

XAUI/XGXS Proposal

By:

Don Alderrou, nSerial; Howard Baumer, Broadcom; Vipul Bhatt, Finisar; Brad Booth, Intel; Kirk Bovil, Blazel; Ed Chang, NetWorth Technologies; Ed Cornejo, Lucent; Robert Dahlgren, SV Photonics; Kevin Daines, World Wide Packets; John Dallesasse, Molex; Joel Dedrick, AANetcom; Thomas Dineen, Dineen Consulting; Schelto van Doorn, Infineon; Steve Dreyer, nSerial; Richard Dugan, Agilent; John Ewen, IBM; Howard Frazier, Cisco; Mark Feuerstraeter, Intel; Eric Grann, Blaze; Steve Haddock, Extreme Networks; Chuck Haymes, IBM; Ken Herrity, Blaze; Jay Hoge, JDS Uniphase; Osamu Ishida, NTT; Pat Kelly, Intel; Van Lewing, QED; David Lynch, Gennum; Jeff Lynch, IBM; Henning Lysdal, Giga; Kreg Martin, Brocade Communications; Ron Miller, Brocade Communications; Shimon Muller, Sun; Bob Musk, JDS Uniphase; Brian Peters, Blaze; Mark Ritter, IBM; Shawn Rogers, Texas Instruments; Koichiro Seto, Hitachi Cable; Dave Simmons, Gennum; Jeff Stai, Qlogic; Daniel Svensson, SwitchCore; Steve Swanson, Corning; Rich Taborek, nSerial; Bharat Tailor, Gennum; Jim Tavecchi, Accelerant Networks; Hemant Thakkar, Kinar Inc.; Tom Truman, Bell Labs/Lucent; Rick Walker, Agilent; Fred Weniger, Vitesse; Tony Whitlow, Molex; Bill Wiedemann, Blaze; Jim Yokouchi, Sumitomo Electric; Jason Yorks, Cielo; Nariman Yousefi, Broadcom;

Presentation Purpose

- Update of March '00 proposal
 - http://grouper.ieee.org/groups/802/3/ae/public/mar00/taborek_1_0300.pdf
- Inclusion of 8B/10B Idle EMI Reduction proposal
 - http://grouper.ieee.org/groups/802/3/ae/public/may00/taborek_1_0500.pdf
- Otherwise, no new material is introduced
- Proposal is ready for Prime Time!

Description

- XAUI = 10 Gigabit eXtended Attachment Unit Interface
- XGXS = XGMII eXtender Sublayer
- 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 RS/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

