IEEE 802.3az
Energy Efficient Ethernet

Task Force Update

Presented to the P802.3ba Task Force

Denver, CO
July 16, 2008

Mike Bennett
mjbennett@ieee.org

Acknowledgement

- Thanks to Dan Dove, Rob Hays, and David Law for their contributions

Discussion

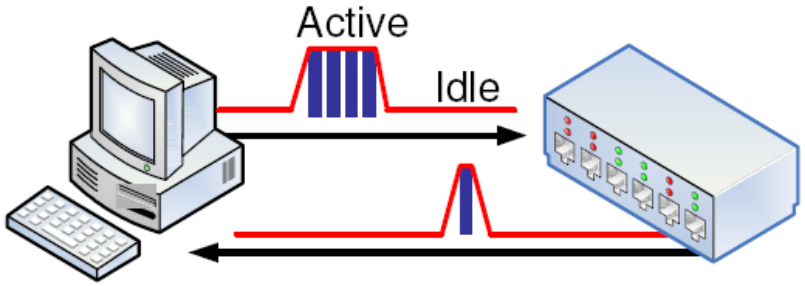
- Brief overview of Energy Efficient Ethernet (EEE)
- IEEE P802.3az – current status
- Things to consider

Briefly, what is EEE?

- EEE is a method to facilitate transition to and from lower power consumption in response to changes in network demand
  - In the process of being defined for these copper PHYs
    - 100BASE-TX (Full Duplex)
    - 1000BASE-T (Full Duplex)
    - 10GBASE-T
  - Expecting to adopt proposals for backplane PHYs this week
    - 10GBASE-KR
    - 10GBASE-KX4
    - 1000BASE-KX
  - Uses Low Power Idle (LPI) to save energy

What is Low Power Idle?

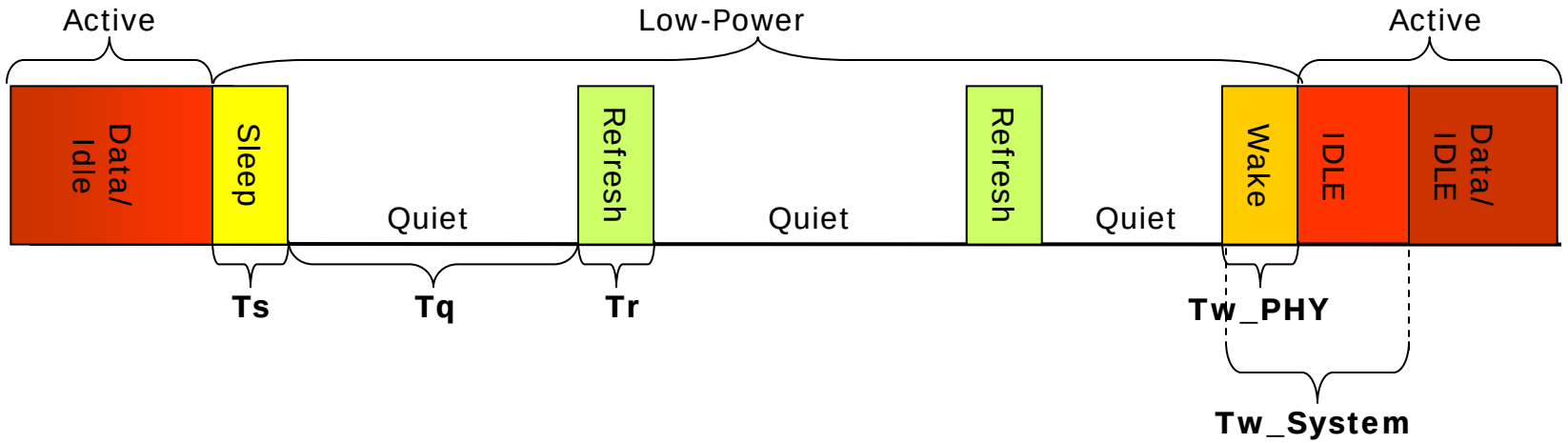
- Concept: Transmit data as fast as possible, return to Low-Power Idle
- Saves energy by cycling between Active and Low Power Idle
  - Power reduced by turning off unused circuits during LPI
  - Energy use scales with bandwidth utilization



What is Low Power Idle?

