need to cross over the 5 critical differential 11.1 Gb/s interface traces on the PCB, a pinout order, as described below, shall be maintained by the SFI-5.2 devices TXDATA_SRC/TXDATA_SNK [3:0] and TXDSC_SRC/TXDSC_SNK pins and also at the according transponder connector pinout.
The same pinout order shall also be maintained in RX direction for signals RXDATA_SRC/RXDATA_SNK[3:0] and RXDSC_SRC/RX_DSC_SNK.
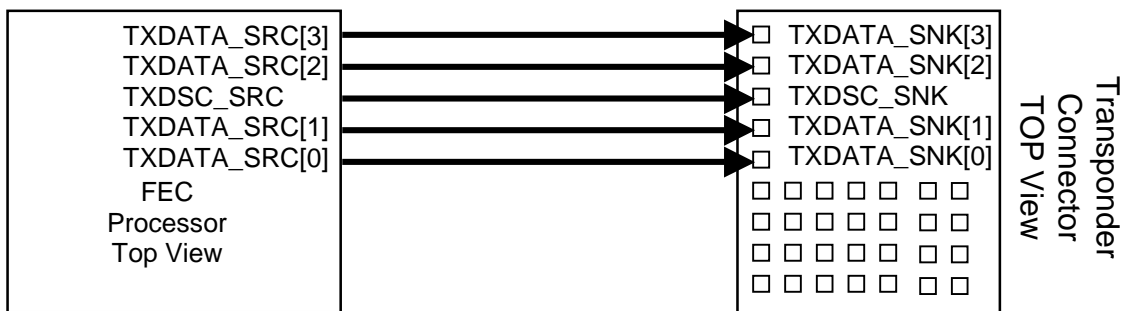
Figure 10: Pinout Order between SFI-5.2 Device and Transponder Connector
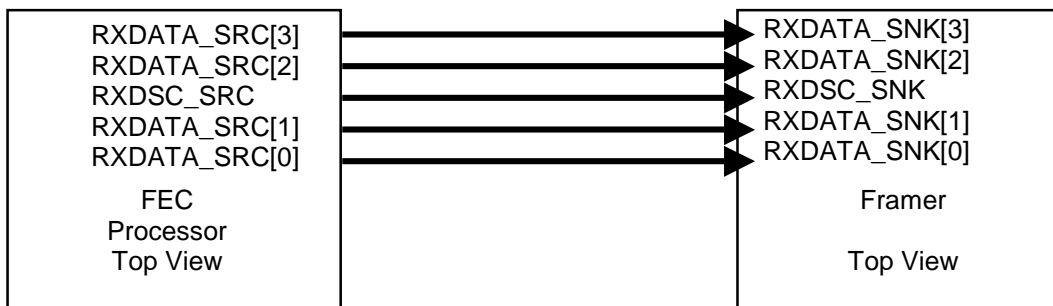
Figure 11: Pinout Order between SFI-5.2 Devices

11   Interface test requirements

It is a requirement of the SFI-5.2 interface that it will operate 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 4 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 4 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-5.2 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-5.2 implementation agreement, but should be included as part of the management function.

12   References

12.1   Normative references

[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   Informative references

[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.

13   Appendix A: Complete Jitter/Wander/Skew Budget
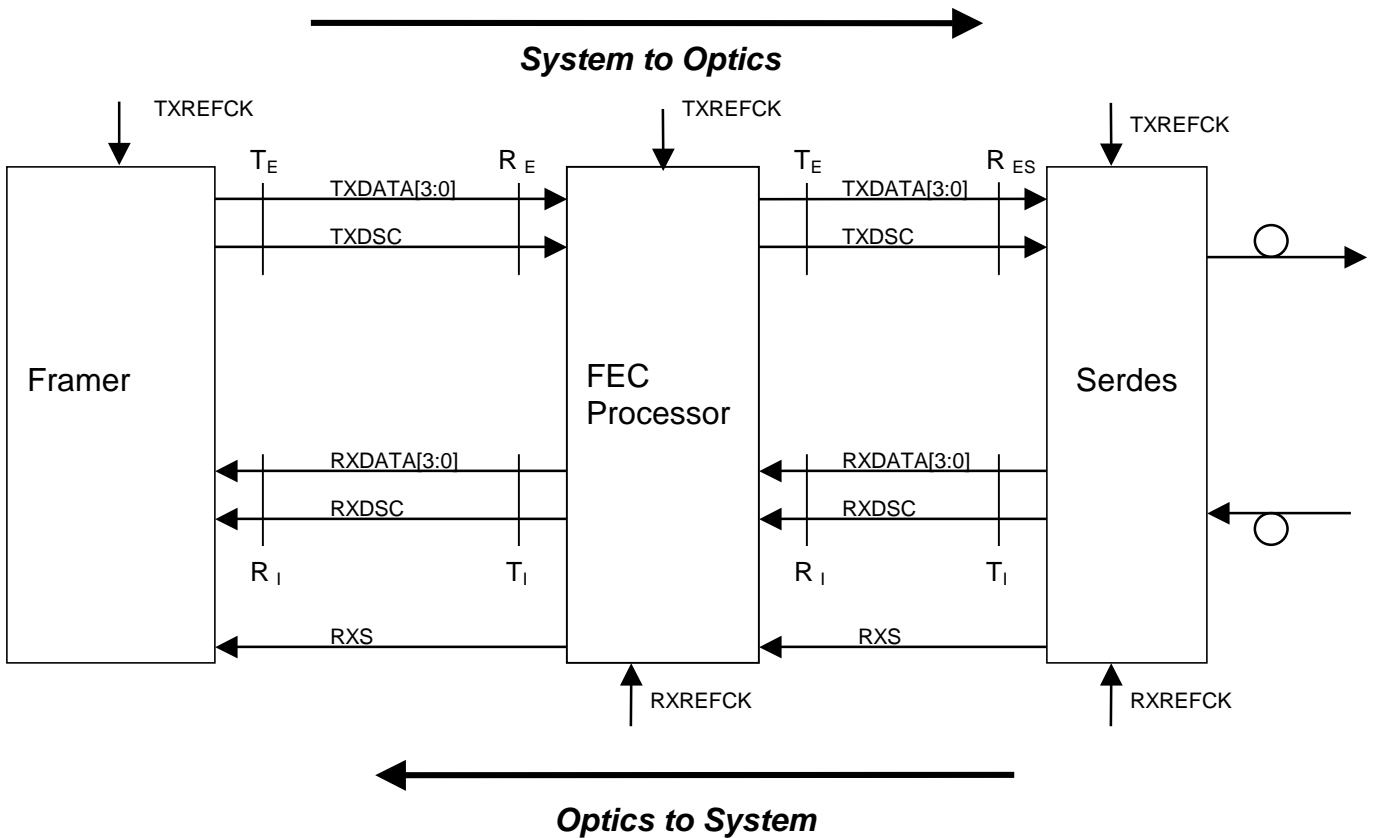
Figure 12: System Reference Model

Parameter	Signal Type	System Points			Units
		Ti/Te	Ri/Re	Res	
Skew	Data	5.50	11.00	6.10	UI peak
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	5.75	5.75	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 6: Complete Jitter/Wander/Skew Budget

14 Appendix B: 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

15 Appendix C: Open Issues / current work items

16   Appendix D: List of companies belonging to OIF when document is approved

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