End-Stations System Requirements and a proposal for EEE Objectives

Ilango Ganga, Intel
David Chalupsky, Intel

Key Messages

- High ROI, if the Network Interface power savings (EEE) is aligned with System Power Management architecture
- Need to study End-Station requirements: desktop, mobile clients and servers
- EEE is not just be about saving a few watts in the network interfaces. The benefits extend into system level power optimization.
- EEE should consider all PHY types: BASE-T, Backplane and Optical
- Strawman proposal for objectives based on End-Station requirements

EEE background

- Change PHY Speed based on Link Utilization
  - Higher speed to Lower speed or vice versa
  - For example: 10Gbps to 1Gbps or 1Gbps to 100Mbps
- A proposal for Rapid PHY Selection (RPS)
  - Some existing NICs use Auto-Neg mechanism to change Link Speed to reduce power
  - Auto-Neg is slow protocol (may take up to 3 sec)
  - Need for a faster protocol to change Link Speed
  - Change speed without bringing down the link (goal 1 ms)
- Control Policy
  - To determine when to transition PHY speed
  - Control policy will not be defined by the Standard

EEE and System Power Management

- High ROI, if the Network Interface power savings (EEE) is aligned with System Power Management Architecture
- Control Policy can be dictated by System Power Management architecture
- Include Blade system requirements, Blade Server market is fastest growing segment
- Define EEE features to be consistent with End-Station needs; include consumer electronics, desktop, mobile and servers

Consider an Electrically Idle Line - “0 Mbps”

- **Some PHY types may find it more expedient to transition between active and electrical idle states than between different link speeds**
  - Many 802.3 PHYs consume the same power regardless of whether data is being transmitted
    - Idle symbols in 100/1000/10GBASE-T, 1000BASE-X, 10GBASE-X & R
- **Working definition of Electrical Idle**
  - Transmitter turned off
    - Receiver may still be active if partner transmitting data, or at reduced functionality if partner is not transmitting. Will vary by PHY type.
- **Examples**
  - Many 10/100/1000BASE-T PHYs implement something similar today when cable disconnected. High resume latency, though, since Clause 28 Auto-Neg issued when receiver detects a link partner.
  - Serial interconnects, like PCI Express, implement this
- **“0 Mbps” merits study. Effectiveness will vary by PHY type.**
- **RPS can be used to indicate transition to “0 Mbps” as well as other speeds.**

End-Station requirements..1

- Speed change consistent with System power states
  - Enable control policies correlated with system power state architecture
  - Higher than 1ms latency may be acceptable for lowest power states
- PHY Requirements
  - Include 0 Mbps speed as an option (e.g., 0M to 1G and vice-versa)
  - Need for Link IDLE and PHY standby modes
  - Rapid wake-up from PHY standby state
  - Protocol
    - Speed selection protocol to accommodate signaling standby modes and proxy messaging
  - Management
    - Need for reading and controlling the PHY power saving capabilities
- Consider Wake On LAN (WOL) and mobile client needs
  - Include 10Mbps speeds, Link IDLE / PHY Standby modes
  - Wake-up signal (not frame based)

End-Station requirements..2

- Control policy based on Link Utilization and System performance
  - Speed change to include following speeds and PHY types
    - Servers 0G, 100M, 1G, 10G, (copper, fiber and backplane PHYs)
    - CE, Desktop, Mobile 0M, 10M, 100M, 1000M (copper PHYs)
    - Devices need not implement all speeds
  - Change speed without bringing the link down
    - Might cause congestion spreading, Need to study flow control requirements
  - Latency goal: 1ms (proposed in CFI)
    - Need further study on feasibility of different PHY types including 100M, 1000M and 10G PHY types
- Management
    - Need for reading and controlling the PHY power saving capabilities

