XAUI/XGXS Proposal

By:

Brad Booth (Intel)

Steve Dreyer (nSerial)

Fred Weniger (Vitesse)

Koichiro Seto (Hitachi Cable)

Shimon Muller (Sun)

Shawn Rogers (Texas Instruments)

Vipul Bhatt (Finisar)

Kevin Daines (World Wide Packets)

Richard Dugan (Agilent)

Howard Frazier (Cisco)

Van Lewing (QED)

Rich Taborek (nSerial)

Steve Swanson (Corning)

Joel Dedrick (AANetcom)

Jay Hoge (JDS Uniphase)

Jason Yorks (Cielo)

Kreg Martin (Brocade Communications)

Robert Dahlgren (SV Photonics)

Mark Ritter (IBM)

Jeff Stai (Qlogic)

Nariman Yousefi (Broadcom)

Schelto van Doorn (Infineon)

Henning Lysdal (Giga)

Jim Tavacoli (Accelerant Networks)

Steve Haddock (Extreme Networks)

IEEE P802.3ae Meeting, Albuquerque, March 6, 2000

Presentation Purpose

- Consolidation of Hari proposals
- Alignment with proposed Layered Model and Document Structure
- No new material is introduced
- It's homeless Hari with a new name and home

Description

- XAUI = 10 Gigabit eXtended Attachment Unit Interface
- XGXS = XGMII eXtender Sublayer
- Based on previous Hari proposals
- CDR-based, 4 lane serial, self-timed interface
- 3.125 Gbaud, 8B/10B encoded over 20” FR-4 PCB traces
- PHY and Protocol independent scalable architecture
- Convenient implementation partition
- May be implemented in CMOS, BiCMOS, SiGe
- Direct mapping of XGMII data to/from PCS
  - XGMII proposed by Howard Frazier, Cisco, et. al.

    http://grouper.ieee.org/groups/802/3/10G_study/public/july99/frazier_1_0799.pdf

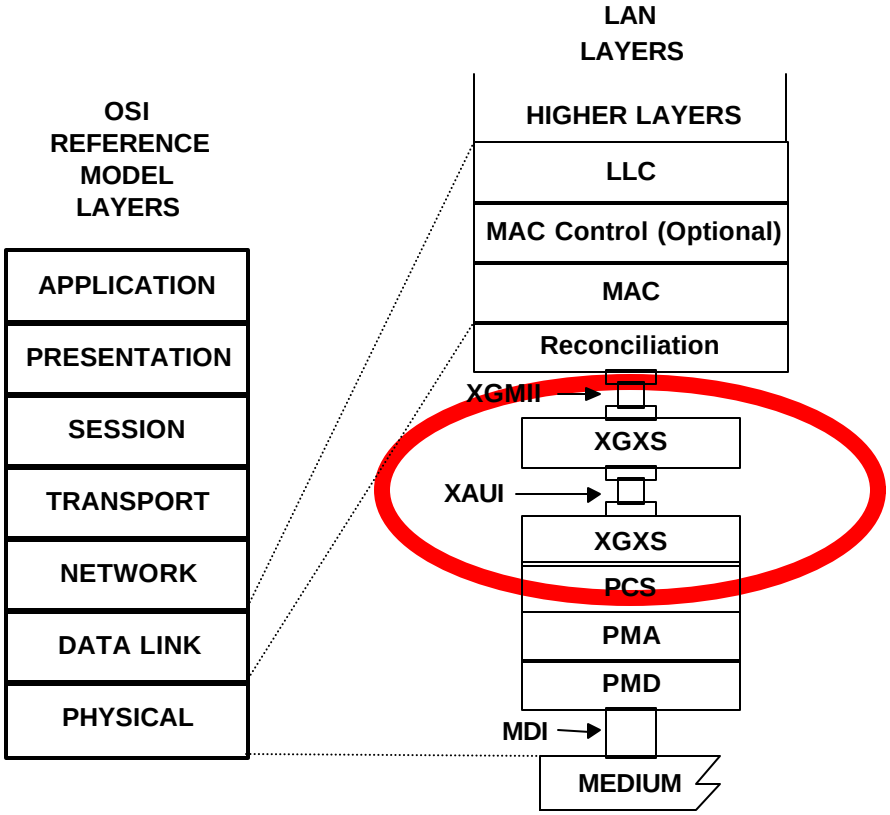
Applications

- Increased XGMII reach
- Low pin count interface = implementation flexibility
- Ease link design with multiple jitter domains
- Lower power consumption re: XGMII
- Common transceiver module interface, enables SFF
- PCS/PMA agent for MultiChannel PHYs
  - Avoids excessive penalties for all other PHYs
- Self-timed interface eliminates high-speed interface clocks

