EEE Modifications
P802.3bj comments 115, 116, 117
P802.3bm – describe fast-wake signaling compatible with OTN mapper

Steve Trowbridge

Alcatel-Lucent

Supporters

Pete Anslow (Ciena)

Ghani Abbas (Ericsson)

Gary Nicholl (Cisco)

Martin Carroll (Verizon)

Sam Sambasivan (AT&T)

Issues

- Compatibility of Fast Wake signaling with OTN mapper

- Unnecessary phases in Fast Wake signaling

- Rapid Alignment Marker insertion is only needed for refresh and wake phases of deep sleep operation and should not be used for "fast wake" as this prevents the PCS with EEE from being used for 40 Gb/s and 100 Gb/s optical interfaces
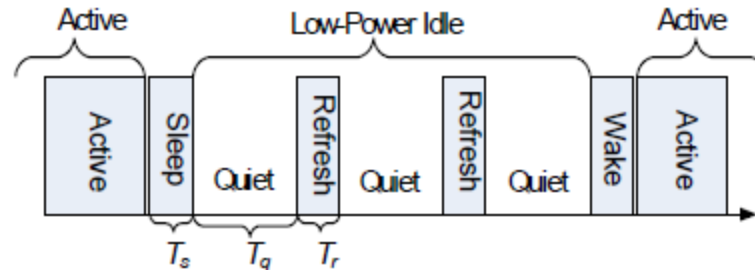
Differences in Fast Wake Signaling

Figure 78–3—Overview of EEE LPI operation

- With "Fast Wake", there is no Quiet period since the transmitter is left on. Refresh is implicit since the transmitter is not disabled. There is no "Sleep" signal needed as the transmitter will never go quiet. Ts, Tq, Tr are all irrelevant

4

Possible Alternate Figure

Figure 78-7-Overview of Fast Wake Operation

- Note that Rapid Alignment Markers are needed for the "Refresh" and "Wake" steps of the "Deep Sleep" process of Figure 78-3. They are not needed at all for the process of Figure 78-7 which can tell that the link is up without "refresh", and can switch from active to fast-wake signaling without the warning needed if the Tx were to be disabled.

OTN Mapping for 40GBASE-R and 100GBASE-R

- Transparent (PCS) and non-transparent (MAC/GFP) mappings
- What is mapped is the PCS output, allowing different PMDs to be used at the OTN ingress/egress
- Restricted to 66B codeword set of Table 82-5 to permit 1024B/1027B transcoding of 40GBASE-R to fit OPU3
- PCS lane recovery and deskew so that combined ingress/egress links do not exceed the skew budget. Lane recovery in G.709 uses block diagrams from Figures 82-10, 82-11 of approved IEEE Std 802.3-2012 to recover the PCS lanes

# Current "fast wake" process not compatible with OTN mapper

- In the current "fast wake" process the position of the alignment markers is different after the LPI section than it was before the LPI section
- The new alignment marker position is related to the point that the signal transitions from LPI
- This means that the mapper has to follow the RAMs so as to re-acquire the position of the normal alignment markers even though the alignment hasn't changed (the transmitter and receiver have been on continuously)
- This is a major change to the OTN mapper / demapper processing without any alignment issue being fixed