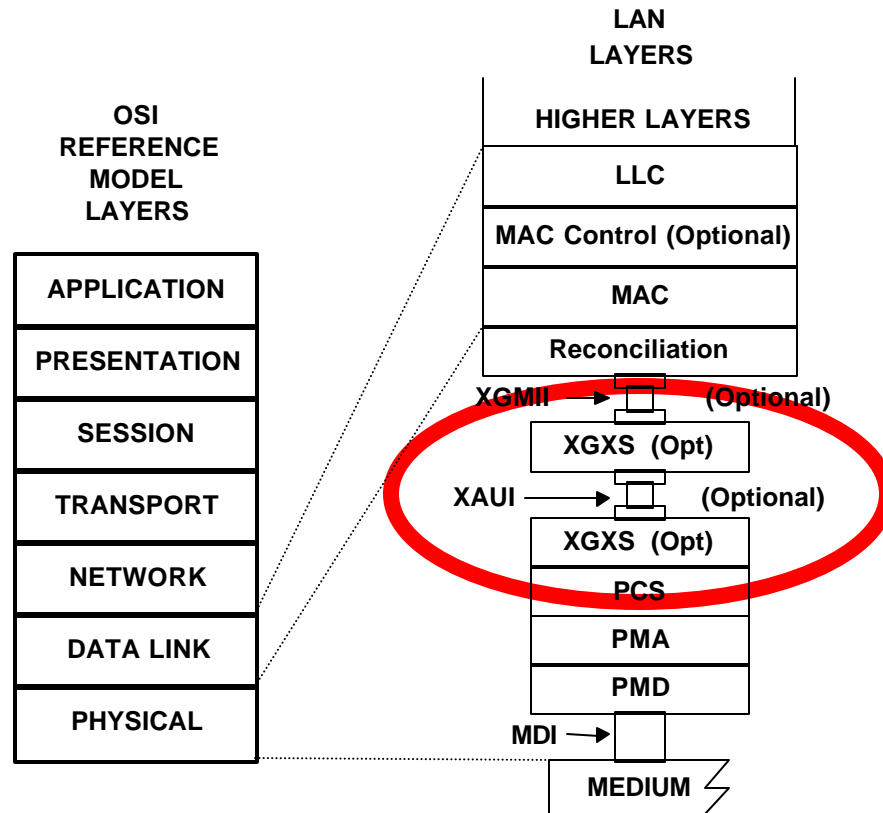
Applications

- Increased XGMII reach
- Low pin count interface = implementation flexibility
- Ease of link design with multiple jitter domains
- Lower power consumption re: XGMII
- Common transceiver module interface, enables SFF
- PCS/PMA agent for WWDM
 - 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)
- Built-in jitter control
 - Chip-to-chip interconnect degrades XGMII source-synchronous clock
 - XAUI self-timed interface enables jitter attenuation at the receiver

