An Overview of Energy-Efficient Ethernet

100G Backplane and Copper Study Group

Michael J. Bennett
Lawrence Berkeley National Laboratory
Topics

• Overview of Energy-Efficient Ethernet
  – Low Power Idle

• How could this apply to 100G BP/CU?

• Considerations and Open Questions
What is Energy-Efficient Ethernet?

• Energy Efficient Ethernet (EEE) is a method to reduce energy used by an Ethernet device during periods of low link utilization
• Specified in IEEE 802.3az-2010™
• The premise for EEE is that Ethernet links have idle time and thus opportunity to save energy
• Specified for copper interfaces
  • “BASE-T’s’
  • Backplane (except 40G)
• The method is called Low Power Idle (LPI)
Where EEE Fits

<table>
<thead>
<tr>
<th>HIGHER LAYERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical Link Control (LLC)</td>
</tr>
<tr>
<td>MAC Control (optional)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Media Access Control (MAC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECONCILIATION</td>
</tr>
</tbody>
</table>

| PCS |
| PMA |
| AUTO-NEGOTIATION |

| MEDIUM |

**ELP**

<table>
<thead>
<tr>
<th>LPI Client</th>
</tr>
</thead>
</table>

**xxMII**
What is Low Power Idle?

- Concept: Transmit data as fast as possible, return to Low-Power Idle
- Saves energy by cycling between Active and Low Power Idle
  - Power reduced by turning off unused circuits during LPI
  - Energy use scales with bandwidth utilization
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: \( T_{PHY}^{(min)} = 11.25 \text{ usec} \)
  - 10GBASE-KX4: \( T_{PHY}^{(min)} = 9.25 \text{ usec} \)
  - 10GBASE-KR: \( T_{PHY}^{(min \text{ w/o FEC})} = 12.25 \text{ usec} \)
  - 10GBASE-KR: \( T_{PHY}^{(min \text{ w/FEC})} = 14.25 \text{ usec} \)

Wait a minimum of \( T_{sys} \) before sending data (\( T_{sys} \geq T_{PHY} \))
Example: 10GBASE-KR
10GBASE-KR

- LPI can save energy at the PHY

- L2 features enable energy savings beyond the PHY
How Could This Apply to 100GBP/CU?

• I am assuming 100GBASE-R PCS
How Could This Apply to 100GBP/CU?

• You could use LPI
  • Distribute Low Power Idle across lane
  • How to deal with the other signals, e.g. sleep, wake, refresh, etc.

• You could choose to do something else
  • May be able to leverage some of the work done in EEESG
  • May do something completely new
Things to consider

• When deciding what to work on in 802.3az, we had to consider the potential energy savings
  – If you plan to do the work in 100GBPcu, a similar analysis should be done

• The amount of work may be substantial
  – How would this impact your schedule?
Things to consider

• 802.3az requires the use of the MAC defined in Annex 4A for simplified full duplex operation (with carrier sense deferral).
  – This provides full duplex operation but uses the carrier sense signal to defer transmission when the PHY is in its low power state

• Latency
  – It takes time to wake up the PHY
    • E.g. 10GBASE-KR has a wake time of 11.25 usec
    • How long is reasonable for a 100G BP PHY?
    • How long is reasonable for a 100G Cu PHY?
Open Questions

• What is the estimated power savings in a low power mode?

• Does Low Power Idle make sense for 100G Backplane? Copper cable?
  – If yes then you get to leverage the work done in 802.3az
  – If not then?

• How long can a 100G Backplane/Copper cable link remain quiet before it degrades to the point it needs to be restarted?
Open Questions

• How much time will it take to wake up 100G backplane PHY? Copper cable?
  – How much time can be tolerated?

• If there is a latency objective, should there be an exception for EEE version of 100G Backplane and / or Copper Cable?
Thank You!