Highlights

- Increased reach
  - XGMII is ~3” (~7 cm)
  - XAUI is ~20” (~50 cm)
- Lower connection count
  - XGMII is 74 wires (2 sets of 32 data, 4 control & 1 clock)
  - XAUI is 16 wires (2 sets of 4 differential pairs)
- Better jitter control
  - XGMII does not attenuate jitter (neither does OIF99.102)
  - XAUI self-timed interface enables excellent jitter control at PCS

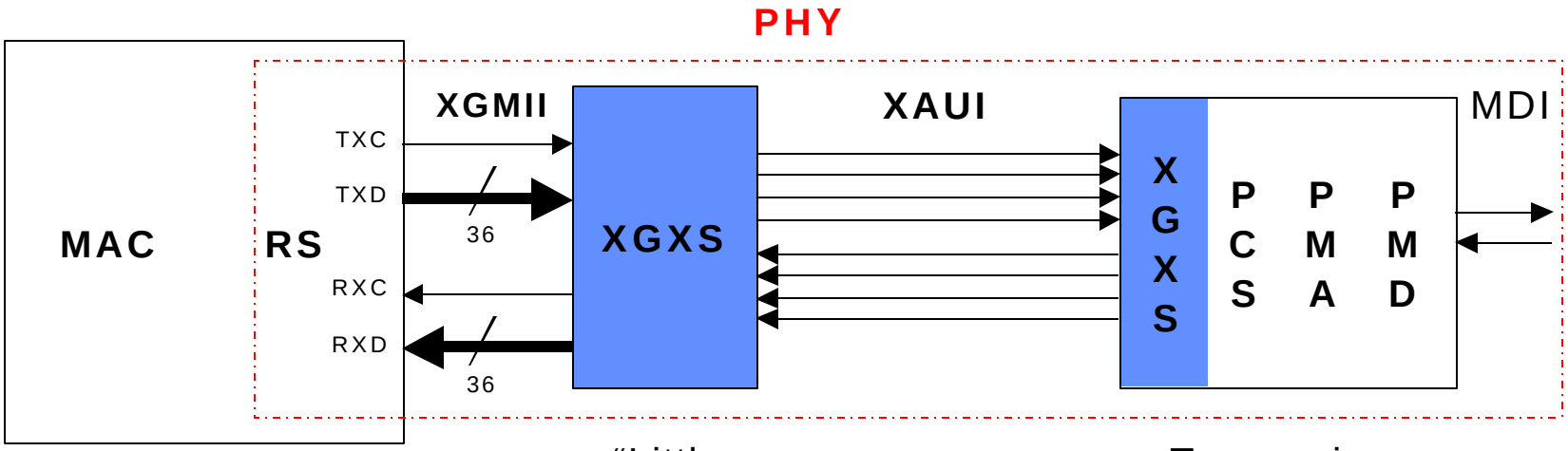
Location - Layer Model



MDI = Medium Dependent Interface
XGMII = 10 Gigabit Media Independent Interface
XAUI = 10 Gigabit Attachment Unit Interface
PCS = Physical Coding Sublayer

XGXS = XGMII Extender Sublayer
PMA = Physical Medium Attachment
PHY = Physical Layer Device
PMD = Physical Medium Dependent

Implementation Example



Layer Model Proposal

- Define the XGMII as the standardized instantiation of the PCS Service Interface and Reconciliation Sublayer interface
- Define the XGXS and the XAUI as an optional extension to the XGMII
  - Define the XAUI as the standardized instantiation of the XGXS Service Interface
  - Interface between PCS and XGXS has no standardized instantiation - implementation specific
- Define the MultiChannel PHYs (e.g. WWDM, Parallel Optics) to optionally use the XGXS CODEC/SERDES as its PCS/PMA