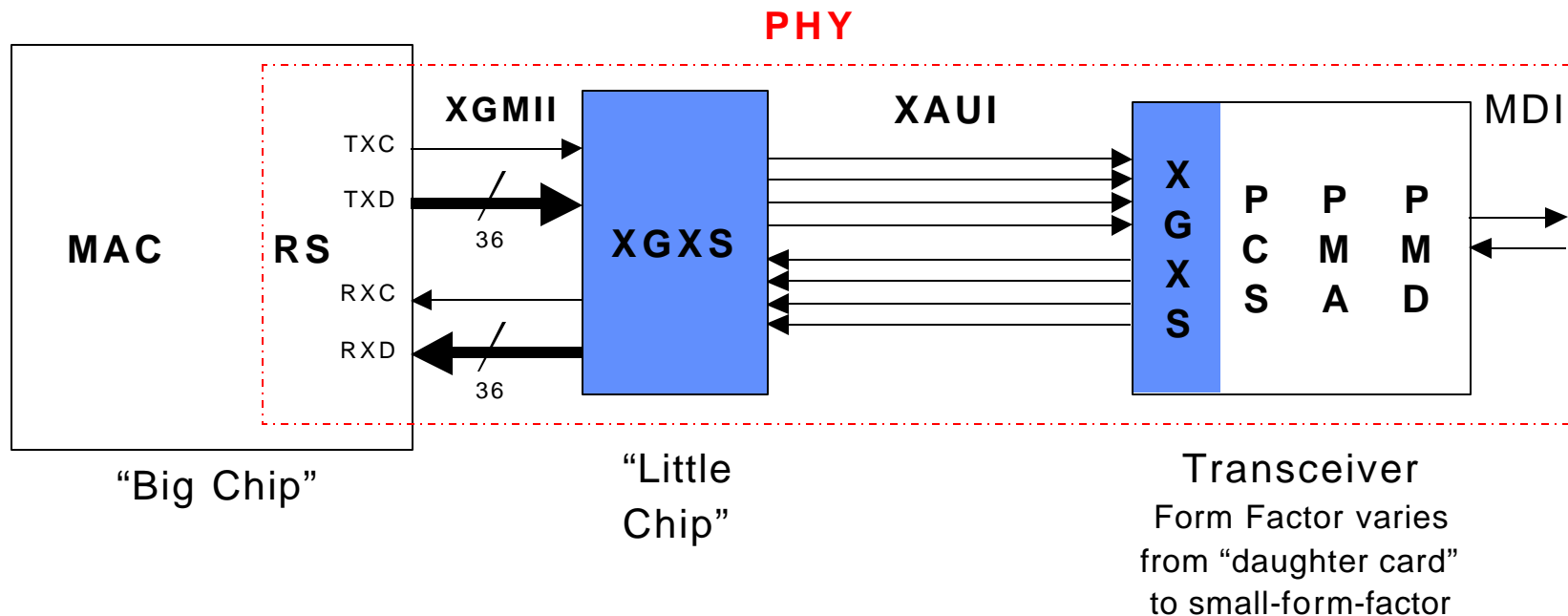
Location - Layer Model



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



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

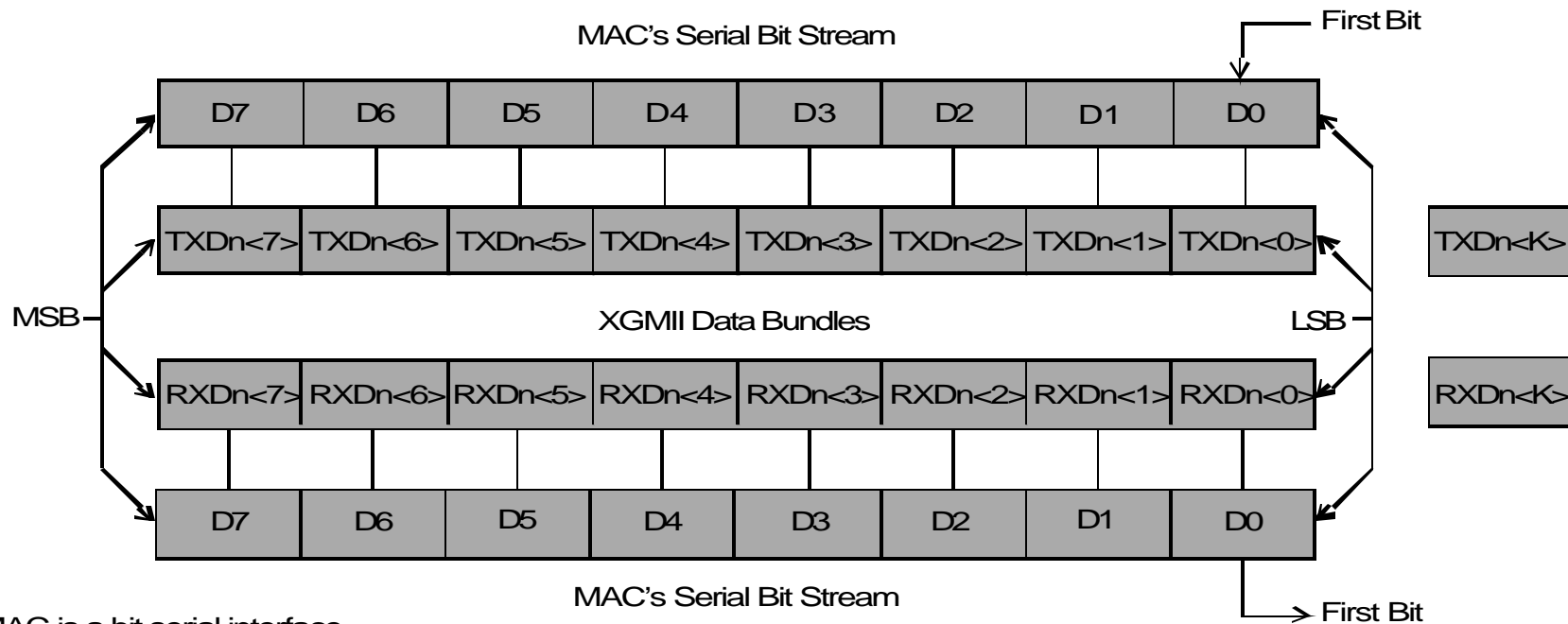
/LS/ K28.1 (Link Signaling) - LSS proposal

