Energy Efficient Ethernet

LSI

1000BASE-T Low-Power Idle

Adam Healey, Presenter

LSI Corporation

IEEE P802.3az Task Force Meeting

January 2008

Contributors: Brian Murray
Albert Molina

Supporters

- Brad Booth, AMCC
- Joseph Chou, Realtek
- Robert Hays, Intel
- Adam Healey, LSI
- Sanjay Kasturia, Teranetics
- Brian Murray, LSI
- Gavin Parnaby, Solarflare
- George Zimmerman, Solarflare

Overview

- Introduction
- 1000BASE-T Low-Power Idle Concept
- 1000BASE-T Low-Power Idle Power Savings
- 1000BASE-T Low-Power Idle Signaling
- Summary

Introduction

- Save power by entering a low-power idle state when there is no data to be transmitted (see hays_01_0108)
  - If there is no data to be transmitted, enter 1000BASE-T low-power idle (LP_IDLE)
  - When there is data to be transmitted return to normal 1000BASE-T idle (N_IDLE) and data
- Low-power idle objectives
  - Very low power; close to “No Link” power and much less than 10/100 power
  - Periodically refresh local/remote timing so they remain in Master/Slave lock
  - Periodically refresh all coefficients to allow for fast transition to data
  - Allow a fast transition back to normal 1000BASE-T idle/data; ideally $< 10\ \mu s$
- High level protocol negotiates capability during auto-negotiation and also negotiates agreement enter low-power idle (see dove_01_0108)
  - Allows NIC/switch port to shut down based on no data for a period of time

1000BASE-T Low-Power Idle Concept

- In LP_IDLE, local/remote PHYs cease transmission
  - All transmit and receive circuits can be turned off
  - All adaptive coefficients are saved and stored
  - Timing circuits free run with acquired frequency
- Periodic refresh cycle where local/remote PHYs transmit 1000BASE-T idle
  - Allows all coefficients and timing to be updated
- Local or remote PHY can initiate return to normal operation at any time by transmitting
  - Energy detect circuit is used to detect transmission and turn on all circuits
  - Use a simple scheme to distinguish between a refresh cycle and return to normal operation
- MAC requests PHY to enter or exit LP_IDLE
  - If the remote PHY initiates exit, local PHY immediately signals to the MAC

1000BASE-T Low-Power Idle Power Savings

- In 1000BASE-T the coefficient adaptation can be very slow as the channel is very stable
  - Master/Slave timing, tight clock tolerance and high performance PLLs mean that local and remote clocks take a long time to drift apart
  - Can run for seconds or even minutes without refresh
  - Only a short period of time is required to refresh coefficients
- This allows a large ratio of OFF time to ON time and thus a very significant power saving
- Using round numbers, estimate LP_IDLE power for 1000BASE-T
  - P_ON = 500 mW - full Gigabit data power (90 nm estimate)
  - P_QUIES = 50 mW - quiescent power when Tx/Rx off (leakage+timing circuit)
  - R_PER = 100 ms - refresh period; could be 100ms to 1000 ms
  - R_ON = 1 mS - refresh on time; could be 0.5 ms to 5 ms

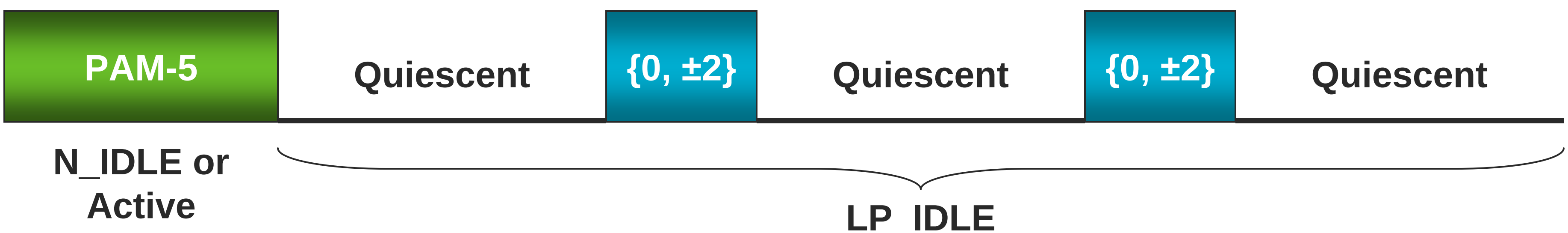
  - P_LPI = P_QUIES + (P_ON − P_QUIES) x (R_ON/R_PER) = **54.5 mW**
    - 50:1 => 59 mW, 500:1 => 51 mW (key is the quiescent power)

1000BASE-T Low-Power Idle Signaling (1/3)