XGXS Functions

- Use 8B/10B transmission code
- Perform column striping across 4 independent serial lanes
  - Identified as lane 0, lane 1, lane 2, lane 3
- Perform XAUI lane and interface (link) synchronization
- Idle pattern adequate for link initialization
- Perform lane-to-lane deskew
- Perform clock tolerance compensation
- Provide robust packet delimiters
- Perform error control to prevent error propagation

Basic Code Groups

- Similar to GbE
  - No even/odd alignment, new Skip and Align

/A/ K28.3 (Align) - Lane deskew via code-group alignment

/K/ K28.5 (Sync) - Synchronization, EOP Padding

/R/ K28.0 (Skip) - Clock tolerance compensation

/S/ K27.7 (Start) - Start-of-Packet (SOP), Lane 0 ID

/T/ K29.7 (Terminate) - End-of-Packet (EOP)

/E/ K30.7 (Error) - Signaled upon detection of error

/d/ Dxx.y (data) - Packet data

“Extra” Code Groups

- The following are included in related proposals:

/Kb/ K28.1 (Busy Sync) - Synchronization/Rate control

/Rb/ K23.7 (Busy Skip) - Clock tolerance comp/Rate control

/O/ K28.2 (FCOS) - Fibre Channel Ordered Set

- The following remaining 8B/10B special code-groups are are not used:

K28.4, K28.6, K28.7*

* (Be careful not to follow /K28.7/ with /K28.x/, /D3.x/, /D11.x/, /D12.x/, /D19.x/, /D20.x/, or /D28.x/, as a comma is generated across the boundaries of the two adjacent code-groups and may result in false code-group alignment)

Data Mapping

- XGMII data is directly mapped through XGXS to PCS
  - XGMII data is 32-bits wide with one control bit “K” per byte
  - XGMII<0:31> maps to parallel bytes/lanes <0:3>; XGMIIn, n=0:3
    - XGMII0<7:0,K> maps to parallel lane 0, XGMII1<7:0,K> maps to…
    - XGMII<7> is MSB, XGMII<0> is LSB,
  - XGMII<K0:K3> maps to parallel lanes 0:3
  - XGMII lanes 0:3 map directly to XGXS lanes <0:3>
  - XGXS lane<n> = XGXSn<7:0,K>
  - Encode input XGXSn<7:0,K> = XGXSn<H,G,F,E,D,C,B,A,K>
  - Encode output XGXSn<7:0,K> = XGXSn<a,b,c,d,e,i,f,g,h,j> = XAUIn<0:9>
  - XAUIn<0:9> serially transmitted/received on XAUI. Bit 0 is first (Tx/Rx)
- See backup slides for an illustration

Data Mapping Example

XGMII

D<7:0,K0>	I	I	S	d_p	d	d	---	d	d	d	d_f	I	I	I	I
D<15:8,K1>	I	I	d_p	d_p	d	d	---	d	d	d_f	T	I	I	I	I
D<23:16,K2>	I	I	d_p	d_p	d	d	---	d	d	d_f	I	I	I	I	I
D<31:24,K3>	I	I	d_p	d_s	d	d	---	d	d	d_f	I	I	I	I	I

XGXS Encoded Data

Lane 0	K	R	S	d_p	d	d	---	d	d	d	d_f	A	K	R	K
Lane 1	K	R	d_p	d_p	d	d	---	d	d	d_f	T	A	K	R	K
Lane 2	K	R	d_p	d_p	d	d	---	d	d	d_f	K	A	K	R	K
Lane 3	K	R	d_p	d_s	d	d	---	d	d	d_f	K	A	K	R	K

Idle Encoding

- Idle (no data to send) is conveyed by the repeating sequence:

  AKRKRKRKRKRKRKRKAKRKR.... on each of 4 XAUI lanes

- /A/ used to deskew and align XAUI lanes at receiver
- /K/ contains a comma. The alternating sequence KRKR contains both running disparity versions of comma (comma+, comma-).
- /R/ selected for its spectral properties when combined with /K/ and sometimes /A/ during Idle. Disparity neutral enabling insertion/removal without affecting lane running disparity.
- /A/, /K/ and /R/ are all a hamming distance of 3 from each other
- “Even/Odd” code-group alignment (GbE) is superfluous and not used