Possible EEE support for PCS codeword transparently mapped Ethernet over OTN
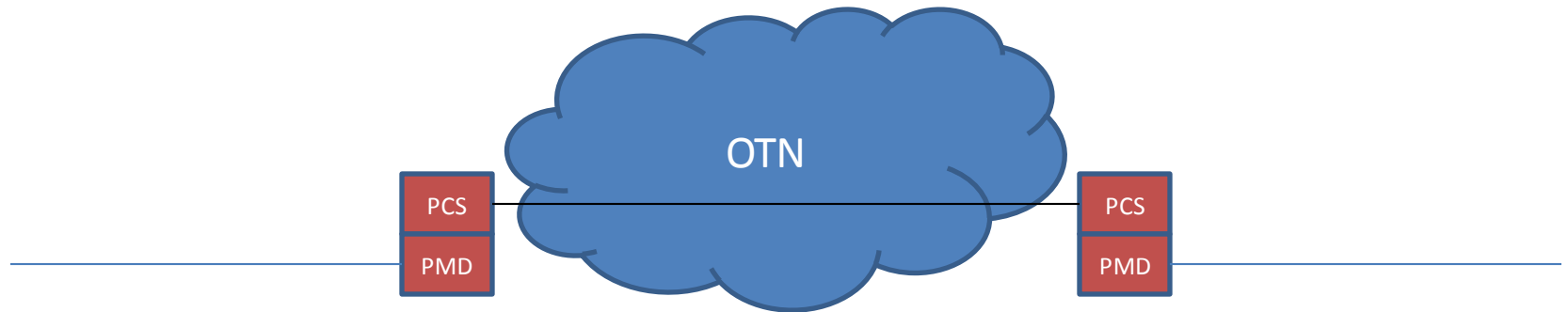
- "Fast Wake" only
- OTN mapper unaware of EEE
- EEE capability exchange between Ethernet endpoints transparently across the OTN (as if the connection were point-to-point)
- LPI assertion during quiescent period passed transparently over the OTN without change to PCS lane framing

Possible EEE support for GFP mapped Ethernet over OTN
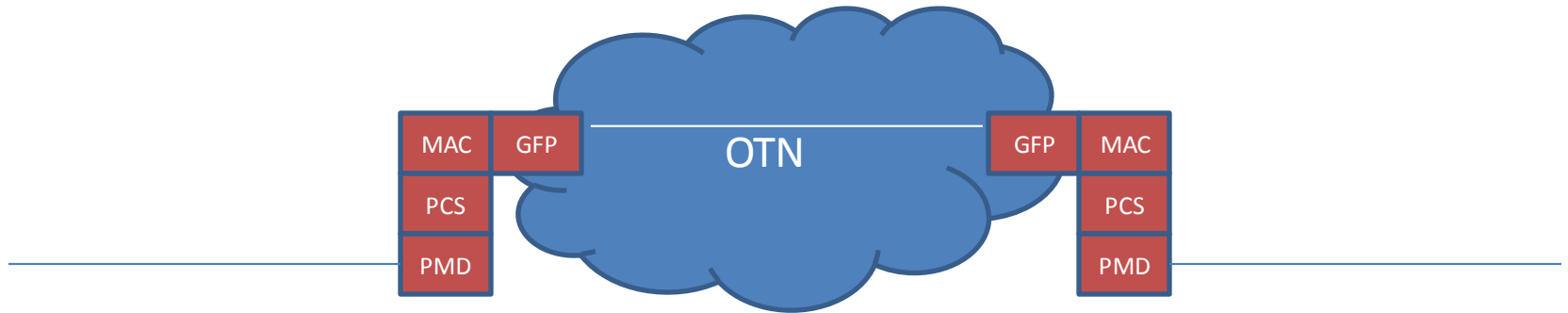
- EEE-capable ports on OTN network elements
- EEE capability negotiation independently on ingress and egress links
- Could propose to add new client-management frame to GFP to convey LPI status between ingress and egress for EPL type services
- Note that services may be more complex than point-to-point EPL, e.g., different VLANS are connected to different egress ports

Proposals

- Insert a warning that "Deep Sleep" EEE operation must be disabled for signals transparently mapped over OTN

- Change "Fast Wake" signaling to send LPI control characters while maintaining normal lane alignment markers to preserve compatibility with the OTN mapper