Strawman Proposal for Objectives

- Mechanism to change PHY speed
  - 10Gbps to 1G/100M, 1G to 100/10M & vice versa, Backplane & Optical PHYs
  - Change speed without dropping link
  - Few milliseconds (1ms goal proposed in CFI may not be feasible for some PHY types, request further study on feasibility from PHY vendors)
- Mechanism for Standby and Rapid Wake-On from standby state
  - Include 0 Mbps speed or Link IDLE state, Wake up signal
  - Power savings for CE/mobile and Rapid Wake-On for Servers
- Compatible with power saving modes at same speed
  - For example: Lower power 10G, based on Link length or Link quality
  - Control protocol should have ability to exchange other parameters
- Define fast frame based Protocol for negotiating speed and standby modes selection
  - Protocol to be extensible to include signaling of standby and proxy messaging
- Management registers to read/control PHY Power saving capabilities
  - Standard format for all EEE PHYs

Backup

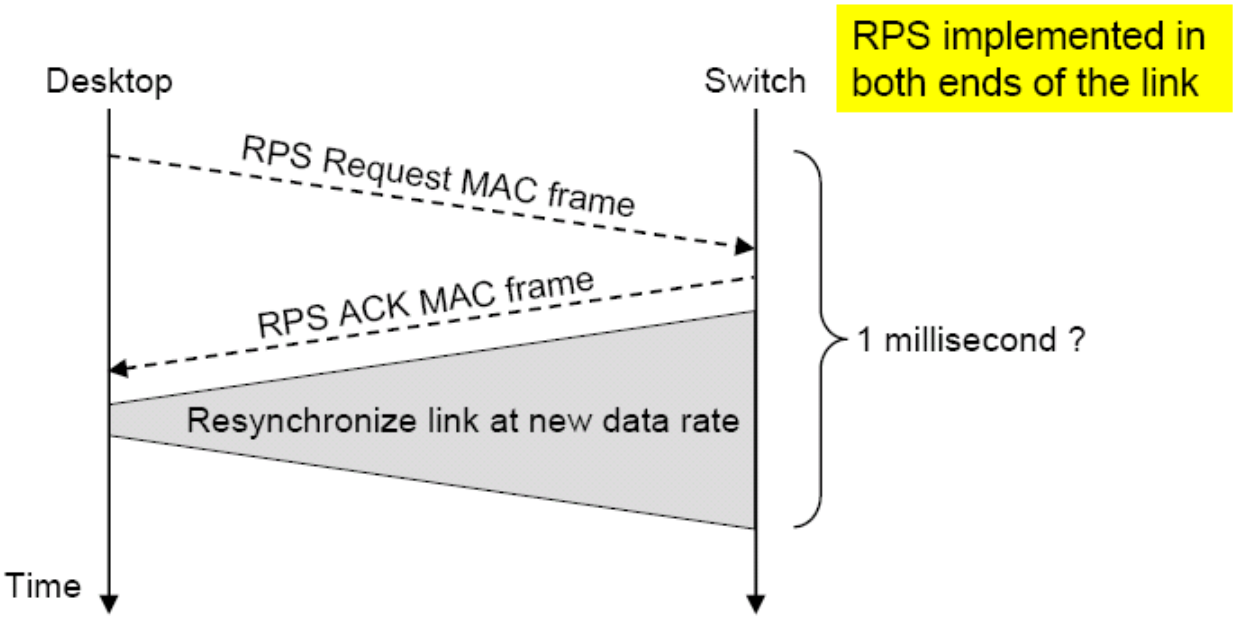
IEEE 802.3 Energy Efficient Ethernet Overview

- IEEE 802.3 tutorial – July 19, 2005 (San Francisco)
  - Reducing Energy consumption for Networked Devices by Bruce Nordman and Ken Christensen
- EEE Call For Interest in Nov 2006 Plenary
  - Approved: EEE Study group formed
- First EEE SG meeting in Jan 2007
  - To discuss Objectives, Goals, and 5 Criteria
- High Level Goals for Energy Efficient Ethernet
  - Reduce power during Low link utilization
  - Be compatible with existing cabling infrastructure
- Benefits of EEE
  - Power savings in Home/Office: Desktop and Mobile clients
  - Power savings in Data Centers: Servers and Switches

A proposal for MAC frame Handshake (CFI)

Rapid PHY Selection

- **One possibility… MAC frame handshake**




intel®
Leap ahead™