- Keep it simple; re-use existing 1000BASE-T signaling and startup approach
- Use 1000BASE-T Idle, PAM-3 {−2, 0, 2}, as refresh signal
- Master is in control of all timing
  - Consistent with 1000BASE-T Master/Slave timing
  - Master clock is always correct
  - Master initiates entry to LP_IDLE, refresh cycles, and return to normal operation
  - Slave can trigger Master to initiate return to normal operation
- Use PAM-3 {−1, 0, 1} to distinguish LP_IDLE exit signal from refresh signal
  - Also allows exit in the middle of a refresh cycle
- Fall back to auto-negotiation if refresh cycle fails to occur or if link pulses are detected

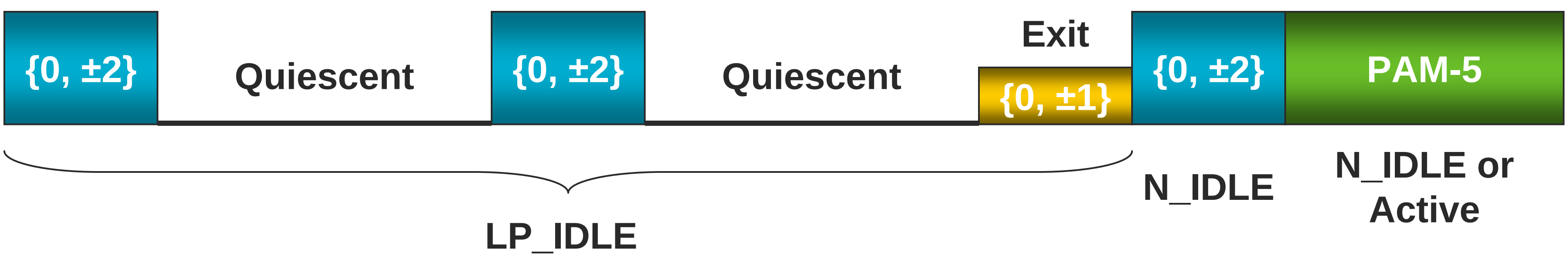
1000BASE-T Low-Power Idle Signaling (2/3)

- Master enters LP_IDLE by ceasing transmission on all pairs
  - Slave acknowledges by also ceasing transmission
- Master starts refresh cycle by transmitting 1000BASE-T Idle
  - Slave acquires correct timing phase
  - Slave responds by transmitting 1000BASE-T Idle
  - After R_ON period has elapsed, Master ceases transmission and Slave acknowledges by also ceasing transmission



1000BASE-T Low-Power Idle Signaling (3/3)

- Masters initiates an exit from LP_IDLE by transmitting {−1, 0, 1}
  - Slave acknowledges with ‘m’ symbols of {0} and ‘m’ symbols of {-1, 0, 1}
  - Slave acquires correct timing phase
  - Slave then starts transmitting 1000BASE-T Idle
  - Master receives {−1, 0, 1}, it responds by transmitting 1000BASE-T Idle
- Slave triggers Master to initiate exit by transmitting {−1, 0, 1}
- Enable data a reasonable time after the Master initiates exit (e.g. 10 microseconds)