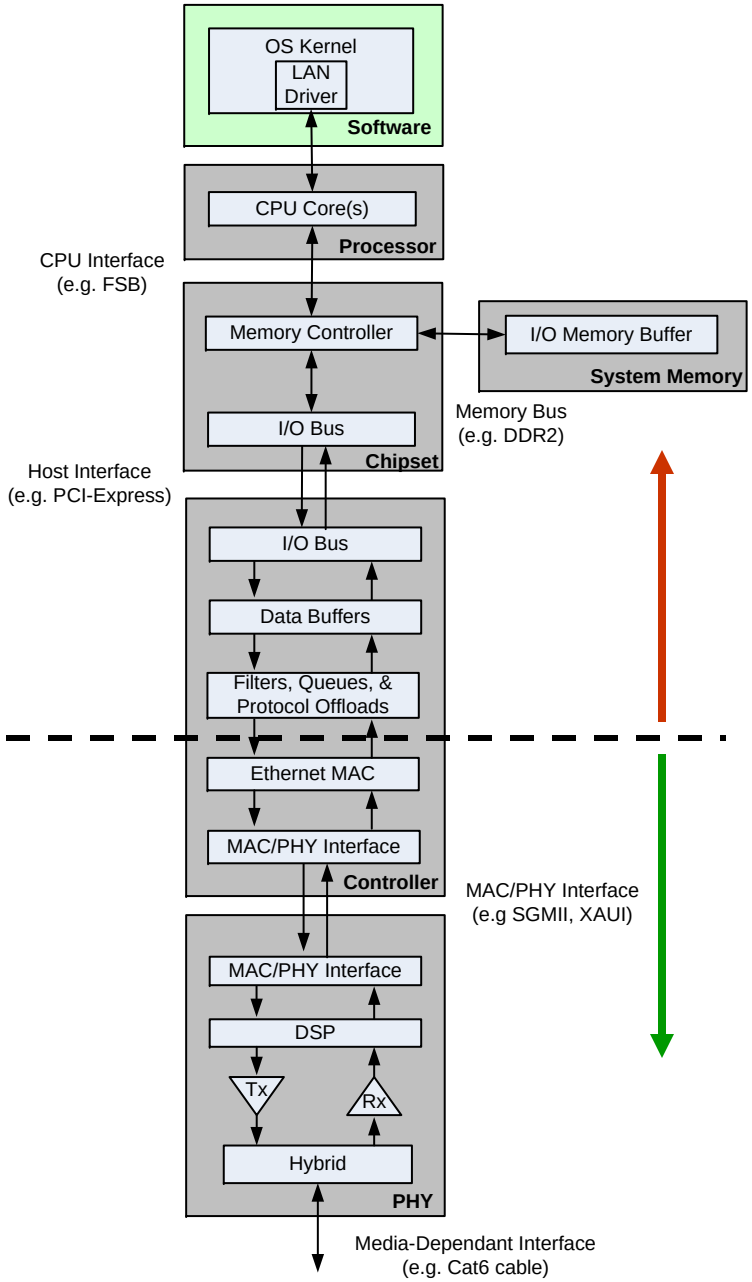
- A closer look
  - We've been talking about Tw values around 10 usec

Term	Description
Sleep Time (Ts)	Duration PHY sends Sleep symbols before going Quiet.
Quiet Duration (Tq)	Duration PHY remains Quiet before it must wake for Refresh period.
Refresh Duration (Tr)	Duration PHY sends Refresh symbols for timing recovery and coefficient synchronization.
PHY Wake Time (Tw_PHY)	Duration PHY takes to resume to Active state after decision to Wake.
System Wake Time (Tw_System)	Wait period where no data is transmitted to give the receiving system time to wake up.

