Energy Efficient Ethernet, Wake Signals and Deep Sleep for Automotive Ethernet

IEEE P802.3ch Multigig Automotive Ethernet
PHY Task Force

George Zimmerman/CME Consulting, Inc

Overview

- EEE – Review
- Power Management States and relation to EEE
- Buntz Wake/Sleep Concept
- Observations
- Recommendations

Sources for EEE Overviews

- "An Overview of Energy Efficient Ethernet" Mike Bennett, youtube talk (TeamNANOG, June 2014)
  - Includes discussion of fast wake/deep sleep, optics, recent projects
- "IEEE 802.3az: The road to energy efficient Ethernet", Ken Christensen, et. Al, IEEE Communications Magazine, December 2010
- "Energy Efficiency and Regulation" Bruce Nordman, et al. (IEEE 802 tutorial, July 2008)
  - (see slides 27-50) EEE and WiFi energy management
- "Baseline Summary", IEEE 802.3az Task Force Home Page, March 2009
  - A nice summary of the 802.3az PHY modes
- "IEEE 802.3az Energy Efficient Ethernet Task Force Update", 802.1/802.3 joint meeting Mike Bennett, July 2008
  - Update on 802.3az approaches, scope related to systems issues
- "IEEE 802 Tutorial – Energy Efficient Ethernet", Hugh Barrass, et al. (IEEE 802 tutorial, July 2007)
  - Objectives, explanation of 802.3az motivations and approaches
- A Brief Tutorial on Power Management in Computer Systems, Dave Chalupsky & Emily Qi (IEEE 802.3 EEE Study Group, March 2007)
  - Covers both link and Systems level power management

EEE – where it started

- Designed for bursty traffic and fast recovery
- Based on “Adaptive Link Rate”:
  - *Ethernet Adaptive Link Rate: System Design and Performance Evaluation,* Gunaratne, C.; Christensen, K.; Proceedings 2006 31st IEEE Conference on Local Computer Networks, Nov. 2006 Page(s):28 -35
- Motivated by data centers
  - Low latency response (microseconds)
- Layered – meant to operate without impact to network applications

IEEE 802.3az Designed for bursty data – not device states

- Designed for bursty data
- Designed for fast, application-transparent recovery
- "The link status shall not change as a result of the transition"

Desktop links have low utilization

- Snapshot of a typical 100 Mb Ethernet link
  - Shows time versus utilization (trace from Portland State Univ.)

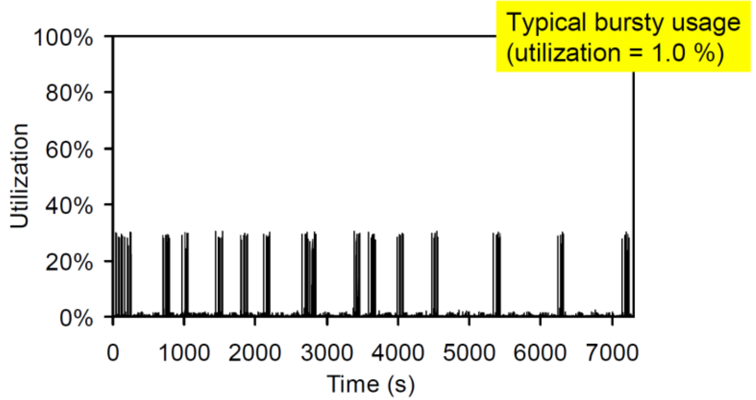


Fig1.xls

July 16, 2007 9

Transition Time Conclusions

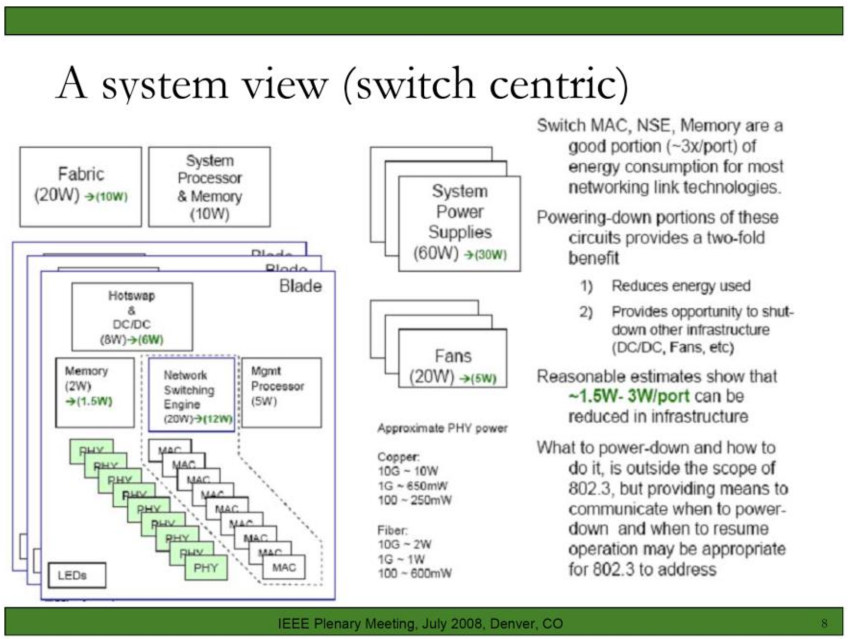
- Applications require sub 10 ms transition time
- Recommend that the EEE TF retain the goal of achieving a transition time of less than or equal to 1 ms

