Proposal of Low-Power Idle 100Base-TX

January 2008
IEEE 802.3az Task Force

Presenter: Joseph Chou
Realtek Semiconductor Corp.

Contributors: Albert Kuo, Dachin Tseng,
Jian-ru Lin, Ken Huang

Supporters

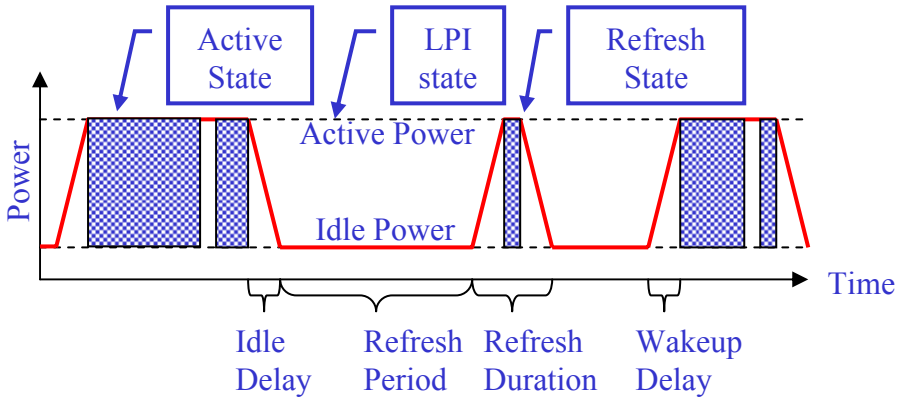
- Robert Hays (Intel)
- Brad Booth (AMCC)
- Adam Healey (LSI)
- Brian Murray (LSI)
- Dan Dove (HP)

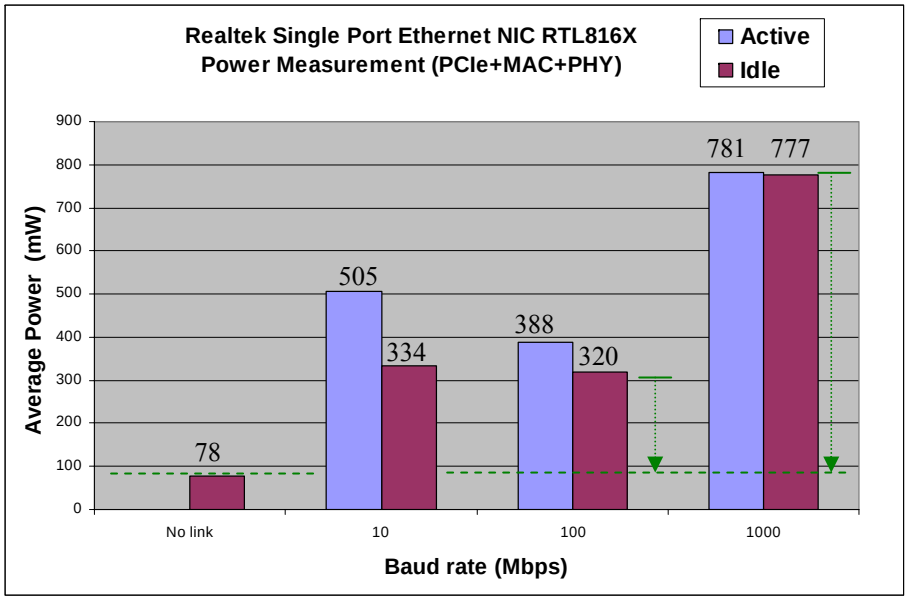
Outlines

- Low-Power Idle (LPI) Mode Features
- Transition Flow Diagrams
  - Transition to Idle State
  - Transition to Wakeup State
  - Refresh Signal
- Transition Timing Diagram
- Simulation Result of Feasibility
- Implementation Considerations
- Single Ethernet PHY Block Diagram
- Estimation of Power Consumptions
- LPI Transition State Diagrams
- Further Discussions
- 802.3 100Base-TX Standard Modifications
  - Auto-negotiation Capability pages & State diagrams
  - Additional 4B/5B (for 100Base-TX) Codeword used
  - PCS State diagrams

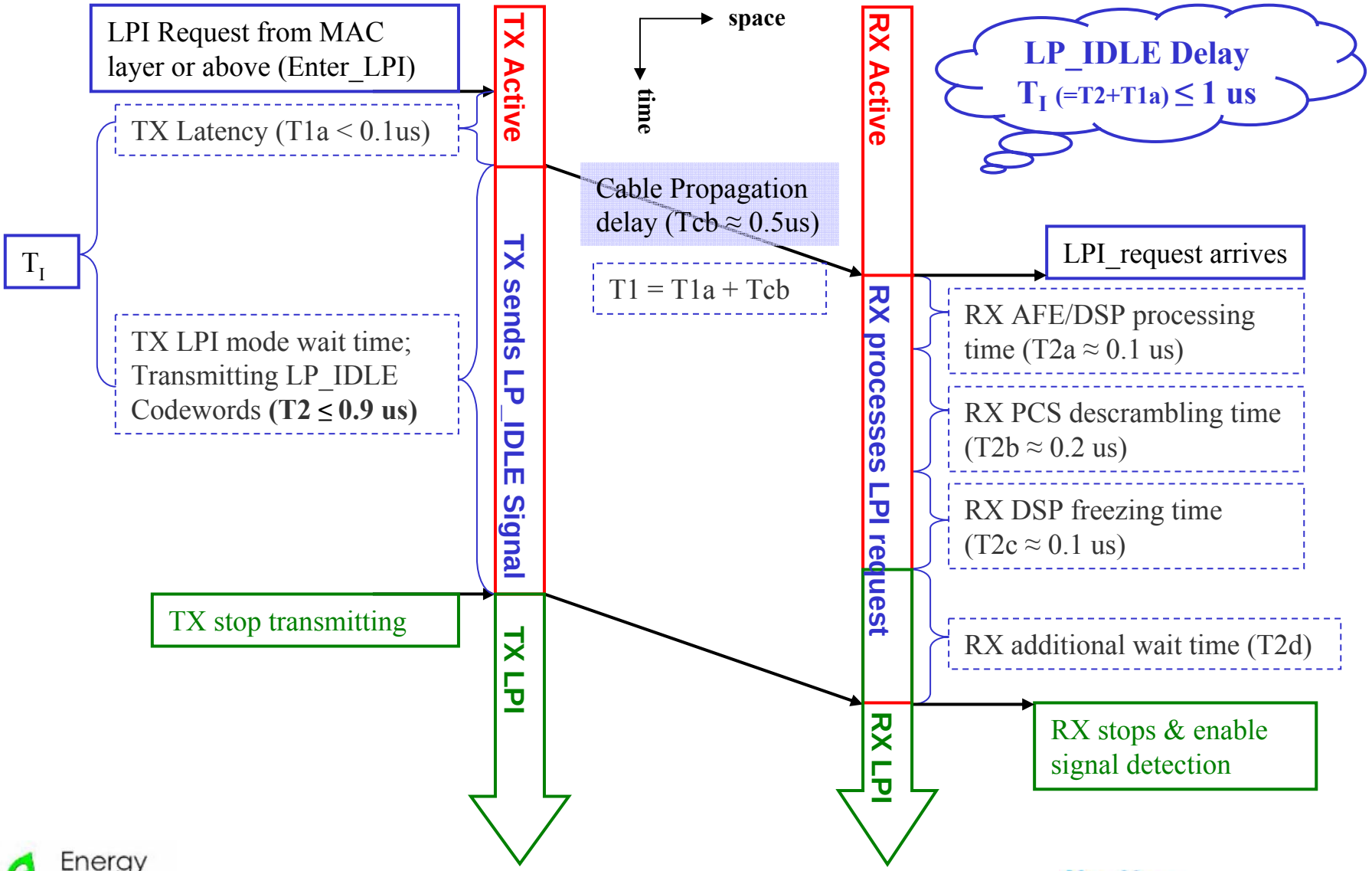
Low-Power Idle (LPI) Mode Features

- Low-Power Idle (LP_IDLE or LPI) mode is a quite line Low Power Idle State that consumes minimum power
- LPI mode allows Asymmetrical operation
- LPI mode may have periodic Refresh (R_DILE) state to allow the maintenance of
  - clock synchronization
  - coefficients of channel equalization & various cancellers
- LPI Request Signal (LP_IDLE Signal) is:
  - A normal carrier with newly defined idle codeword
- LPI Refresh Signal (R_IDLE Signal) is:
  - A normal carrier with newly defined idle codeword (same with LP_IDLE Signal)
- LPI Wakeup Signal is:
  - A normal carrier with Normal IDLE (N_IDLE) codeword
- LPI mode is initiated and terminated from the layer on or above MAC
- Support different level of LPI modes on system level
- Only one LPI mode is supported in PHY