Implementation Considerations
OTN Mapper

- A new OTN mapper supporting a new P802.3bj PMD would need to terminate the FEC, trans-decode the 257B blocks back to 66B, and create a P802.3ba PCS as the "characteristic information" to map over an ODU4. The PMD types at the OTN ingress and OTN egress may be different, and all should be able to interconnect.
- P802.3bm will define (a) new 100GBASE-SR4 PMD(s) which likely has the same digital properties and implications as new P802.3bj PMDs.
- P802.3bm new 100Gb/s SMF PMD(s) TBD whether they require a new OTN mapper or can use an existing mapper.
- P802.3bm will define a new 40GBASE-ER4 PMD which is almost certain to use the same OTN mapper as 40GBASE-LR4
- A new OTN mapper in principle could implement new features, e.g., RAM if necessary (but this is not what is proposed)

Implementation Considerations
OTN mapper - continued

- P802.3bj will add EEE capability for three P802.3ba PMDs (40GBASE-KR4, 40GBASE-CR4, 100GBASE-CR10). P802.3bm will add EEE FW capability for six P802.3ba PMDs (40GBASE-SR4, 40GBASE-LR4, 40GBASE-FR, 100GBASE-SR10, 100GBASE-LR4, 100GBASE-ER4).
- Risk of not making a change: EEE-capable implementations of P802.3ba PMDs can be interconnected over existing OTN mappers, can negotiate EEE capabilities over LLDP (higher layer and invisible to OTN) and the mapping fails when a link enters LPI.

Simple Solution

- Fast Wake signaling maintains normal alignment marker spacing, continuously asserting LPI control characters.

- Wake indication changes from asserting LPI control characters to sending idle control characters.

- Transmission of data packets can resume after Tw expires from the time idle control characters are asserted.

- No other solution has been identified that would be compatible with the OTN mapper for the nine P802.3ba PMDs

Simple Solution Pro/Con

- Simple Solution for OTN-mapped PMDs supporting FW operation is as good as possible: OTN mapper must decode any RS FEC anyway because it may be interconnecting different PMD types.

- For a non-OTN Ethernet link with FW signaling, may be differences in how much can be turned off in the Ethernet Rx based on whether Wake is signaled in the alignment markers only or by an LPI->Idle transition in the data stream

Possible Savings and Alternate Solutions

- FW signaling needs to maintain 66B (or 257B) block alignment anyway as this is a prerequisite to alignment marker lock.

- Not looking at the bits between alignment markers could save you descrambling and lane interleave (negligible) and, where applicable, RS FEC decoding (possibly significant)

Can you recognize an LPI->Idle transition without FEC decoding?

- Yes! Hamming distance between an LPI control block (0x1e, 0x6, 0x6, 0x6, 0x6, 0x6, 0x6, 0x6, 0x6) and an idle control block (0x1e, 0x0, 0x0, 0x0, 0x0, 0x0, 0x0, 0x0, 0x0) is 16. Hamming distance between a 257B transcoded block comprised of four LPI control blocks or four Idle control blocks is 64. If those are the only expected receive values during FW signaling, it is easy to distinguish which is "closer" without decoding FEC, even in a high BER environment. Since the FEC code is systematic, data bits are not changed by the added parity bits.

Time to recognize transition out of LPI

- Existing draft mechanism requires countdown across five RAMs to signal "Wake" to Ethernet Rx. Across 20 PCS lanes for 100GBASE-R, this requires 512ns. Across 4 lanes of a 100 Gb/s P802.3bj interface, this requires 102.4ns. Across 4 lanes of a 40GBASE-R interface, this requires 256ns. Time to signal wake can be longer than Tw(min) from Table 78-4.

- After lane interleave, the 66B block time for 100GBASE-R is 640ps and the 257B block time is 2.4ns. For 40GBASE-R, the 66B block time is 1.6ns. Even if several blocks are allowed to recognize the LPI->Idle transition, this can be done significantly faster than the RAM countdown method.

- The standard would not need to specify how the transition is recognized or whether FEC is decoded − a time could be specified that is required to recognize the LPI->Idle transition that could accommodate a variety of implementations.