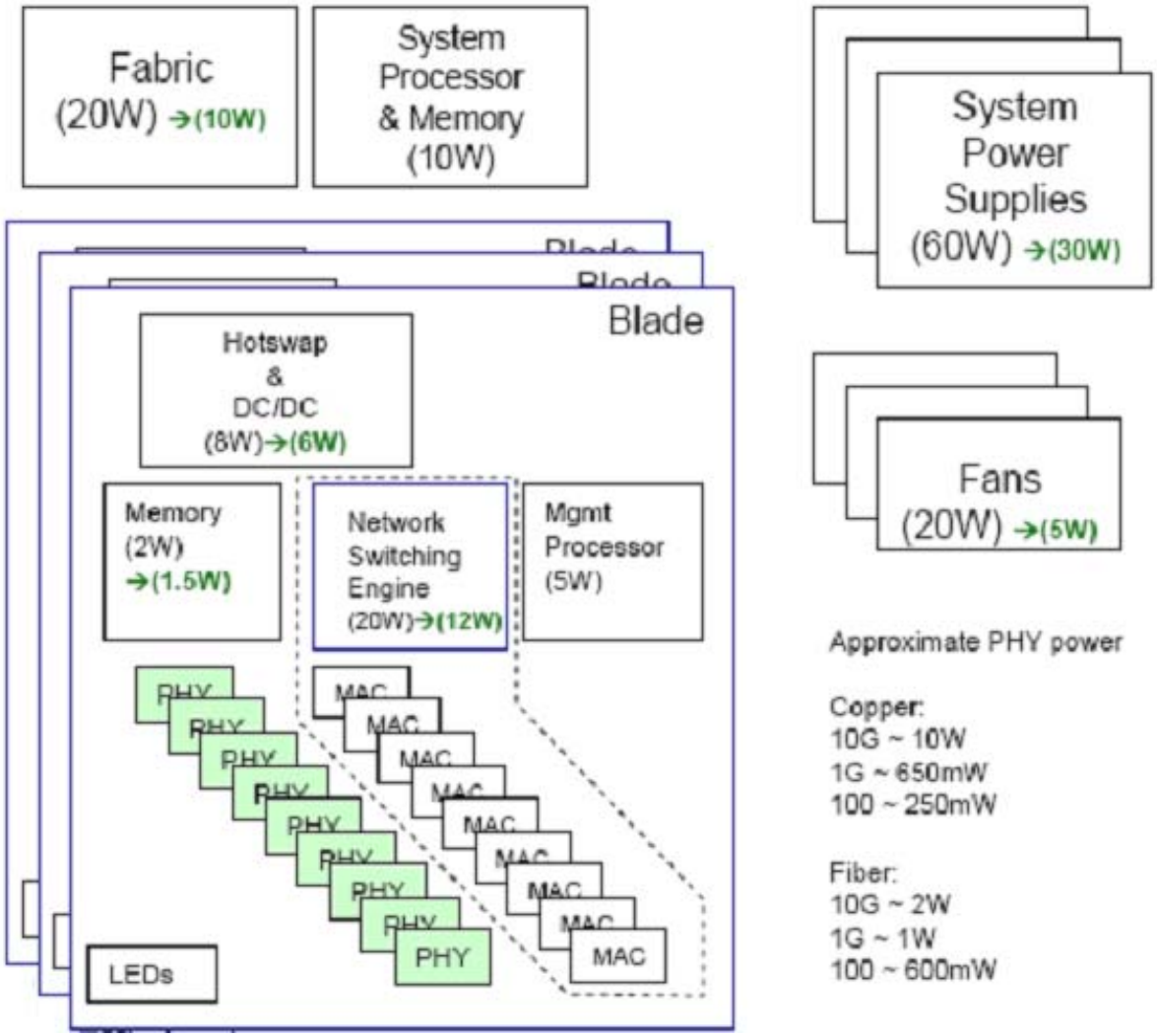


A system view

- Outside the scope of our work
  - Control policy
  - System power savings
- In scope
  - PHY power savings
  - Auto-negotiation
  - Management
  - Protocol to communicate parameter changes