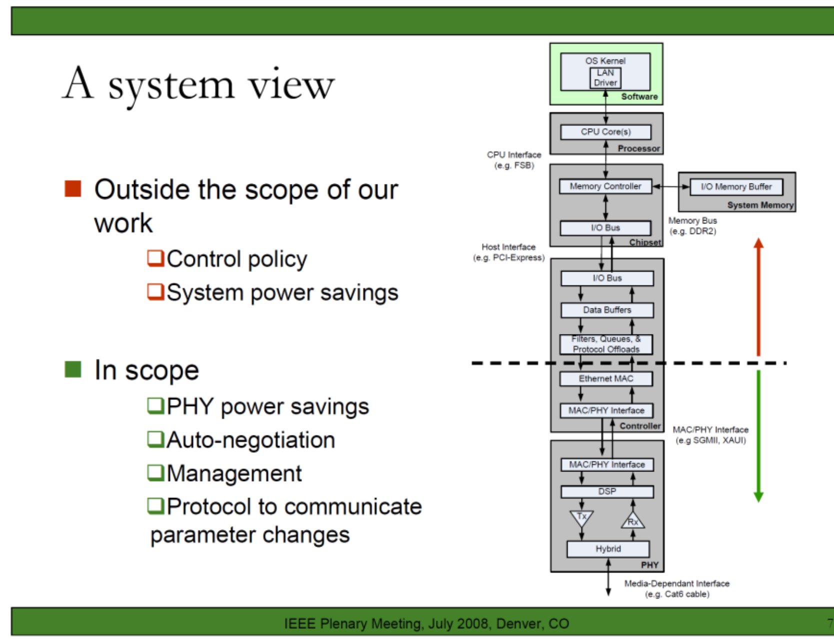
July 16, 2007 37

Source: "IEEE 802 Tutorial – Energy Efficient Ethernet", Hugh Barrass, et al. (IEEE 802 tutorial, July 2007)

EEE Scope

- System power savings & control are out of scope
- Focus on the PHY/MAC
- Signaling starts and ends with the MAC interface





Source: “IEEE 802.3az Energy Efficient Ethernet Task Force Update”, 802.1/802.3 joint meeting
Mike Bennett, July 2008

EEE - Principles

- “LPI Client” controls transitions to/from low-power states
- MAC/RS Layers (802.3)
  - MAC: maintains link
  - RS supports new code: LPI
  - Signalling above the MAC is unchanged
- PHY Layer (802.3) –
  - Defined signaling during LPI (mostly off)
  - (Sometimes) special signals: “WAKE”, “ALERT”, “SLEEP”
  - Capabilities negotiated at startup or via LLDP
    - Deep sleep – TX quiet/refresh
    - Fast wake – optical concept
      - TX is on, only higher layer PHY functions suspended
    - Long wake times may be negotiated