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

K28.2¹, K28.4, K28.6, K28.7

¹ Reserved for Fibre Channel usage in NCITS T11 10 GFC project proposals

Data Mapping: MAC to XGMII

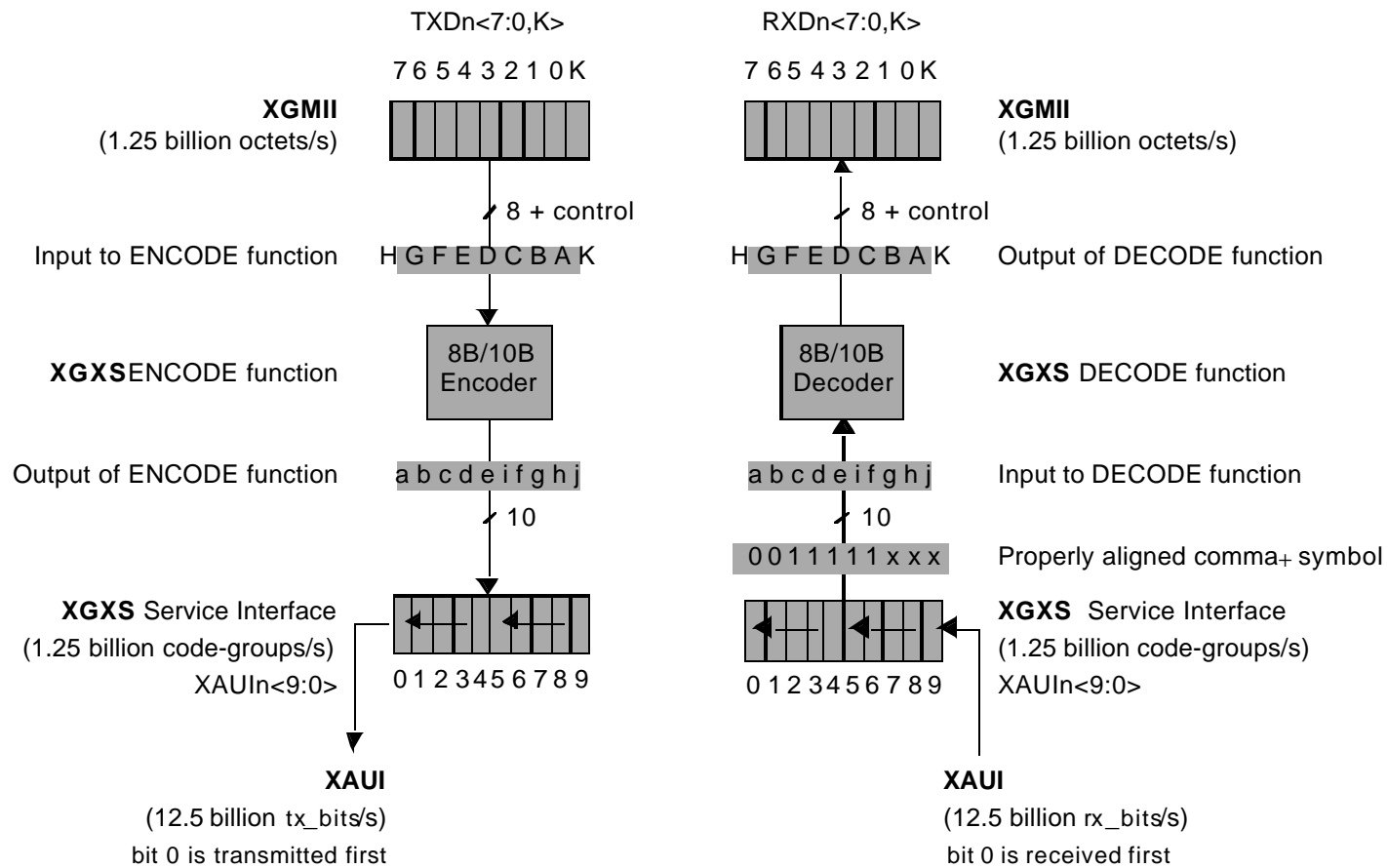


MAC is a bit serial interface

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

Data Mapping: XGMII to XAUI



Data Mapping Example

RS/XGMII Encoded Data

| | | | | | | | | | | | | | | | |
|-------------|---|---|----------------|----------------|---|---|-----|---|---|----------------|----------------|---|---|---|---|
| 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 _p | 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 _p | d | d | --- | d | d | d _f | K | A | K | R | K |

Idle Encoding

- Idle (no data to send) is conveyed by the randomized pattern /A/ K/R/:
+T-A+K-R-K+R+R+K-K+R+K-R-K+R+R+R+K (example pattern)
-K+R+R+K-R-R-K+R+K-R-R-K+R+K-K+R+A ... on each XAUI lane
- /A/ spacing is randomized: 16 min, 32 max (80-bit deskew capability)
- /K/R/s between /A/s randomly selected (no discrete spectrum)
See http://grouper.ieee.org/groups/802/3/ae/public/may00/taborek_1_0500.pdf for additional details
- /A/, /K/ and /R/ are all a hamming distance of 3 from each other
- Minimum IPG pattern is /A/K/R/ sequence, in order