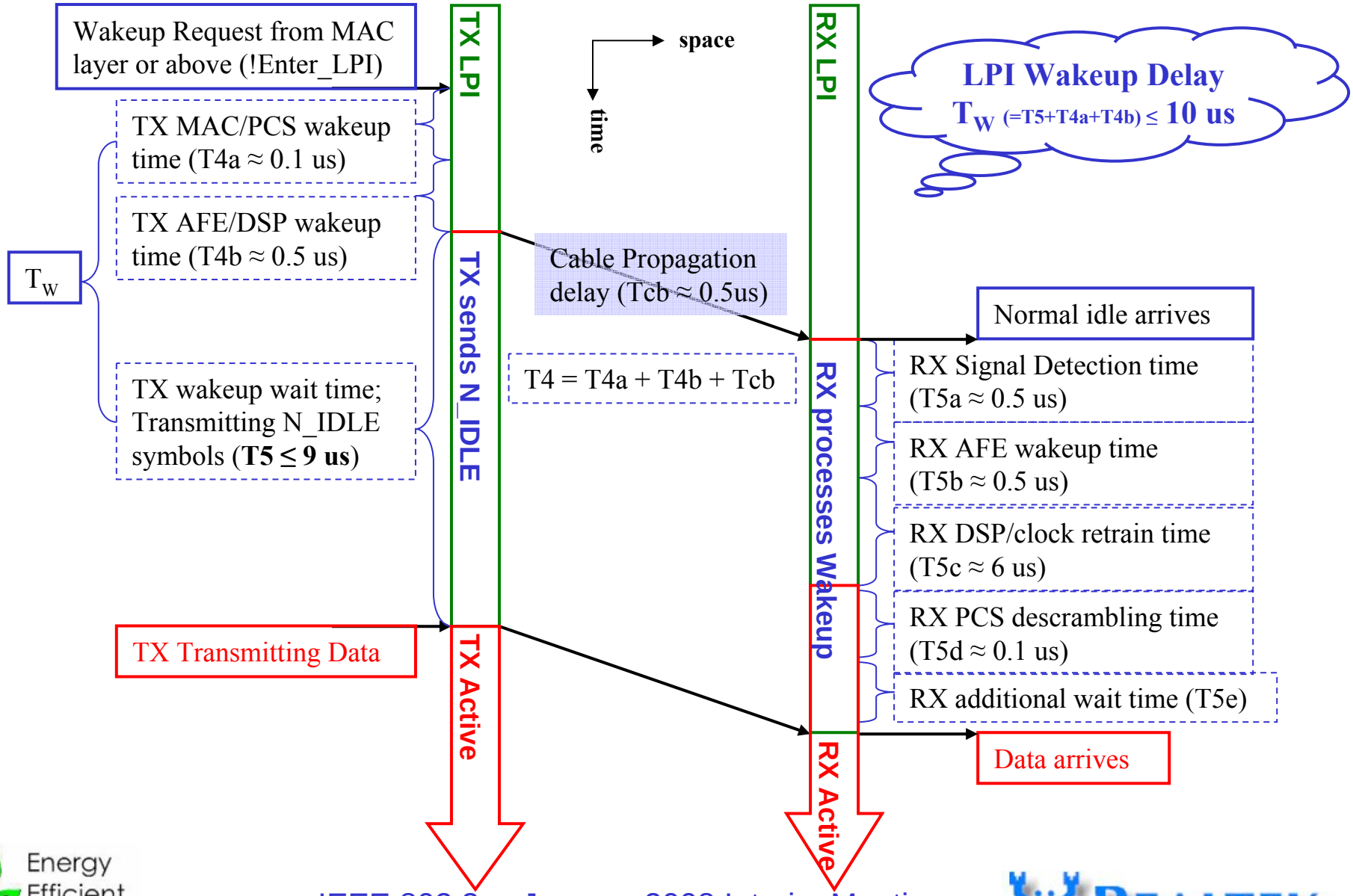




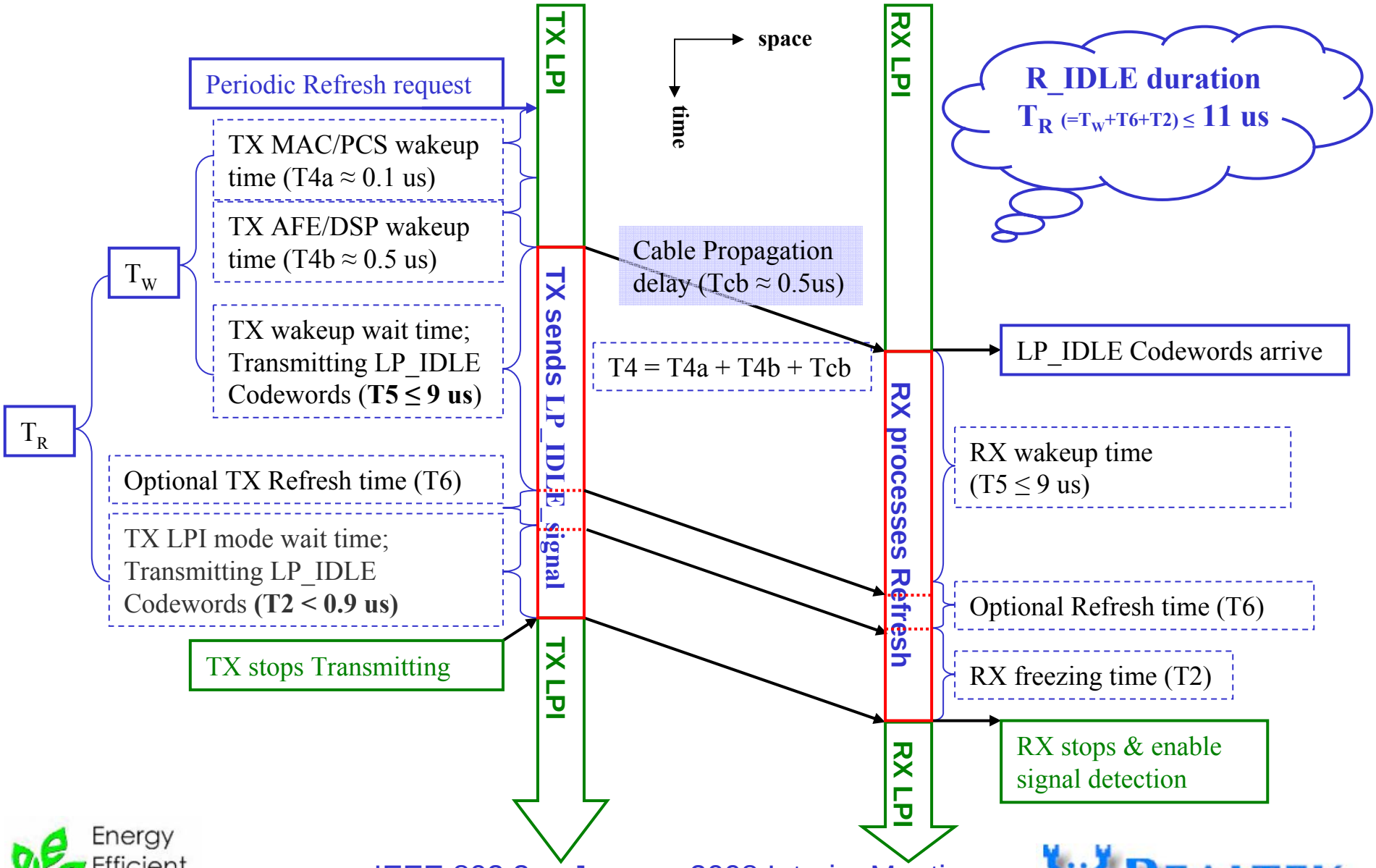
LPI Transition Diagram (Idle)