A system view (switch centric)



Approximate PHY power

Copper:
10G ~ 10W
1G ~ 650mW
100 ~ 250mW

Fiber:
10G ~ 2W
1G ~ 1W
100 ~ 600mW

Switch MAC, NSE, Memory are a good portion (~3x/port) of energy consumption for most networking link technologies.

Powering-down portions of these circuits provides a two-fold benefit

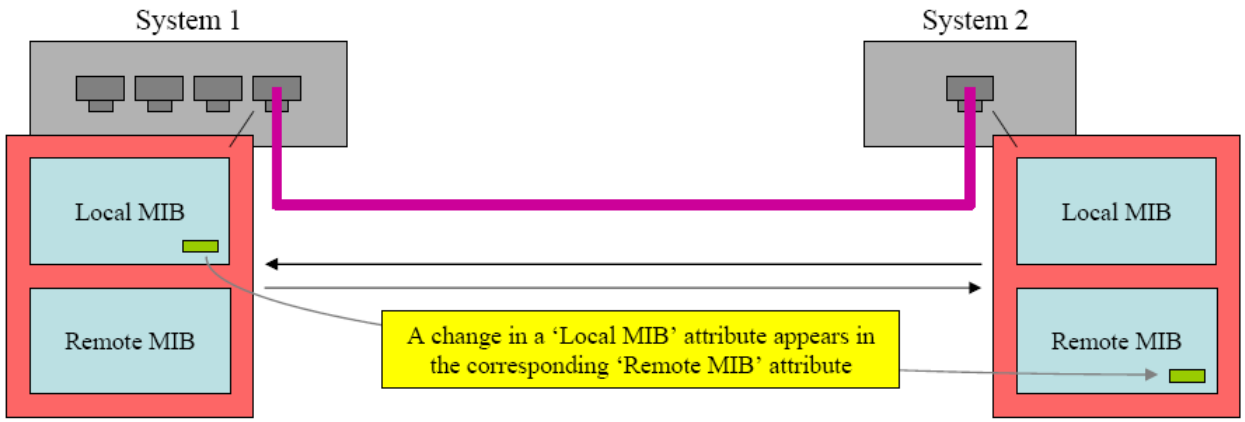
1) Reduces energy used
2) Provides opportunity to shut-down other infrastructure (DC/DC, Fans, etc)

Reasonable estimates show that **~1.5W- 3W/port** can be reduced in infrastructure

What to power-down and how to do it, is outside the scope of 802.3, but providing means to communicate when to power-down and when to resume operation may be appropriate for 802.3 to address