Synchronization

- XAUI 4-lane link synchronization is a 2 step process
 1. Acquire sync on all 4 lanes individually;
 2. 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
 - Use hysteresis to preclude false sync and Loss-of-Sync due to bit errors
 - Re-synchronize only upon Loss-of-Sync (i.e. no “hot-sync”)
 - 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

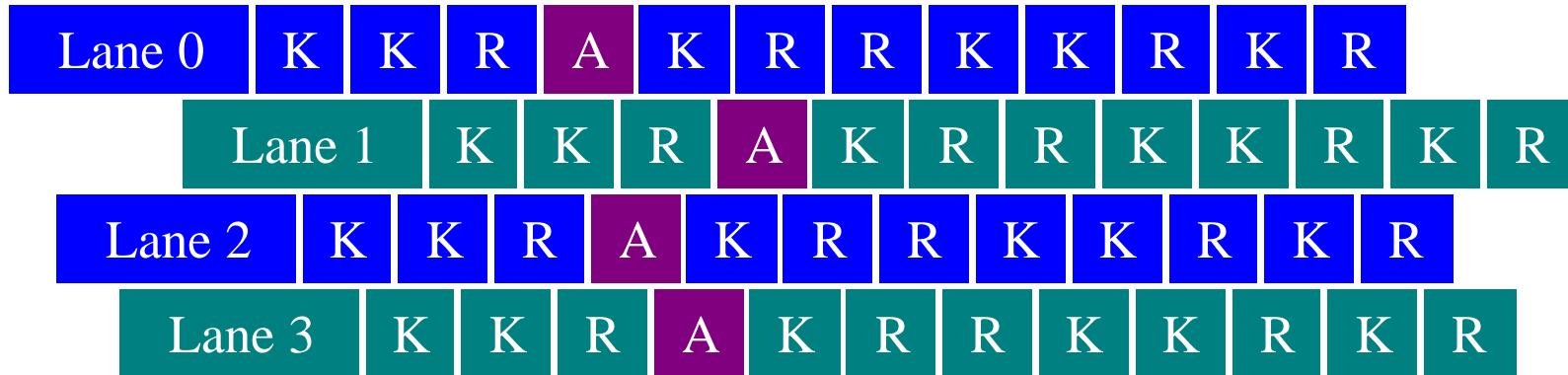
- Skew is imparted by active and passive link elements
- XGXS deskew accounts for all skew present at the Rx
- Lane deskew performed by alignment to deskew pattern present in Idle/IPG stream: Align /A/ code-groups in all lanes

| 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
- /A/ column Idle/IPG spacing is 16 columns (160 bits) minimum

XGXS Deskew

Skewed data at receiver input. Skew ~18 bits



Deskew lanes by lining up Align code-groups



Clock Tolerance Compensation

- The XGXS must restore the temporal fidelity of the signal by:
 - a. Repeating by amplifying and/or reshaping the signal w/<100% jitter transfer;
 - b. Retiming the data to a timing reference other than the received data.
- Idle pattern Skip (/R/) columns may be inserted/removed to adjust for clock tolerance differences due to retiming only
 - 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 (~13) 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 |

Skip column inserted 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 are chosen to ensure that running disparity errors are detected
- Rule: Signal Error code upon detected error or in column containing EOP if the error is detected in the column following the EOP.
 - Error is signaled per lane since disparity is checked per lane
- XGXS checks received packets for proper formation

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 |
| Iout 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 | 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 FR4 grade may reduce loss by as much as 50%.

| PCB Condition | 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 ⁻¹³ 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

- Meets HSSG Objectives and PAR 5 Criteria
- Provides PHY, Protocol & Application independence
- Based on generic 10 Gbps chip-to-chip interconnect
- Resembles simple and familiar 1000BASE-X PHY
- Low Complexity, Low Latency, Quick Synchronizing
- May be integrated into MAC/RS ASIC, eliminating XGMII

Backup Slides

- XGXS Synchronization state diagrams