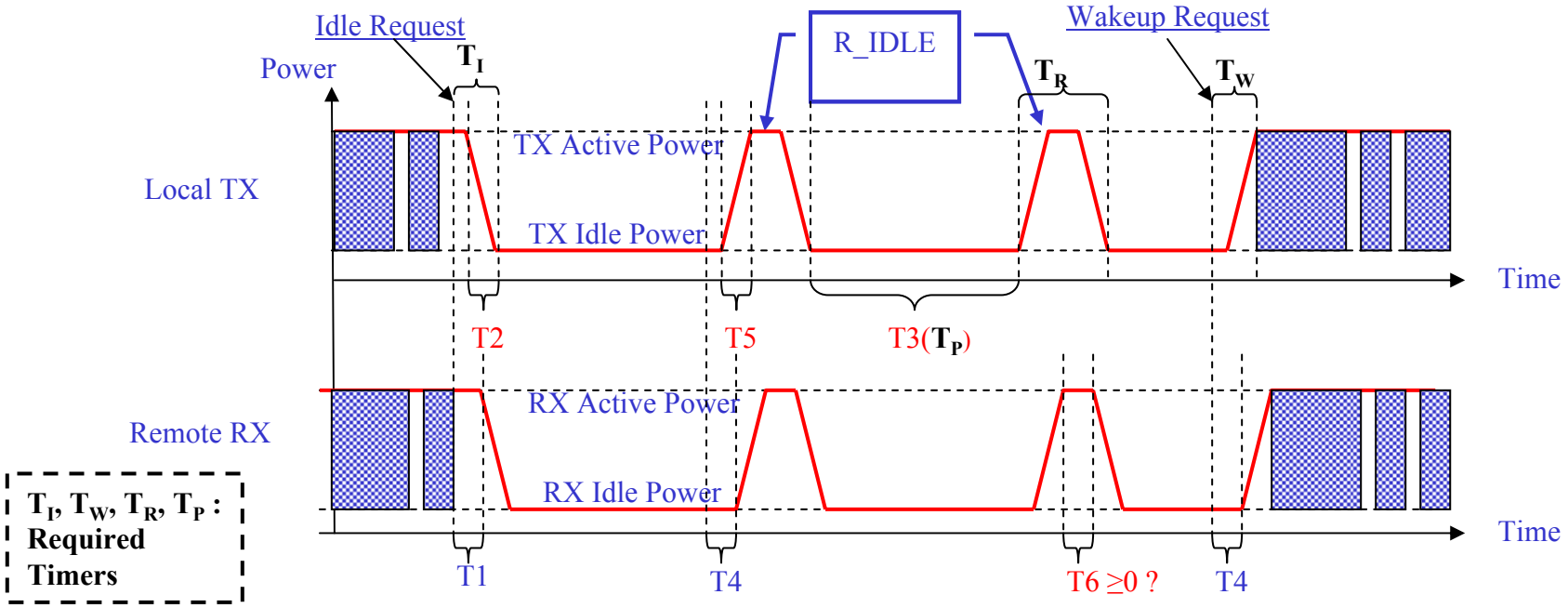
LPI Transition Diagram (Wakeup)



LPI Transition Diagram (Refresh)



LPI Transition Timing Diagram



T1：LPI request latency = TX idle latency + Cable Propagation delay < 1 us

T2： LP Idle wait time = Time to allow remote RX to freeze before turning TX off < **1 us**

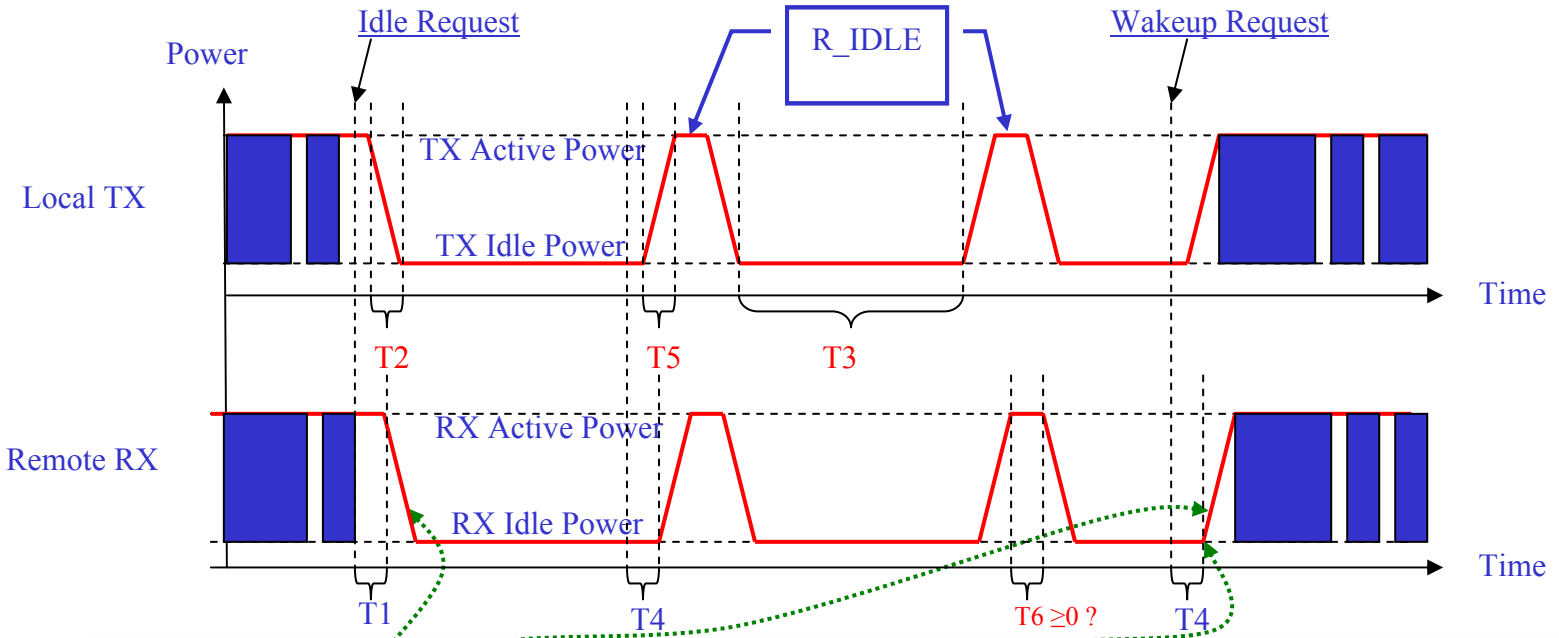
T3： LP Refresh Period = Inter sync-frame time during LPI mode. For 100BaseT, **100 ms** is very safe based on the lab measurement. (**1 hour** is ok to maintain the 100BaseT coefficients and clock in sync)

T4：LPI wakeup request latency = TX wakeup latency + Cable Propagation delay < 2 us.

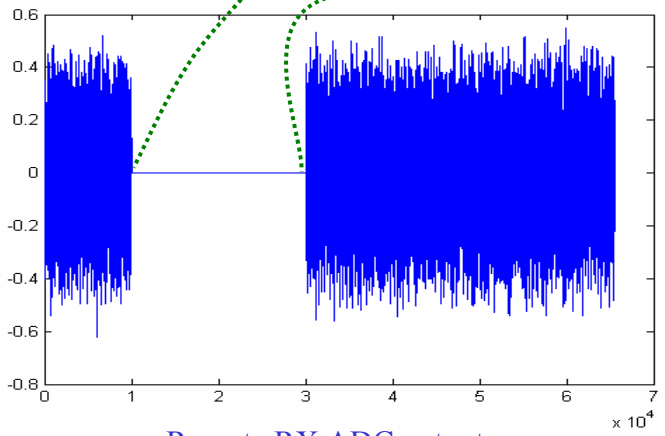
T5：LP Idle Wakeup time = Time to allow remote RX to restore equalizer and to synchronize clock < **9 us**

T6 : Optional Additional Refresh time = Optional additional time in Sync Frame for RX before returning to LPI state; vendor dependent; may be used for MAC control frame ≥ 0 us