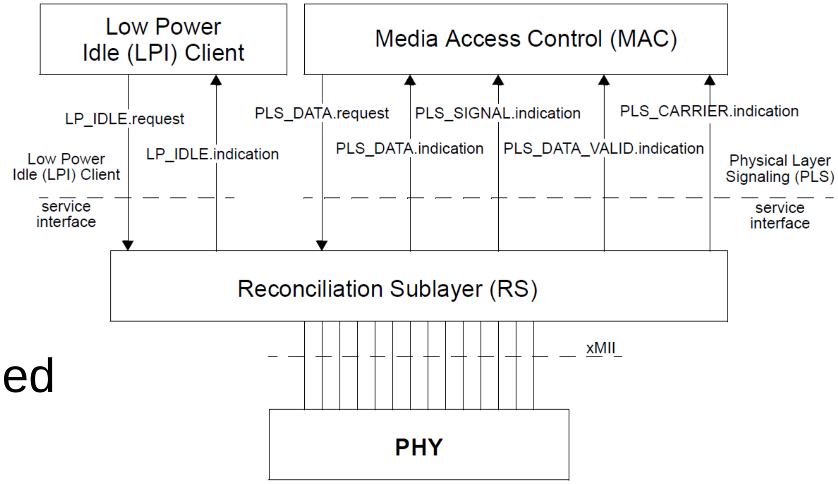


Figure 78–1—LPI Client and RS interlayer service interfaces

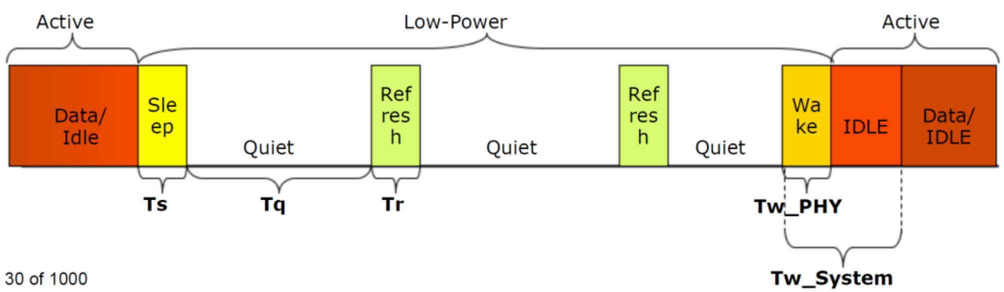
At the PHY: Designed for Fast Recovery

- “Refresh” keeps loop timing and link security
- Thermal constants drive “Fast Wake” for optics

What is Low Power Idle?

- A closer look
  - PHY Wake Time values are in the order of 10’s of microseconds

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.**



Slide 30 of 1000

Source: “Energy Efficiency and Regulation” Bruce Nordman, et al. (IEEE 802 tutorial, July 2008)

Refresh – different for different PHYs

- 10GBASE-T – most sensitive
  - Frequent refresh
  - Sensitivity to timing (e.g., echo cancellers) drives refresh timing
    - 100BASE-TX/1000BASE-T have longest T_quiet at 24msec
- Backplane
  - Refresh resyncs timing/FEC
- BUT: transparency to applications means link is never taken down

Power Management States

- Active Power Management States are defined at various levels, usually by system vendors
- Hierarchical
  - Global power states
  - System power states
  - Device power states
  - Link power states
- NOT application transparent (systems are “put to sleep” at various levels)
  - Transitions can be long

Global System State Definitions

G States reflect the User’s perception of the machine.

G0 Working (S0)
- **A computer state where the system fully operational.**
- **It is not safe to disassemble the machine in this state.**

G1 Sleeping (S1-S4)
- **Power consumption is small and the system “appears” to be off.**
- **Work can be resumed without rebooting the OS.**
- **Large elements of system context are saved by the hardware and the rest by system software.**

G2 Soft Off (S5)
- **The computer consumes a minimal amount of power.**
- **This state requires a large latency in order to return the Working state.**
- **The system’s context will not be saved. The system must be restarted.**

G3 Mechanical Off
- **This state that is entered by a mechanical means (i.e. power switch).**
- **The OS must be restarted and no hardware context is retained.**
- **Except for the real-time clock, power consumption is zero.**

20 March 11, 2007 File: chalupsky_01_0307 Energy Efficient Ethernet intel

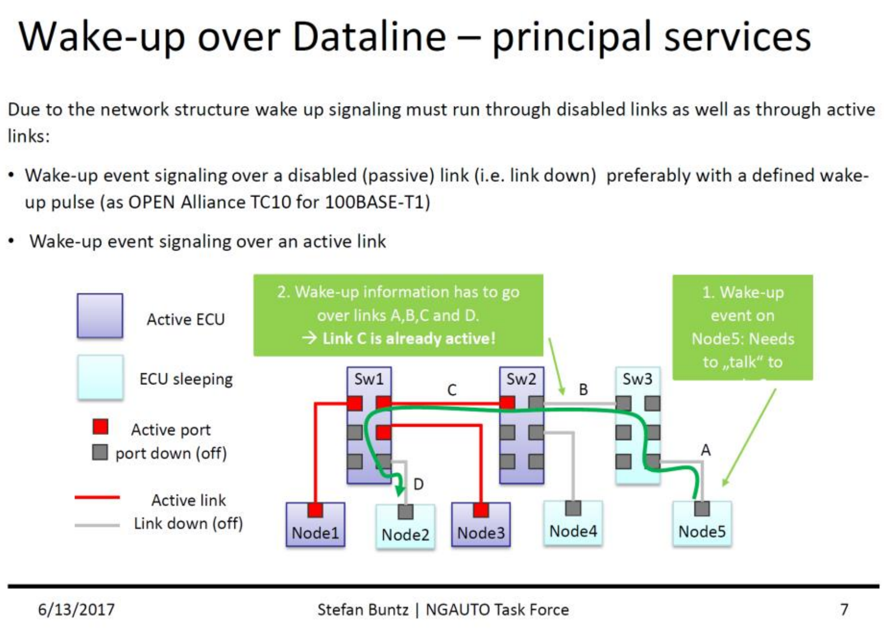
Source: A Brief Tutorial on Power Management in Computer Systems, Dave Chalupsky & Emily Qi (IEEE 802.3 EEE Study Group, March 2007)