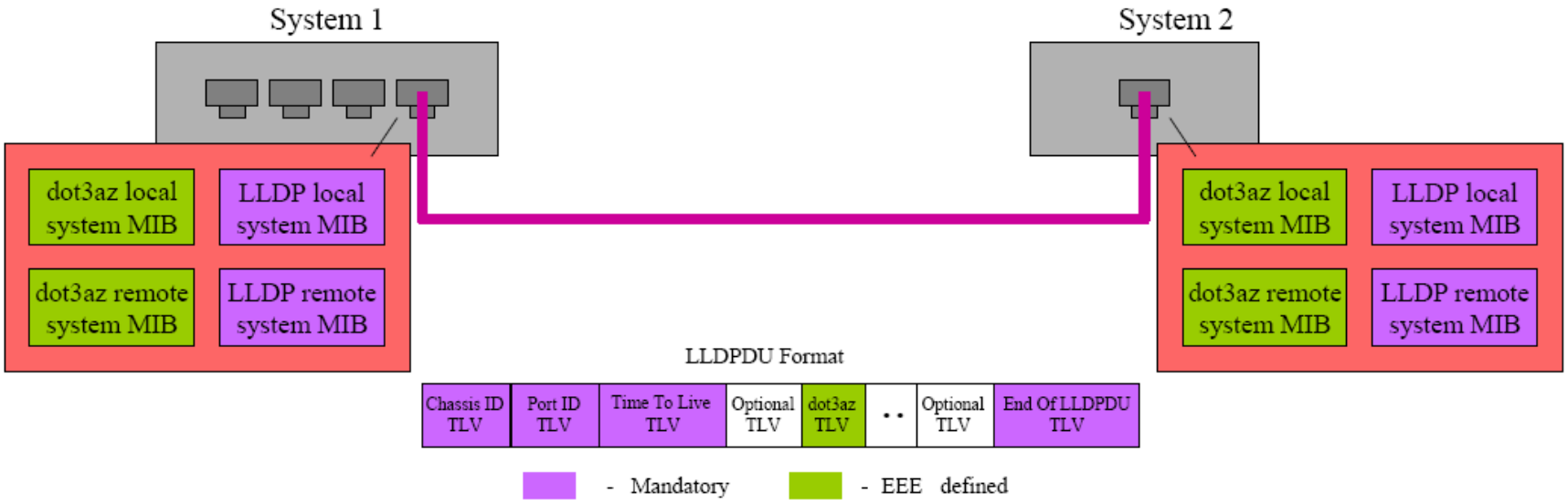
Link partner communication

- Will use auto-negotiation to notify link partner of EEE capabilities
- Will use LLDP to notify link partner of parameter changes
  - E.g. control policy
    - User can choose energy savings preferred over performance or vice versa



LLDP

- We'll need to define LLDP MIB extension and TLVs as well as an EEE MIB



Current status

- **We've adopted several proposals for our baseline**
  - Reduced Amplitude 10BASE-T
  - Low Power Idle (LPI) for 100BASE-TX
  - LPI for 1000BASE-T
  - LPI for 10GBASE-T
  - Use of LLDP to communicate between link partners after auto-negotiation
- We have an editor's draft incorporating these proposals

Current status

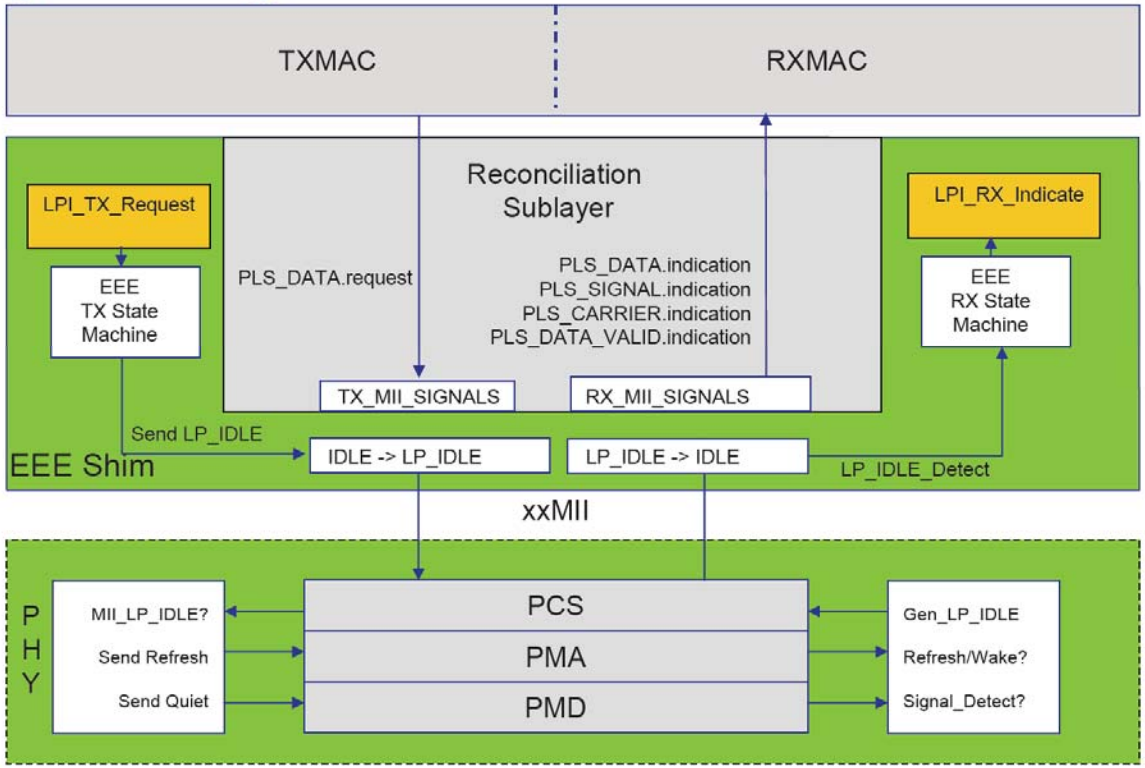
- **We're considering backplane proposals for our baseline**
  - 10GBASE-KR
  - 10GBASE-KX4
  - 1000BASE-KX
    - The work is being done, will likely add an objective