Synchronization

- XAUI 4-lane link synchronization is a 5 step process
  - 1-4 acquire sync on all 4 lanes individually
  - 5 align/deskew synchronized lanes
- Loss of sync on any lane results in XAUI link loss-of-sync
- Lane sync acquisition similar to 1000BASE-X PCS
  - Employ hysteresis to preclude false sync and loss-of-sync due to bit errors
  - “Hot-sync” not an appropriate implementation technique
  - Periodic Align (/A/-column) check a good link health check
- XAUI Link Sync is fast, straightforward and reliable
- See backup slides for an illustration

Deskew

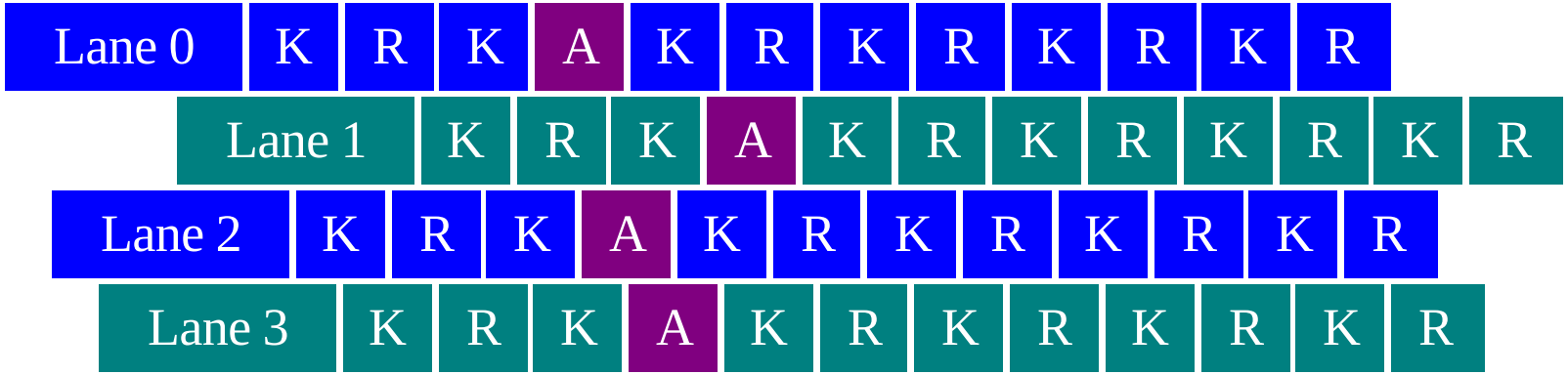
- Skew is imparted by active and passive link elements
- XGXS deskew accounts for all skew present at the Rx
- Lane deskew performed by alignment to deskew pattern present in Idle stream: Align code-groups in all lanes
- /A/ columns are issued as the first and every 16th column of the IPG

Skew Source	#	Skew	Total Skew
SerDes Tx	1	1 UI	1 UI
PCB	2	1 UI	2 UI
Medium	1	<16 UI	<16 UI
SerDes Rx	1	20 UI	20 UI
Total			<39 UI

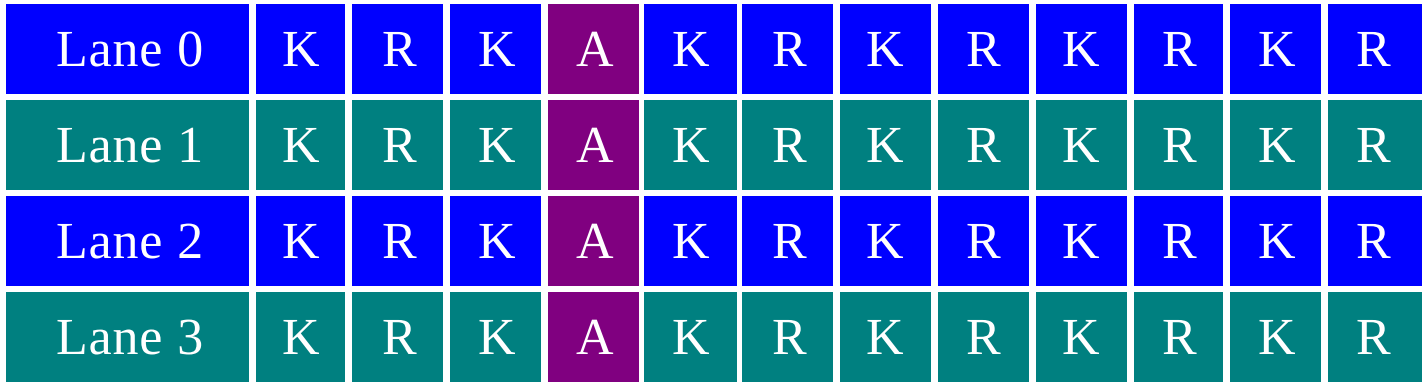
- 40 UI deskew pattern needs to be 80 bits. Idle is 160 bits