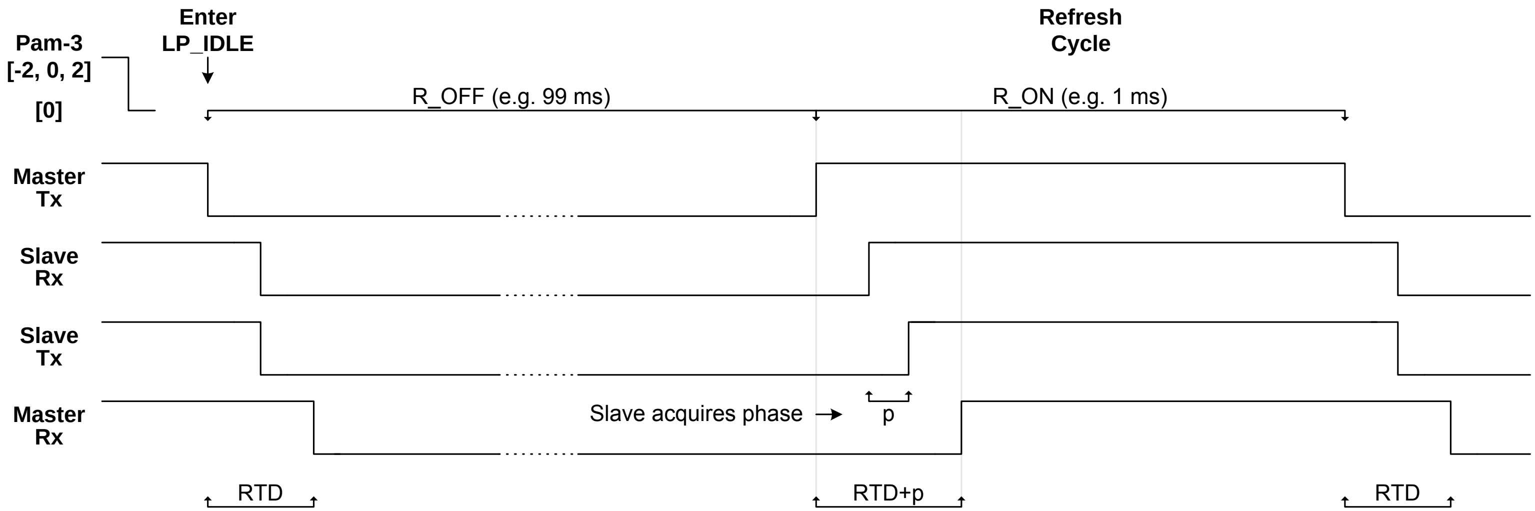
Summary

- 1000BASE-T low-power idle is described
- Allows a fast transition back to normal operation and very low power
- Allows timing to remain in Master/Slave lock
- Allows coefficients to be periodically updated
- Signaling defined for LP_IDLE entry, refresh, and LP_IDLE exit
- Simple protocol that reuses current 1000BASE-T signaling, startup approach

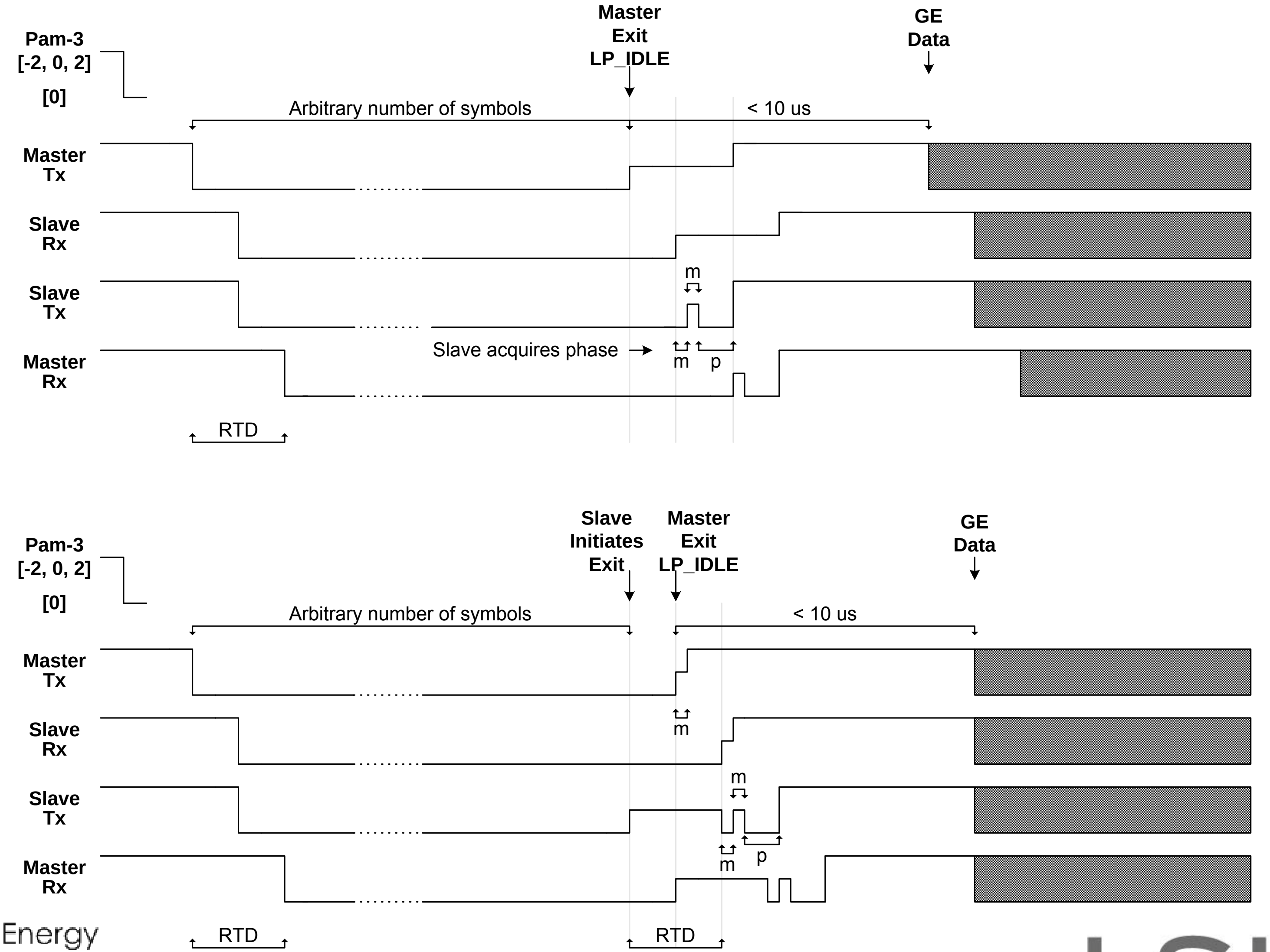
Energy
Efficient
Ethernet

Back-up

Timing Diagrams – Enter LP_IDLE



Timing Diagrams – Exit LP_IDLE



Timing Diagrams – Exit LP_IDLE During Refresh