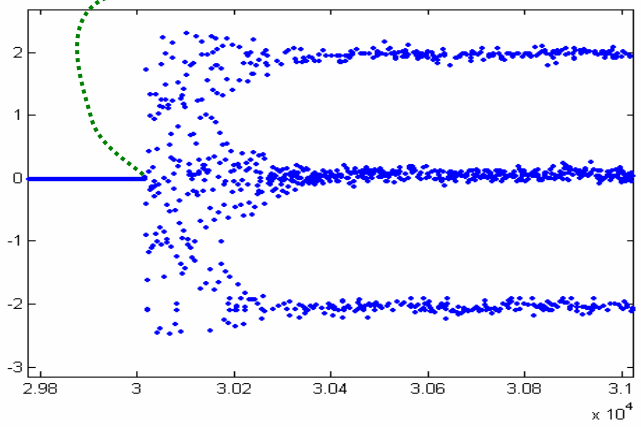
LPI Transition Simulations



Note: Local and Remote Frequency offset = 600ppm



Remote RX ADC output



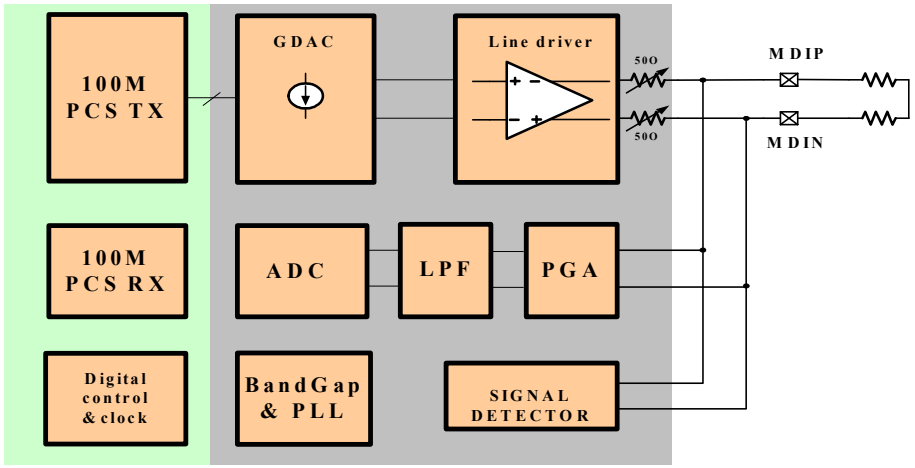
Remote RX Eye Diagram

LPI Mode Implementation Considerations

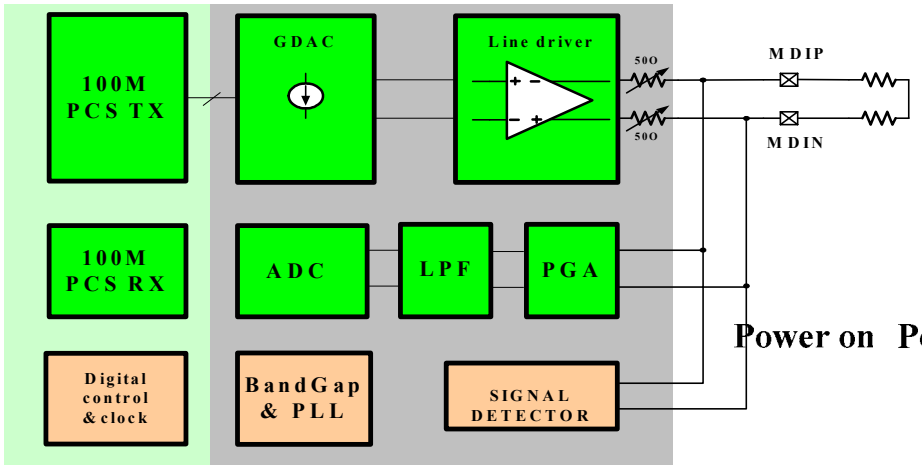
- LPI mode power saving depends on:
  - Idle power consumptions
  - Ratio of R_IDLE duration (T_R) vs. Refresh Period during LP_IDLE (T_P)
  - Multiple system sleep levels (with only one Phy LPI)
- Timers and Parameters
  - Timers for LPI state transitions and failure recovery
  - Transition delays for various sleep level
    - Idle delay (T_I)
    - Active (Wakeup) delay (T_W)
  - R_IDLE duration (T_R) and Refresh Period (T_P)
  - Advertised in new AN capability pages during Auto-Negotiation
- Control Policy and Communications between layers of MAC and PHY
- Decision of Low-Power Idle/Active states between PHY and PHY
  - Initiated from MAC layer or above, but implemented in PHY level
  - Use Special Codeword for LPI request and Refresh Signal
  - Use Normal IDLE codeword to wakeup
- Modifications of Existing 802.3 100Base-TX standard State Diagrams
  - Auto-Negotiation State Diagram and AN registers/pages
  - Modifications of LPI State Diagram for each Phy type
  - Modifications of State Diagram of TX and RX

Single Ethernet PHY Block Diagram

Fully Active (P_full)

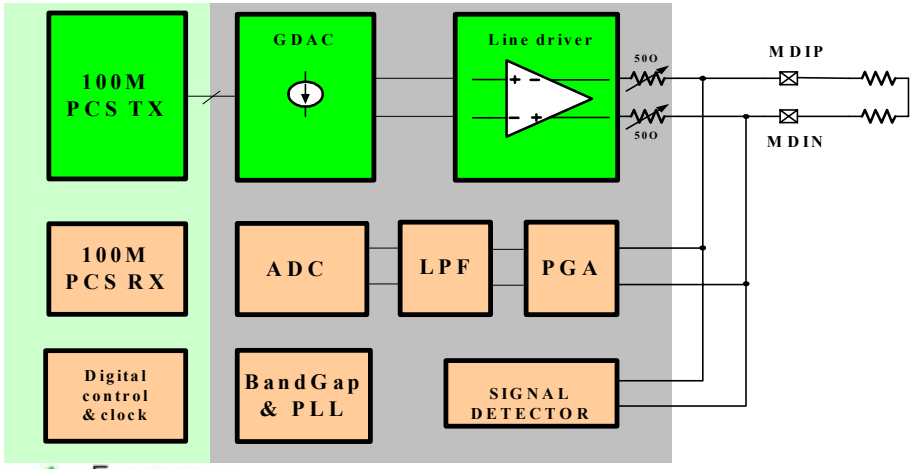


Fully Idle (P_LPI)

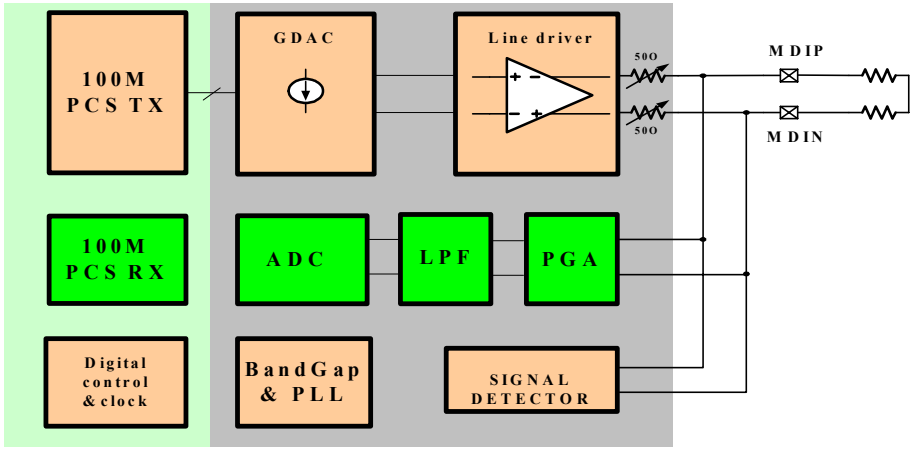


Power on Power off

TX Idle (P_LPI_TX)