XGXS Deskew

Skewed data at receiver input. Skew ~18 bits



Deskew lanes by lining up Align code-groups



Clock Tolerance Compensation

- The XGXS provides jitter and noise isolation by retiming or attenuating jitter by the use of a high quality repeater
- Retiming is optional, but may help control jitter
- Idle pattern Skip (/R/) columns may be inserted/removed to adjust for clock tolerance differences due to retiming
  - Skip columns may be inserted anywhere in Idle stream
  - Proper disparity Skip required in each Lane
  - Any Skip column may be removed
- Clock tolerance for 1518 byte packet @ ±100 ppm is 0.76 UI/lane
  - A few bytes of elasticity buffering is sufficient to wait for many frames in case a Skip column is not available for removal

Skip Column Insert Example

Lane 0	K	R	S	d_p	d	d	---	d	d	d	d_f	A	K	R	K
Lane 1	K	R	d_p	d_p	d	d	---	d	d	d_f	T	A	K	R	K
Lane 2	K	R	d_p	d_p	d	d	---	d	d	d_f	K	A	K	R	K
Lane 3	K	R	d_p	d_s	d	d	---	d	d	d_f	K	A	K	R	K

PMD Inserts Skip column here

Lane 0	K	R	S	d_p	d	d	---	d	d	d	d_f	A	R	K	R
Lane 1	K	R	d_p	d_p	d	d	---	d	d	d_f	T	A	R	K	R
Lane 2	K	R	d_p	d_p	d	d	---	d	d	d_f	K	A	R	K	R
Lane 3	K	R	d_p	d_s	d	d	---	d	d	d_f	K	A	R	K	R

Error Control

- Packets with detected errors must be aborted
  - 8B/10B code violation detection may be propagated forward
  - IPG special code-groups stop error propagation (e.g. /A/, /K/, /R/)
- Rule: Signal Error code upon detected error or in column containing EOP if the error is detected during the IPG. Error signaling is a lane function since disparity is checked per lane.
- XGXS checks received packets for proper formation
  - Rules TBD, should be PHY/Protocol independent

Electrical

- Electrical interface is based on low swing AC coupled differential interface
- AC coupling is required at receiver inputs
- Link compliance point is at the receiver
- Transmitter may use equalization as long as receiver specifications are not exceeded

XAUI Rx/Tx & Interconnect

Transmitter Parameter	Value
Vo Dif(max)	800 mv
Vo Dif(min)	500 mv
Voh	AC
Vol	AC
Input nominal	6.5 ma
Differential Skew(max)	15 ps

Receiver Parameter	Value
Vin Dif(max)	1000 mv
Vin Dif(min)	175 mv
Loss 50W	9.1 dB
Differential Skew(max)	75 ps

Interconnect Parameter	Value
Tr/Tf Min, 20%-80%	60 ps[1]
Tr/Tf Max, 20%-80%	131 ps[1]
PCB Impedance	100 ±10W
Connector Impedance	100 ±30W
Source Impedance	100 ±20W
Load Termination	100 ±20W
Return Loss	10 dB[2]

1. Optional if transmitter meets the receiver jitter and eye mask with golden PCB
2. SerDes inputs must meet the return loss from 100 MHz to 2.5 GHz (0.8 x 3.125 Gbaud)

XAUI Loss Budget

Item	Loss
Connector Loss	1 dB
Next + Fext Loss	0.75 dB
PCB Loss	7.35 dB
Loss Budget	9.1 dB

PCB Condition 1	Normal	Worst
MSTL Loss Max (dB/in)	0.32	0.43
Max Distance (in)	23”	17.1”

Normal PCB was assumed with loss tangent of 0.22, worst case it was assumed high temperature and humidity 85/ 85. Better grade of FR4 may reduce the loss by as much as 50%.

