Simple Link Protocol (SLP)

A zero-overhead packet delineation for 10Gb Ethernet LAN-PHY

IEEE 802.3 Albuquerque meeting
March 6-10, 2000

Kamran Azadet, Lei-lei Song, Tom Truman, Mark Yu …… Lucent Technologies
Paul Bottorff, Norival Figueira, David Martin ……….. Nortel Networks
Fred Weniger ................................................................. Vitesse
Tom Palkert ................................................................. AMCC
Jim Yokouchi .............................................................Sumitomo
Don Larson ................................................................. TriQuint
Goals

- For packet-delineation & line coding scheme (PCS)
  - Zero overhead
  - No packet length required
  - No rate conversion required (10.000Gb/s and 10.000GBaud)
  - Serial- and WDM-PMD agnostic
  - LAN friendly - simple and low cost
  - Suitable for other non-ethernet packet-based protocols

- For chip-to-chip interconnect (XAUI)
  - EMI mitigation
  - No code conversion required
  - ... Can work with 8b/10b HARI or SUPI
Key Concepts for SLP

- Standard Ethernet frames have sufficient information to find packet/byte boundaries
  - Control characters are embedded in Inter-Packet Gap (IPG)

- Encoding: zero-overhead scrambling
  - Data payload is scrambled bit serialization of Ethernet packets
  - Control characters: EMI-reducing scrambling or DC-balanced control chars
SLP Frame Delineation

- PHY state machine tracks “context”: {HUNT, DATA, CONTROL}
  - HUNT state: unsynchronized, looking for repeated IDLE characters

- Start of Packet
  - Look for SOP byte
  - When found, enter DATA mode
  - Look for frame termination pattern

- Frame termination pattern: (T-FLAG)
  - Look for {EOP, 11 x IDLE}
  - Enter CONTROL state
  - All bytes are control characters until SOP is found

- Detection of error packet (E-FLAG):
  - Drop current packet and Enter CONTROL state
Detection of control sequences

- There are 2 critical control sequences, each at least 12-byte long
  - End of normal packet: T-FLAG
  - End of error packet: E-FLAG

- Occurrence of these flags inside payload is once every 2,000 billion years (match over 12Bytes is very unlikely)

- If the end-of-packet sequence, the T-FLAG, is found inside payload (not a lucky day!)
  - Transmitter generates E-FLAG
  - Packet is re-transmitted with a different scrambler state

- Allowing up to 3-bit mismatches for detection of the control sequences to increase the likelihood of matching patterns
What about false match?

- Pattern matching with 3 bit error tolerance means = difference between two words is less than 3 bits

- Likelihood of matching patterns with 3 bit error tolerance is greater than with no error tolerance, but still extremely small

- Probability of false match: ~ once every 15 million years (see supporting calculations in full presentation)
Calculating Probability of False Match

- Chance that a 96-bit segment $A$ matches desired pattern $B$ to within 2 bits

$$\Pr\left[ X = \sum_{i=0}^{96} A_i \oplus B_i \leq 3 \right] = \frac{1}{2^{96}} \left( \binom{96}{0} + \binom{96}{1} + \binom{96}{2} + \binom{96}{3} \right)$$

$$= 1 \times 10^{-24}$$

- Number of 96-bit patterns per year (@10Gbit/s) =

$$10^{10} \text{ bit/s} \times 1/8 \text{ bits} \times (3600 \times 24 \times 365) \text{ s/year} ; \ 4 \times 10^{16} / \text{year}$$

- Thus, frequency of false match is once every 15 million years
Scrambling Scheme

- Similar to scrambling scheme used in SONET (proposed by Bottorff, Martin & Figueira)
  - $x^{43} + 1$ is self-synchronous
  - $x^7 + x^6 + 1$ is periodically synchronized
- Dual scrambler
  - Inner scrambler prevents malicious attack
  - Outer scrambler defines spectral characteristics
- Scrambling improves EMI characteristics
IDLE Pattern Spectral Measurements at “lane” Rate

- Green: PRBS $2^7-1$ @ 2.5gbit/sec
- Red: 8b/10b @ 3.125gbit/sec

10-20 dB improvement for scrambling relative to 8b/10b
IDLE Pattern Spectral Measurements at Line Rate