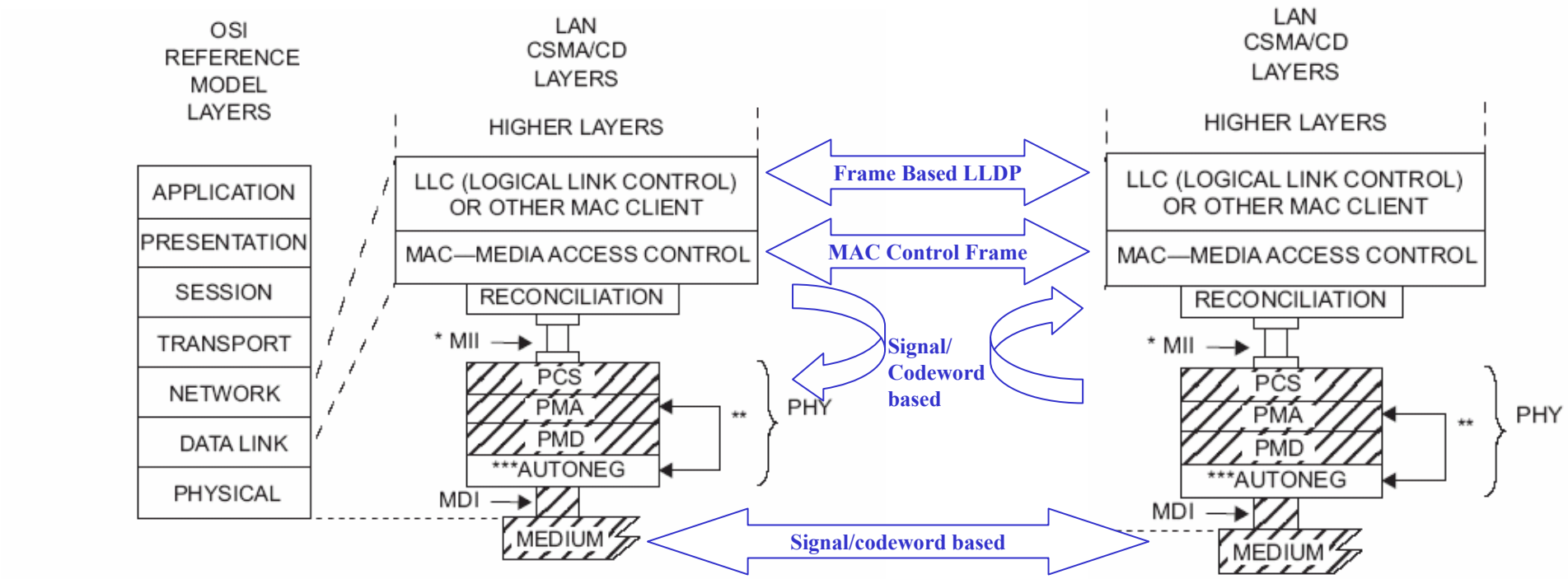
RX Idle (P_LPI_RX)



Power estimations (100Base-TX PHY)

- Power estimation is based on Realtek RTL811x's simulation data.
  - Ptx: TX active power = 64 mW
  - Prx: RX active power = 125 mW
  - Pctl: Control, and clock driver power (w/o MAC, system bus) = 60 mW
- P_full = Pcontrol + Ptx + Prx = 60+64+125
  = **249 mW**
- P_LPI ≒ Pcontrol + (Ptx+Prx)*(T5+T2)/(T3+T5+T2)
  ≒ 60+(64+125)*10us/(10us+100ms)
  ≈ **60 mW** (TX and RX are in LPI)
- P_LPI_TX ≒ Pcontrol + Ptx*(T5+T2)/(T3+T5+T2) + Prx
  ≒ 60+64*10us/(10us+100ms)+125
  ≈ **185 mW** (TX is in LPI)
- P_ LPI_RX ≒ Pcontrol + Ptx + Prx*(T5+T2)/(T3+T5+T2))
  ≒ 60+64+125*10us/(10us+100ms)
  ≈ **124 mW** (RX is in LPI)
- P_MAC(TX) ≒ **29 mW**; P_MAC(RX) ≒ **43 mW**

Control Policy and Communications between layers of MAC and PHY



MDI = MEDIUM DEPENDENT INTERFACE
MII = MEDIA INDEPENDENT INTERFACE

PCS = PHYSICAL CODING SUBLAYER
PMA = PHYSICAL MEDIUM ATTACHMENT
PHY = PHYSICAL LAYER DEVICE
PMD = PHYSICAL MEDIUM DEPENDENT

Negotiation between MAC and PHY

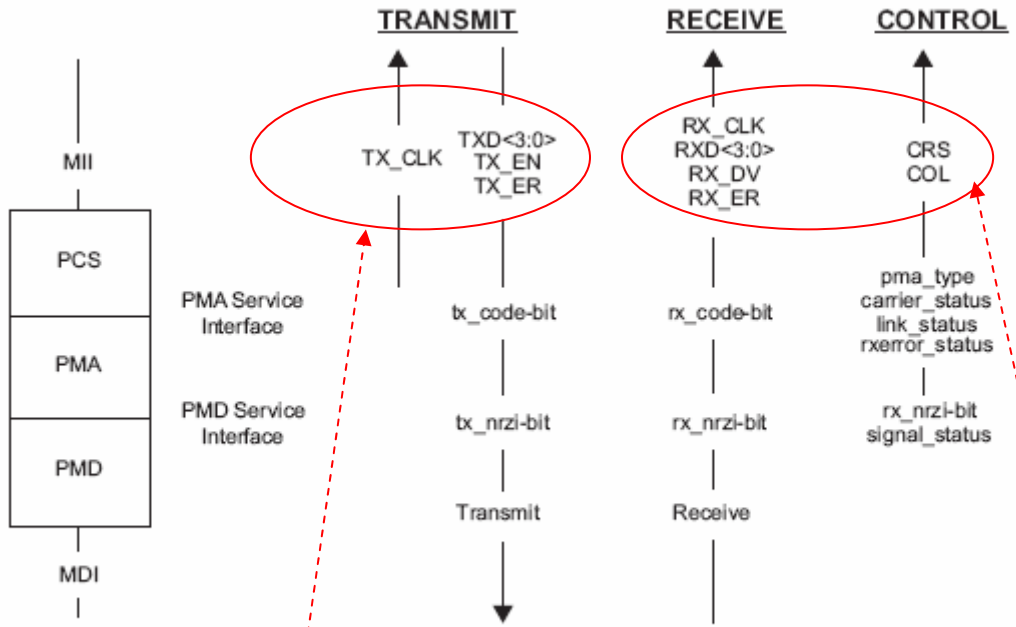


Figure 24–3—Interface mapping

MAC to PHY

TX Signaling
*** Enter_LPI**
RX Signaling
*** Receive_LPI**
*** Receive_NI**

PHY to remote PHY

TX Signaling
*** LP_IDLE Signal**
*** N_IDLE Signal**

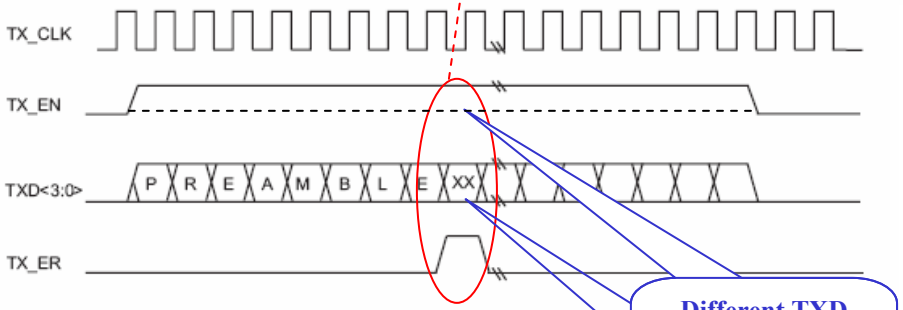


Figure 22–5—Propagating an error

Different TXD value at TX_EN=0 to convey LPI state

Figure 22–8 shows the behavior of RX_ER, RX_DV and RXD<3:0> during a False Carrier indication.