PCB Condition 2	Normal	Worst
STL Loss Max (dB/in)	0.41	0.55
Max Distance (in)	18”	13.4”

HP test measurement for 20" line showed 5.2 dB loss or 0.26dB/ in based on the eye loss, the loss assumed here is very conservative.

XAUI Jitter

Jitter Compliance Point	Tx[1]	Rx
Deterministic Jitter	0.17 UI	**0.41 UI**
Total Jitter	0.35 UI[2]	**0.65 UI**
1-sigma RJ @ max DJ for 10^{-12} BER[3]	4.11 ps	**5.49 ps**
1-sigma RJ @ max DJ for 10^{-12} BER[3]	3.92 ps	5.23 ps

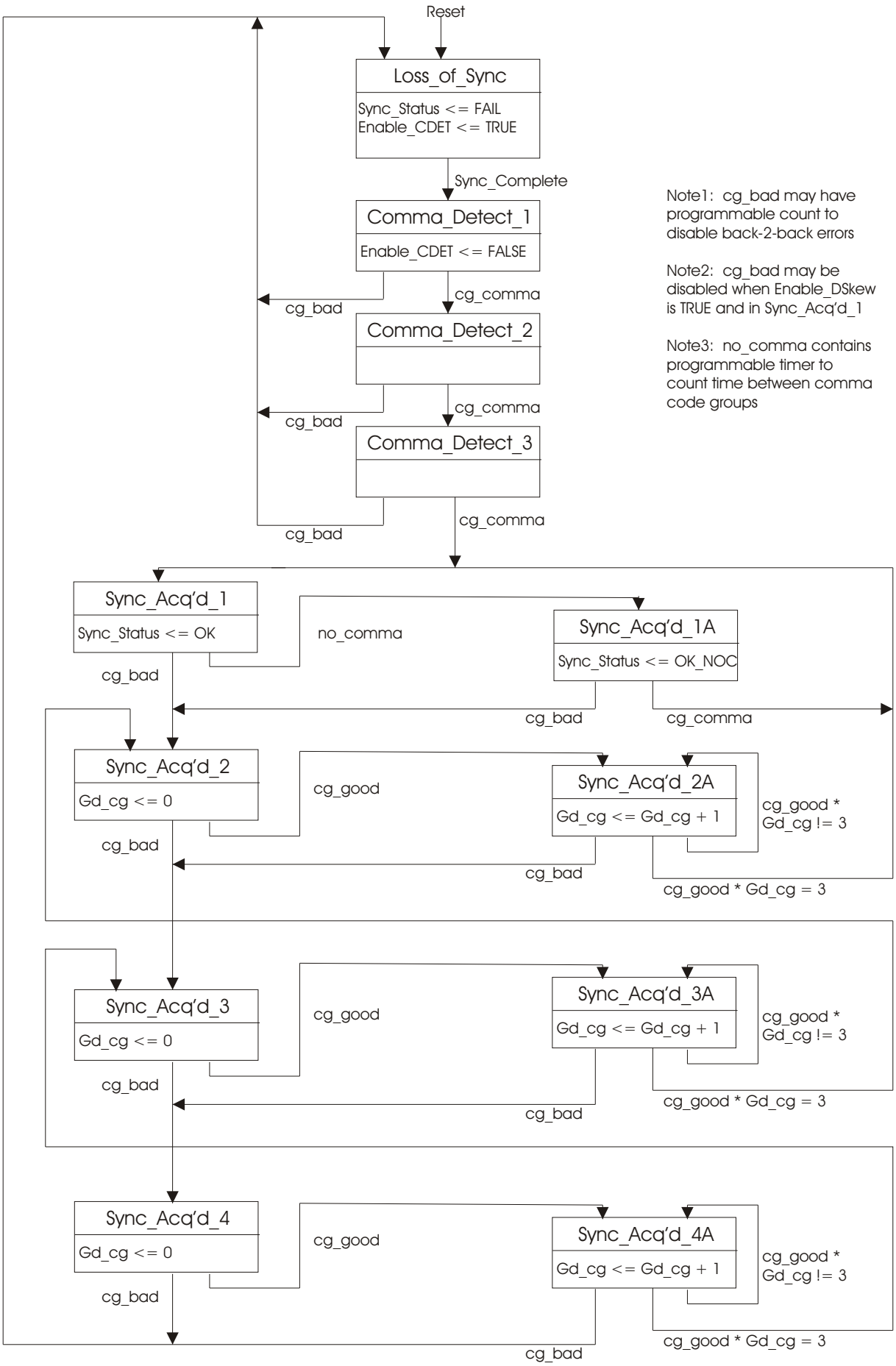
1. Tx point is for reference. Rx point is for compliance.
2. The SerDes component should have better jitter performance than specified here to allow for system noise.
3. 1-Sigma value listed here are at maximum DJ, if the DJ value is smaller then the 1-Sigma RJ may increase to the total jitter value.