- Green: PRBS $2^7-1$ @ 10.0 Gb/s
- Red: 8b/10b @ 12.5 Gb/s
Control characters

- S: SOP
- T: EOP
- E: Error
- I: Idle
- BI: Busy Idle

- 4 bit Hamming distance => 1 bit error correction, or 3bit error detection
- Can have up to 16 control codes with 4b Hamming distance – plenty of spares
The following extended Hamming (8,4) code can be used for representing up to 16 control characters:

\[(0 0 0 1, 0 1 1 1), (0 0 1 0, 1 1 0 1), (0 0 1 1, 1 0 1 0),
(0 1 0 0, 1 1 1 0), (0 1 0 1, 1 0 0 1), (0 1 1 0, 0 0 1 1),
(0 1 1 1, 0 1 0 0), (1 0 0 0, 1 0 1 1), (1 0 0 1, 1 1 0 0),
(1 0 1 0, 0 1 1 0), (1 0 1 1, 0 0 0 1), (1 1 0 0, 0 1 0 1),
(1 1 0 1, 0 0 1 0), (1 1 1 0, 1 0 0 0),
(0 0 0 0, 0 0 0 0), (1 1 1 1, 1 1 1 1).\]

Property: -- minimum Hamming distance of 4

The first 14 code words are DC-balanced, hence can be used directly for systems without scrambler.
Loss of Synchronization

- Loss of synchronization occurs when:
  - PHY control state machine in DATA mode and no IPG after timeout period
  - Timeout period can be chosen as maximum frame length (1518+8 bytes) for Ethernet
  - For non-Ethernet applications, other timeout period can be used (e.g., 8 Kbytes)

- Synchronization is achieved immediately at the next IPG (provided that optical link is up)

- No handshaking required between Transmitter & Receiver

- Synchronization time < maximum packet length
Application of SLP to PCS/PMA Interface

- **16-bit Parallel (625MHz)**
  - Interface of choice for serial PMD
  - 16 bit interface is formed by grouping two bytes at XGMII
  - Phase ambiguity in serial PMD results in constant bit offset between mux & demux – easily resolved by searching for T-FLAG using sliding window to find byte-sync

- **Scrambled 4-bit Serial (2.5GBaud / lane)**
  - SLP could also be used for transmission over 4-channel WWDM
  - Results in 25% lower Baud-rate for only 2.5% penalty due to DC wander
  - Each of the 4 bytes of the encoder is directly sent serially on four lanes
  - Deskewing can be done using “ALIGN” control code periodically embedded in IPG
Summary of Simple Link Protocol

- Zero overhead (10.0000GBaud)
- No rate conversion, “gear box”, or elastic buffers
- SOP is not restricted to Lane 0: no impact on IPG
- Direct mapping of Ethernet (LAN friendly)
- Data rate does not decrease to 9.29Gb/s (as for 64b/66b)
- Preserves byte boundaries (WAN friendly)
- Suitable for both serial and WDM
- Low latency (could be used in serial Infiniband)
- Scrambling allows DC balance and is good for EMI
- Best performance, at lowest baud-rate
- Can work with or without HARI (as serial I/O)
Demonstration System

- 16-bit LVDS @ 622 MHz interface into Opto module
- < 20k gates (~3% of FPGA logic resources)
- VHDL code will be made available to SLP supporters
Back-up slides
Case of Transmit Error / Abort

- Support of cut-through packet switching
- Error can occur inside the payload:
  - when error occurs terminate packet, then:
  - insert at least 4 error symbols (desirable in some applications)
IDLE Pattern Spectral Measurements

- Data Pattern
  - 1 megabit in length using PRBS
- 8B10B Pattern
  - 8B10B Encode Data Pattern
- Scrambled Pattern
  - Scrambled Data Pattern based on proposed $X^{58}+X^{19}+1$ Sequence
IDLE Pattern Spectral Measurements

- Test setup used in spectral measurements
PCS latency

- Less than XGMII clock latency at encoder (direct mapping of XGMII)
- 3 XGMII clock cycles at the decoder to detect EOP (12Bytes)
  - can be reduced by using less than 12Bytes to detect EOP
  - ex: when using 8Bytes pattern matching, latency can be reduced to 2 cycles
  - Probability of false match with 8bytes and 1b error tolerance is $3.52 \times 10^{-18}$
  - False match over 8bytes with 1b error tolerance occurs ~ once a year
- In the conservative case (12Bytes) latency is 16 Byte times
- total PCS latency is = **12.8ns**
During loss of sync or initialization enter HUNT mode

Look for sequence of repeated IDLE characters

Output of descrambled IDLEs should give a long chain of 000….

Initially open the loop of the side-stream descrambler, and after observing N zeros, enable sync bit (then close the loop of the descrambling shift register)

100BASE-TX uses a similar scheme