- Assuming we get through the remaining proposals this week
  - We'll bring in the material from the new proposals for review in September
  - Focus on
    - filling in TBDs and holes in the draft
    - Possibly Modifications to the Reconciliation Sublayer and MIBs

Things to consider

- LPI is “architecture agnostic”
  - It will work on a serial interface
  - It will work on a multi-channel interface

Layer Diagram



Things to consider

- The question about applying EEE to optics has come up a few times
  - We don't have objectives for optical PHYs
    - Doesn't necessarily mean this can't be applied to optical PHYs
  - Suggested in Onn Haran's *Applicability of EEE to Fiber PHYs* http://www.ieee802.org/3/eee_study/public/sep07/haran_1_0907.pdf
    - Interesting idea – just may have been premature
  - Can you "gate" optics and still achieve reliable operation?
    - We didn't study this

Things to consider

- Latency
  - EEE operational mode adds additional latency to be considered by the network designer
  - When at Low Power Mode, PHY device is not available immediately for data transmission request.
    - System has to wake it up by sending normal idle code on the MAC interface.
    - Following IDLE code reception on the MAC interface, PHY starts waking up process.
    - The max PHY recovery time Tw is different for different PHY types

Things to consider

- Latency
  - We will specify the maximum values of Tw as soon as we have agreed on them

Table 4—

Protocol	Tw
10GBASE-KR	TBD
10GBASE-KX4	TBD
10GBASE-T	TBD
1000BASE-T PHY	TBD
100BASE-TX	TBD

Thank You!

Back-up

Objectives

- Define a mechanism to reduce power consumption during periods of low link utilization for the following PHYs
  - 100BASE-TX (Full Duplex)
  - 1000BASE-T (Full Duplex)
  - 10GBASE-T
  - 10GBASE-KR
  - 10GBASE-KX4
- Define a protocol to coordinate transitions to or from a lower level of power consumption
- The link status should not change as a result of the transition
- No frames in transit shall be dropped or corrupted during the transition to and from the lower level of power consumption
- The transition time to and from the lower level of power consumption should be transparent to upper layer protocols and applications
- Define a 10 megabit PHY with a reduced transmit amplitude requirement such that it shall be fully interoperable with legacy 10BASE-T PHYs over 100 m of Class D (Category 5) or better cabling to enable reduced power implementations
- Any new twisted-pair and/or backplane PHY for EEE shall include legacy compatible auto negotiation