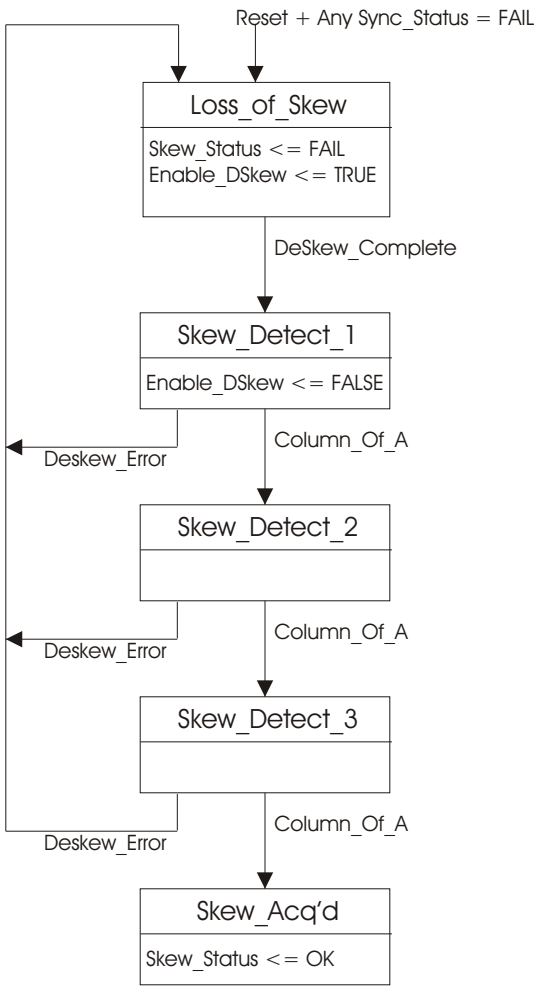
Summary

- PHY and Protocol independent scalable architecture
- XAUI/XGXS provide PHY, Protocol & Application independence
  - Common interface/rules for 10 GbE, 10 GFC, InfiniBand™ & UniPHY
  - Based on generic 10 Gbps chip-to-chip interconnect architecture
- Architecture resembles simple and familiar 1000BASE-X PHY
- Low complexity, low latency, quick synchronizing reliable interface
- Enabler for early 10 GbE PHYs
- May be integrated into MAC/RS ASIC, eliminating XGMII

