100 Gb/s Ethernet and EEE

Mark Gustlin

IEEE P802.3bj Task Force

Chicago September 2011

Contributors

Hugh Barrass – Cisco

Mike Bennett - LBNL

Wael Diab - Broadcom

Dan Dove – HP

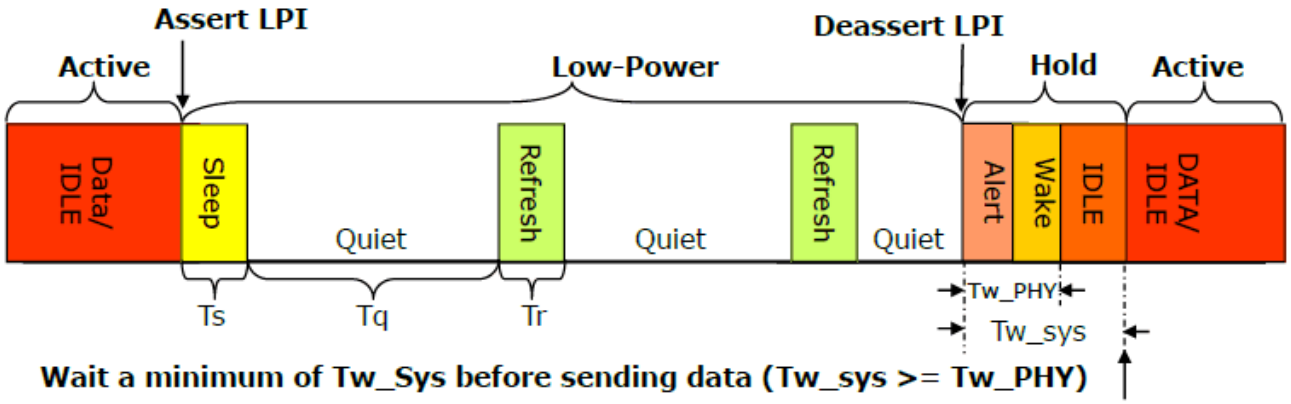
EEE for 100 Gb/s Overview

- This presentation has some early thoughts on EEE and 100 Gb/s backplane and copper Ethernet

- If we decide to support EEE within the 802.3bj project, what issues do we have to address?
  - What mechanism is used? Do we re-use Low Power Idle?
  - Is there anything inherent to the 802.3ba protocol that will cause problems with LPI?
  - The main issue identified is that the Alignment Marker lock time is very long, a possible solution to address this concern is described

EEE Review

LPI Overview



- LPI – PHY non-essential circuits shut down during idle periods
- During power-down, maintain coefficients and sync to allow rapid return to Active state
- Wake times for the respective backplane PHYs:
  - 1000BASE-KX: $Tw\_PHY_{(min)}$ = 11.25 usec
  - 10GBASE-KX4 $Tw\_PHY_{(min)}$ = 9.25 usec
  - 10GBASE-KR: $Tw\_PHY_{(min\ w/o\ FEC)}$ = 12.25 usec
  - 10GBASE-KR: $Tw\_PHY_{(min\ w/FEC)}$ = 14.25 usec