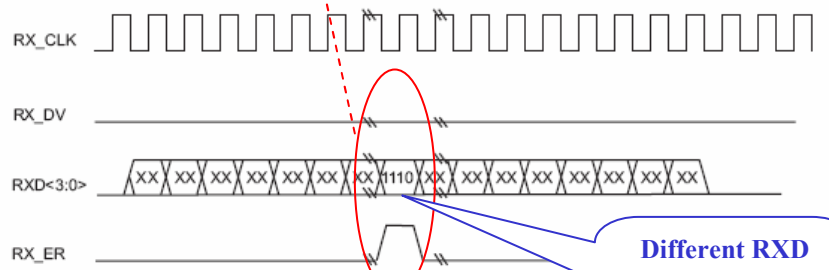


Figure 22–8—False Carrier indication

Different RXD value at RX_ER=1 to convey LPI state

Negotiation between PHYs - Idle/Wakeup Request Codeword

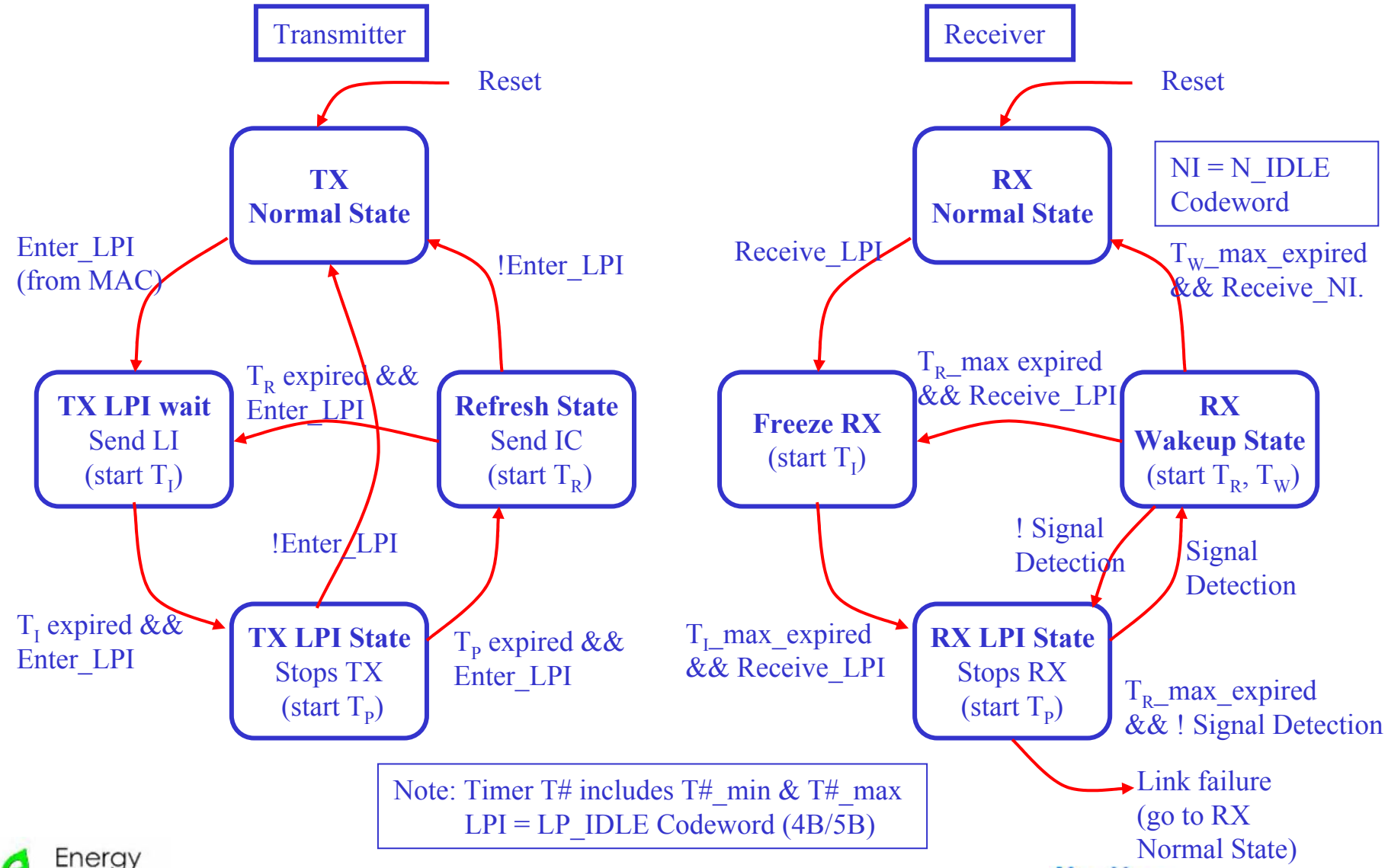
Table 24–1—4B/5B code-groups

	PCS code-group [4:0] 4 3 2 1 0	Name	MII (TXD/RXD) <3:0> 3 2 1 0	Interpretatio
	1 1 1 1 1	I	undefined	IDLE; used as inter-stream fill code
INVALID	0 0 1 0 0	H	Undefined	Transmit Error; used to force signaling errors
INVALID	0 0 0 0 0	V	Undefined	Invalid code
INVALID	0 0 0 0 1	V	Undefined	Invalid code
INVALID	0 0 0 1 0	V	Undefined	Invalid code
INVALID	0 0 0 1 1	V	Undefined	Invalid code
INVALID	0 0 1 0 1	V	Undefined	Invalid code
INVALID	0 0 1 1 0	V	Undefined	Invalid code
INVALID	0 1 0 0 0	V	Undefined	Invalid code
INVALID	0 1 1 0 0	V	Undefined	Invalid code
INVALID	1 0 0 0 0	V	Undefined	Invalid code
INVALID	1 1 0 0 1	V	Undefined	Invalid code

Normal IDLE codeword to wakeup the Remote PHY

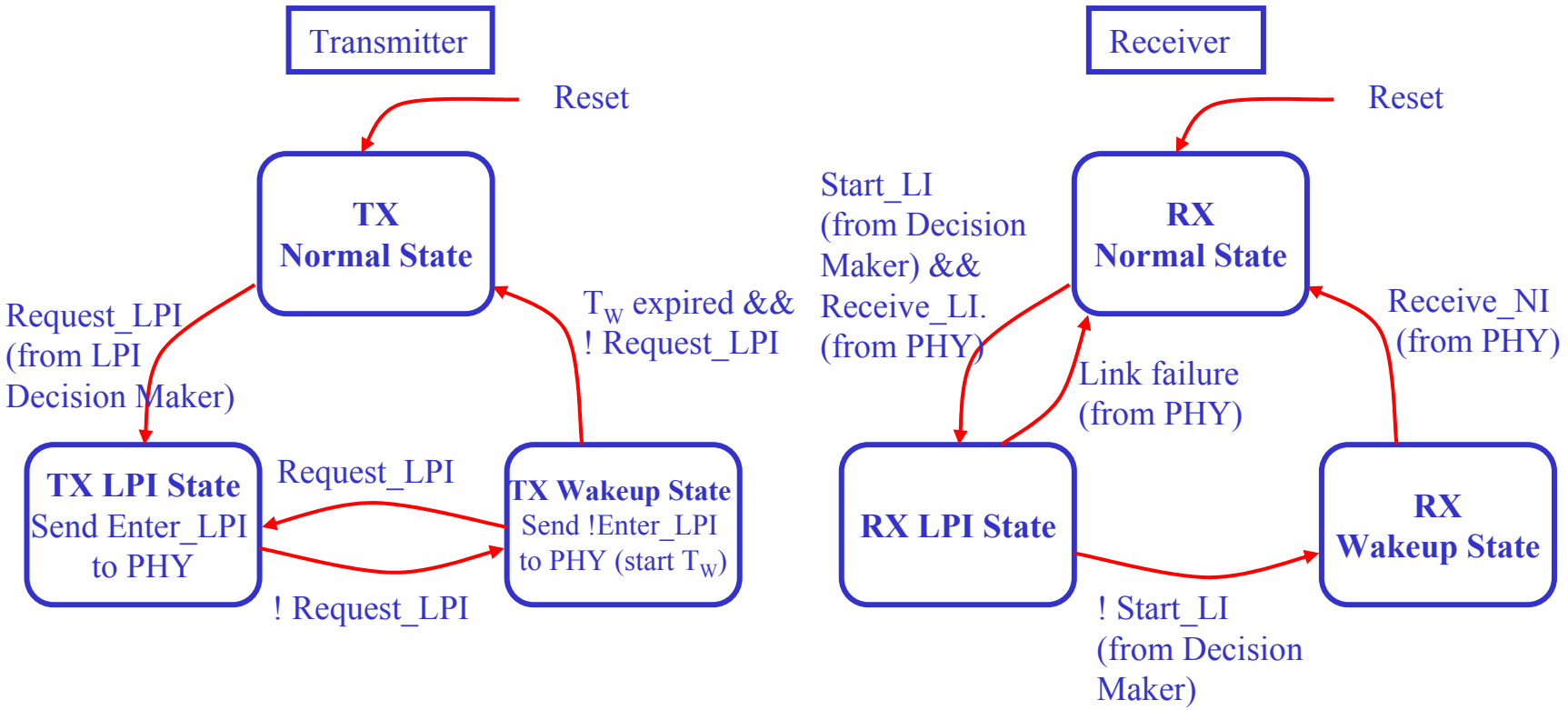
Unused codeword for LP_IDLE Signal to Request LPI mode or Refresh the Remote PHY

LPI Transition State Diagrams (PHY)



Note: Timer T# includes T#_min & T#_max
LPI = LP_IDLE Codeword (4B/5B)

LPI Transition State Diagrams (MAC)



Further Discussions

- Control Policy Maker
  - How many level of LPI modes? (at least two: MAC and PHY)
  - Who makes the LPI decision? MAC or upper layers?
  - Where is the Store and Forward Buffer located during LPI transition, PCS, MAC or system ? (depending on the LPI State Machine)
- Idle mode negotiation
  - Protocol or Signaling used between MACs and PHYs
- Is Asymmetrical LPI feasible for all PHYs?
- How to determine the range of Refresh Period? (measurements?)
- No Link (line disconnect) Detection during LPI mode.
- Criteria of Signal Detection
- Low Power state of System level (ex. PCIe)
  - It takes different order of time to power down and power up system bus depending on the system sleep level
  - It should not be included in the standard of EEE of Ethernet MAC/PHY.