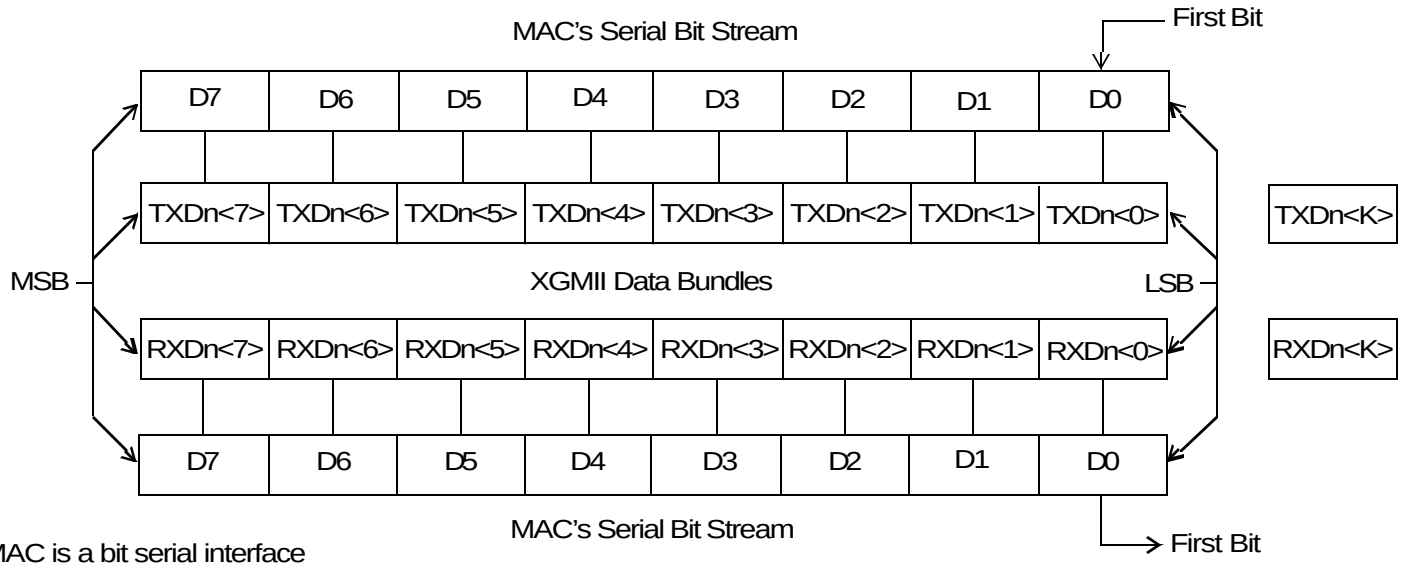
Backup Slides

- XGXS Synchronization State Diagrams
- XGXS/XAUI Reference Diagram


Reset
Loss_of_Sync
Sync_Status <= FAIL
Enable_CDET <= TRUE
Sync_Complete
Comma_Detect_1
Enable_CDET <= FALSE
cg_bad
cg_comma
Comma_Detect_2
cg_bad
cg_comma
Comma_Detect_3
cg_bad
cg_comma
Note1: cg_bad may have programmable count to disable back-2-back errors
Note2: cg_bad may be disabled when Enable_DSkew is TRUE and in Sync_Acq'd_1
Note3: no_comma contains programmable timer to count time between comma code groups
Sync_Acq'd_1
Sync_Status <= OK
no_comma
Sync_Acq'd_1A
Sync_Status <= OK_NOC
cg_bad
cg_bad
cg_comma
Sync_Acq'd_2
Gd_cg <= 0
cg_good
Sync_Acq'd_2A
Gd_cg <= Gd_cg + 1
cg_good *
Gd_cg != 3
cg_bad
cg_bad
cg_good * Gd_cg = 3
Sync_Acq'd_3
Gd_cg <= 0
cg_good
Sync_Acq'd_3A
Gd_cg <= Gd_cg + 1
cg_good *
Gd_cg != 3
cg_bad
cg_bad
cg_good * Gd_cg = 3
Sync_Acq'd_4
Gd_cg <= 0
cg_good
Sync_Acq'd_4A
Gd_cg <= Gd_cg + 1
cg_good *
Gd_cg != 3
cg_bad
cg_bad
cg_good * Gd_cg = 3


Reset + Any Sync_Status = FAIL
Loss_of_Skew
Skew_Status <= FAIL
Enable_DSkew <= TRUE
DeSkew_Complete
Skew_Detect_1
Enable_DSkew <= FALSE
Deskew_Error
Column_Of_A
Skew_Detect_2
Deskew_Error
Column_Of_A
Skew_Detect_3
Deskew_Error
Column_Of_A
Skew_Acq'd
Skew_Status <= OK

Relationship of XGMII data bundles to MAC serial bit stream



MAC is a bit serial interface
XGMII is a 32 bit data + 4 control bit interface
MAC octets represented by D7:0 map to 4 consecutive XGMII Data Bundles in rotating fashion of n=0:3

XGXS/XAUI Reference Diagram