EEE is used within power management

- Link power state
- Link power state may inform other system power savings states
- Management of system power and out-of-band signaling is outside 802.3 scope
- Hooks for transitioning, signaling and managing link state are within scope

Buntz: Wake/Sleep concept

- Full PHY power down
  - Loss of timing
  - Minimal power
- Wake-up event signaled on PHY link
  - Long wake (100ms) allowed
- Wake-up propagated to other links/ECUs
  - Startup of other PHYs in parallel



Source: "NGAUTO Sleep/Wake Concept", S. Buntz Daimler (IEEE 802.3ch ad hoc, June 2017)

Other 802 standards have similar needs

- Wireless standards have constraints similar to automotive
  - Battery operation drives need for very low current draw for long time periods

802.11 Power Save Features

- **Power Save (PS-Poll) – 802.11-2007**

  A STA enters PS mode, and while in in PS mode it listens to selected beacons. If beacon indicates packets buffered for that STA, the STA sends a PS-Poll frame to the AP to fetch the packets.

- **Automatic Power Save delivery (APSD) – 802.11-2007**

  The power save mechanism for Quality-of-Service (QoS) enabled STA and APs and allows per-stream (Access Category) power save.

- **Fast BSS Transition – Pub. 802.11r-2008/802.11mb**

  A STA can roams between 2 APs in a power efficient manner. Fast BSS transition allows use of a security key hierarchy and QoS negotiation during the fast transitions process.

Approved

Slide 47 of 1000

802.11 Power Save Features

- **Proxy ARP – 802.11v, WiFi Alliance (WFA)**

  The AP has the ability to proxy ARP frames for the STA. This is intended to enable the STA to remain in power-save for longer periods of time.

- **TIM Broadcast – 802.11v**

  AP periodically transmits a TIM frame, which is shorter than Beacons and transmitted at a higher rate, to indicate traffic buffered for a STA.

- **WNM Sleep Mode - 802.11v, WFA**

  WNM Sleep Mode is an extended STA power save mode in which a STA need not listen for every DTIM Beacon frame, and AP does not perform GTK/IGTK updates.

In-Development

Slide 48 of 1000

Source: “Energy Efficiency and Regulation” Bruce Nordman, et al. (IEEE 802 tutorial, July 2008)

OPEN Spec (ref. Buntz)

- Disclaimer: I am not affiliated with the OPEN Alliance, and do not speak for them.
  - The below is based on my reading of the TC10 specification
- Defines a “Low Power Sleep” state and a request protocol to transition
  - Shuts link down “without generating interrupts”
    - Link state at MAC is unclear
  - Defines wakeup pulses (training codes w/energy detect) used for inactive links
    - Principle similar to EEE WAKE or ALERT signaling
  - Defines wakeup signaling for active links
  - Budgets timing and current draw across components
  - Requires ‘selective wakeup forwarding’ on multi-PHY devices
  - Describes ‘inhibit pin’ and out-of-band wakeup pin signaling

Observations

- OPEN specification at the PHY level is like EEE w/o refresh
  - Additional multi-port device and pin specifications out of 802 scope
- OPEN specification mixes and blurs product vs. protocol specifications, does not show MAC interface
- Buntz requirement suggests another power management state beyond EEE
  - Low power with application-impacting recovery times
    - Allowed to drop synchronization and retrain
    - No “asymmetric” signaling?
- PHY-level signaling to SLEEP/WAKE/ALERT is already well proven in EEE
- Lots of work needed to develop this power management, much may be beyond 802.3’s scope
- System level management and parallel wake-up is more like power management states than EEE, and is application-impacting
  - Outside 802.3 scope, System-specific, already well-trodden path in IT industry

Recommendations

- Consider PHY signaling to support fast start
  - 100msec already in objectives
  - Consider PHY signaling to save state in prep for fast start
- Consider new CFI for ‘extra deep sleep’ EEE
  - Leverage IT industry “power management” work.
  - Enable fast restart from a saved state
  - Likely have uses in other-than-automotive (IMHO)
  - Not a primary job of 802.3ch
- Discuss here, consider an NEA application