Questions?

Thank you

Backup

100Base-TX Standard Modification - Next Page for EEE capability

- Add new next page to exchange EEE capability in Auto-negotiation.
  - 100M/10M: After base page
  - 1G: After UP2

PAGE 3 (Message Next Page) for LPI		
M10:M0	X	
PAGE 4 (Unformatted Next Page) for LPI		
U0	LPI_Enable	

100Base-TX Standard Modification

- Add state

EEE_MP_TX

EEE_UP_TX

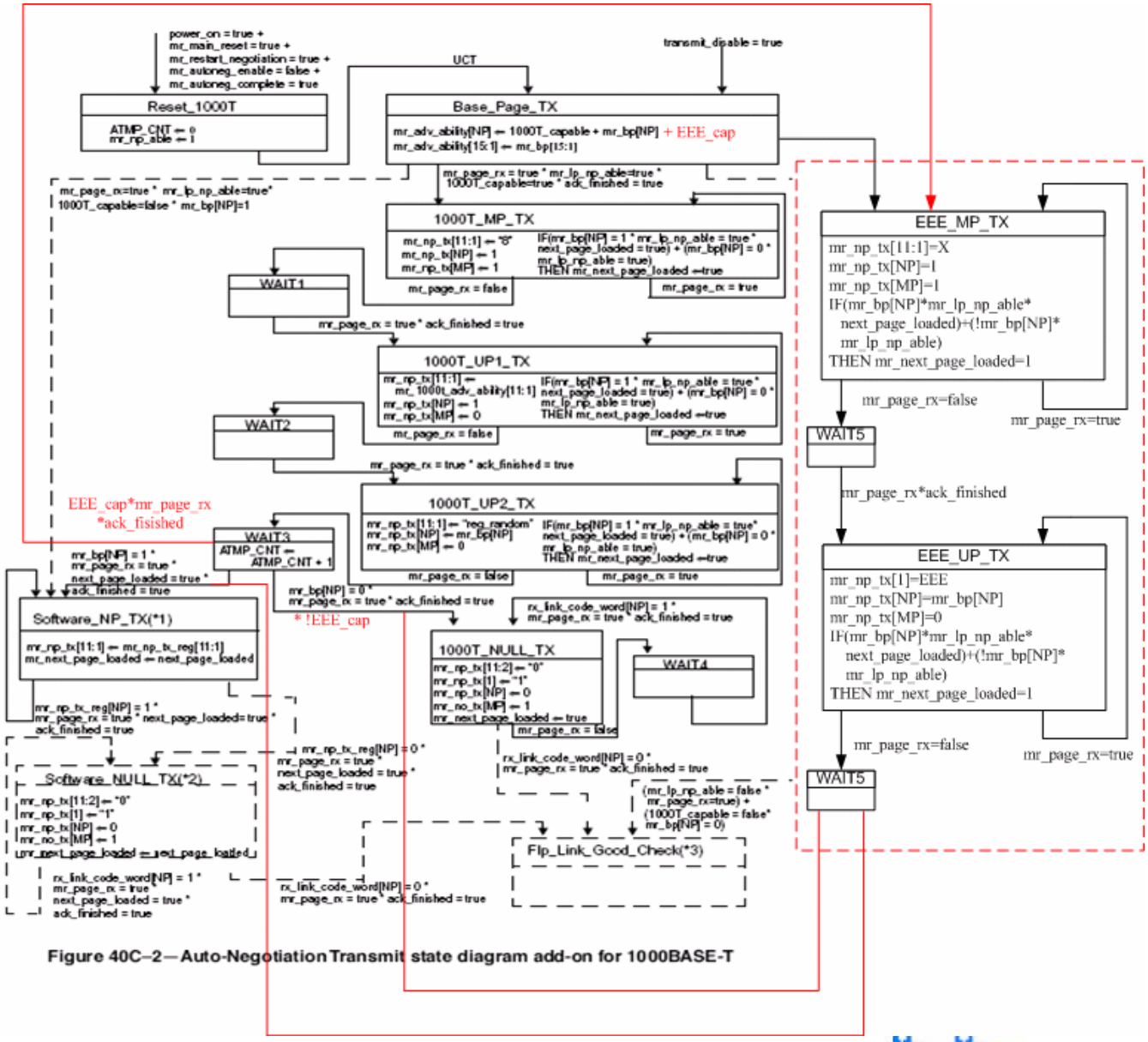


Figure 40C–2—Auto-Negotiation Transmit state diagram add-on for 1000BASE-T

100Base-TX Standard Modification