From: IEEE Presentation bennett_01_0311

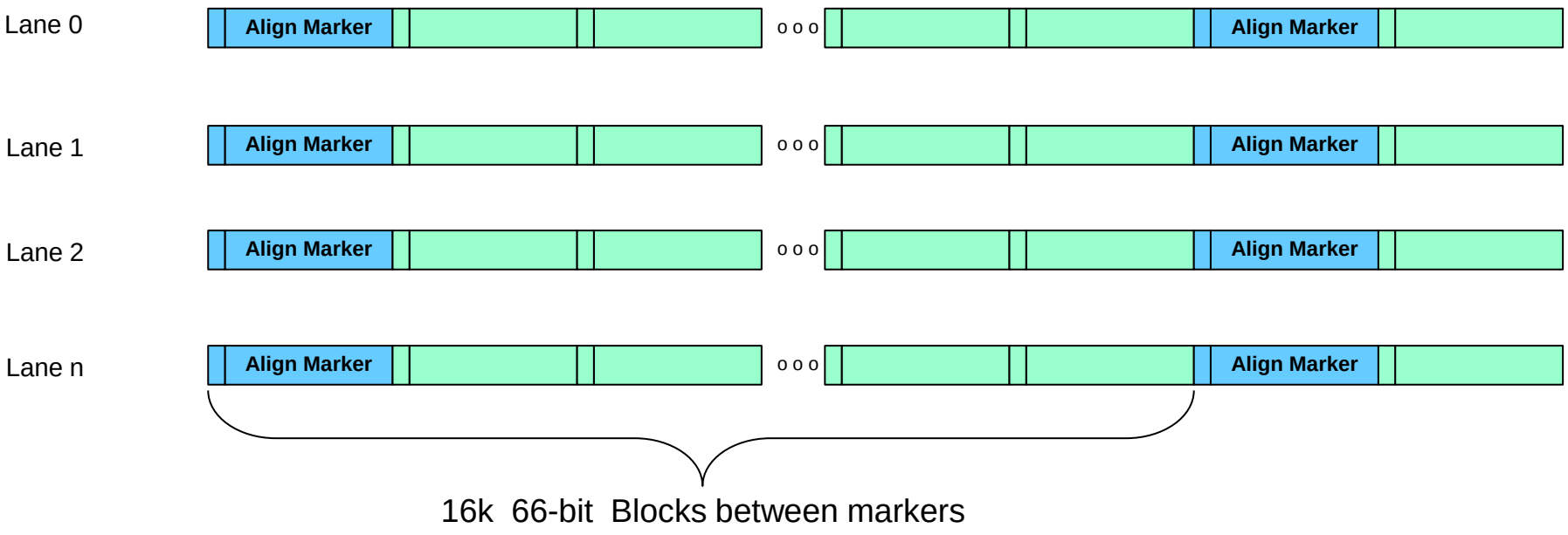
EEE Overview Continued

- Wake time range is 9 to 14usec
- Note that today wake time does not scale down with speed even though data accumulates faster at higher interface speeds
- So for 100 Gb/s should we shoot for a wakeup time of 10usec?
  - Note that in 10usec, 1Mb of data accumulates

- LPI seems to work well for lower speeds, and I don't see any issues on why it can not be re-used
- Are there any concerns in the 100/40 Gb/s PCS that would prevent us from supporting a 10usec wakeup?
  - Alignment marker lock is >> 10usec, the next few slides look at this issue

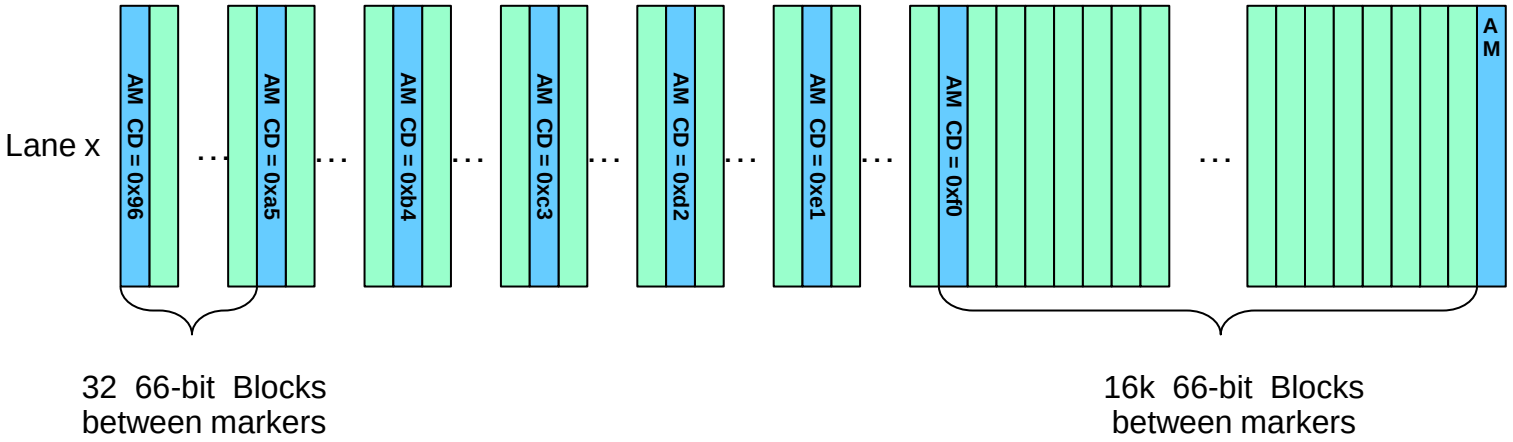
100GE Standard Alignment Marker Distance

- The alignment markers are widely spaced for 100 Gb/s and 40 Gb/s, 16k blocks apart on each PCS lane
- The alignment marker lock SM looks for two that match in a row before declaring lock and allowing alignment, so that is (16384*2 * 66 * 194psec) = 419usec (for 100GE)
- This means that startup would take > 400usec today!
  - Ok for 802.3ba where startups occurs rarely
  - Not ok for a EEE interface where the interface goes up and down frequently



Start-up AM Distance Reduction

- When the lanes are starting up, reduce the distance between AMs temporarily
- The allowed minimum distance is dependent on the skew that is allowed (14 66-bit words for 100G at the RX PCS, 28 words for 40G)
- So let's say every 64 words there is a marker until startup is finished, then revert to normal distance
- Alignment Marker lock would now take at least 64 * 2 * 66 * 194psec = 1.64usec (for 100GE), it can take longer with errors



EEE for 100 Gb/s Summary

- The majority of the EEE protocol that was developed for 10 Gb/s could be applied directly to EEE at 100 Gb/s (using the Low Power Idle)
- The main protocol hurdle that 100 Gb/s has to overcome in order to support EEE is the Alignment Marker distance and startup time due to this distance
- Presented is one possible and simple change to the Alignment Marker protocol which will greatly reduce the start up time and enable efficient EEE support at 100 Gb/s