- Add state

EEE_MP_RX

EEE_UP_RX

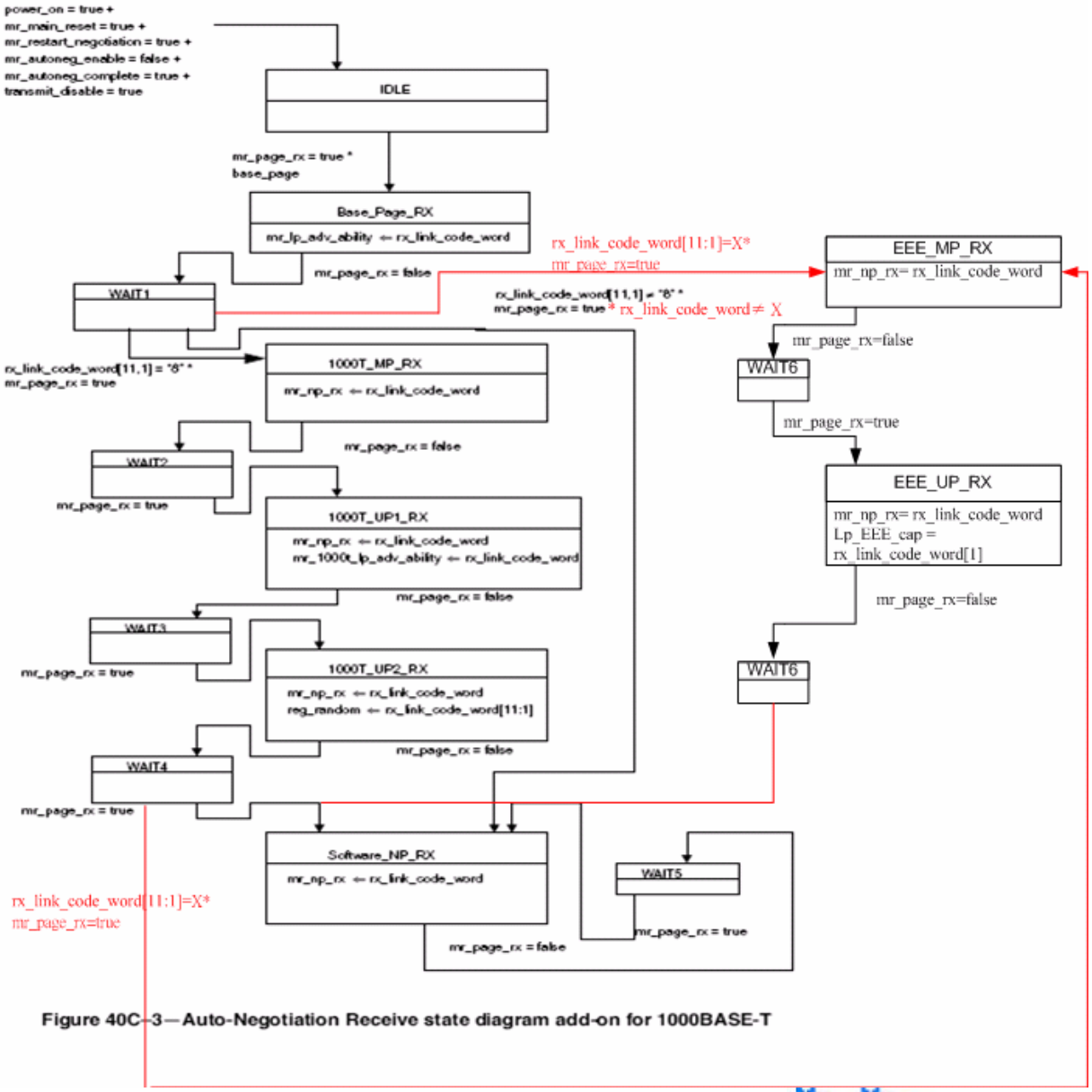


Figure 40C–3—Auto-Negotiation Receive state diagram add-on for 1000BASE-T

100Base-TX Standard Modification
PCS TX FSM

- Add

TX I1 N times state

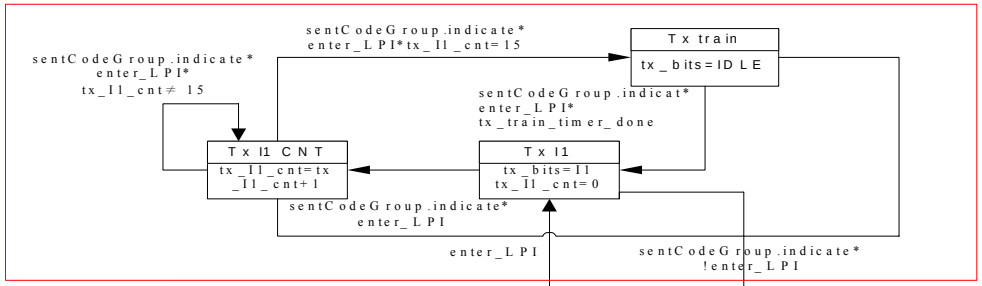
TX train state

(IDLE +I1)

- tx_I1_cnt =25

  => 40ns*25=1us

- tx_train_timer= 9us



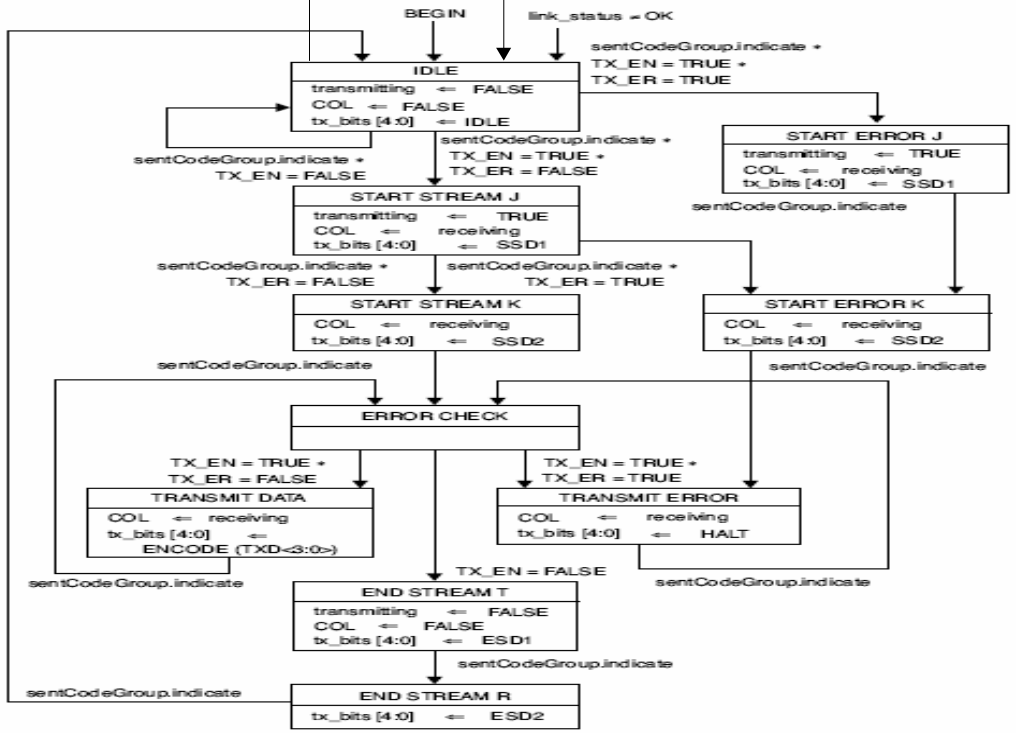


Figure 24–8 —Transmit state diagram

100Base-TX Standard Modification
PCS RX FSM

- Add
  Rx I1 state

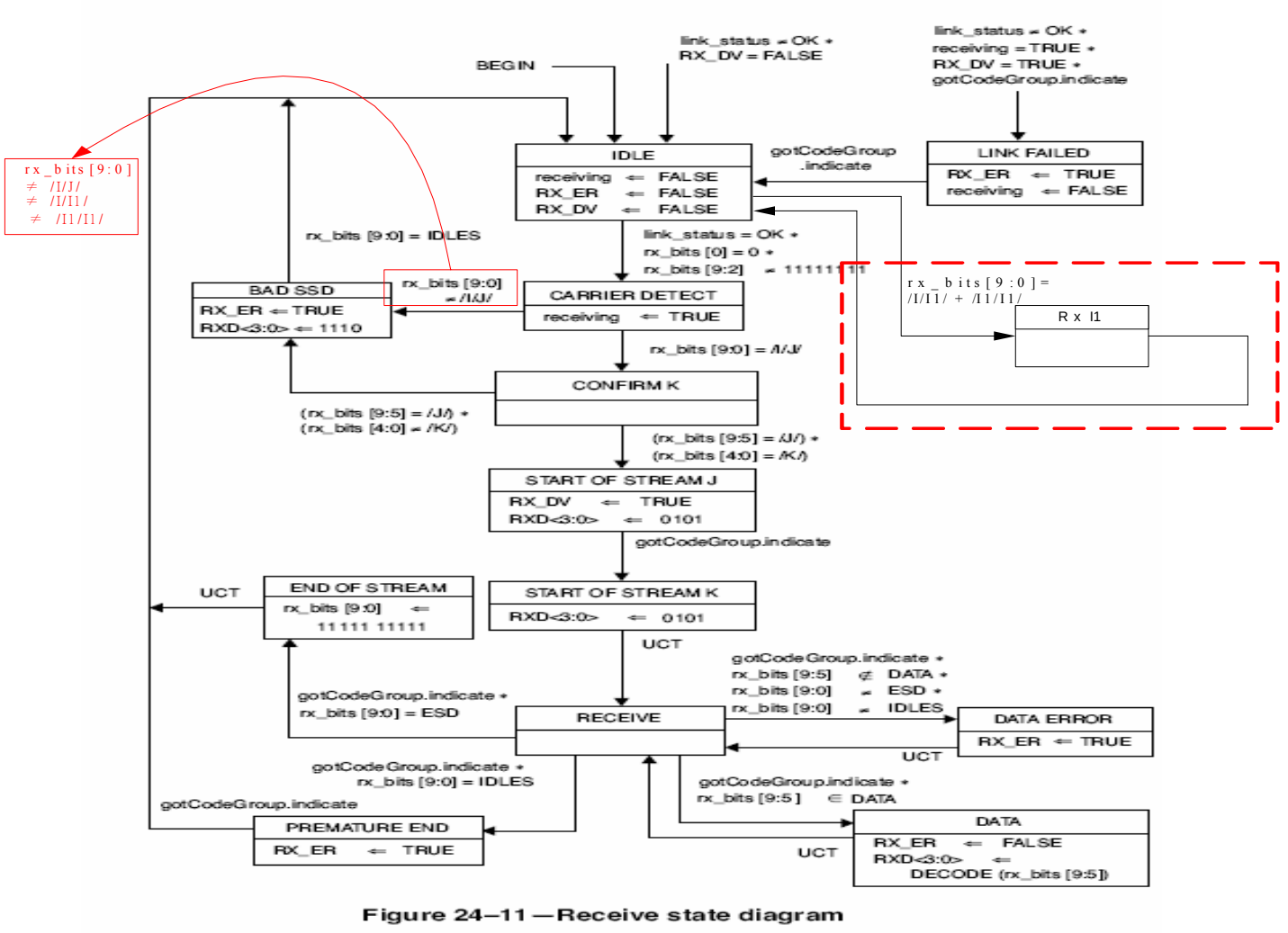
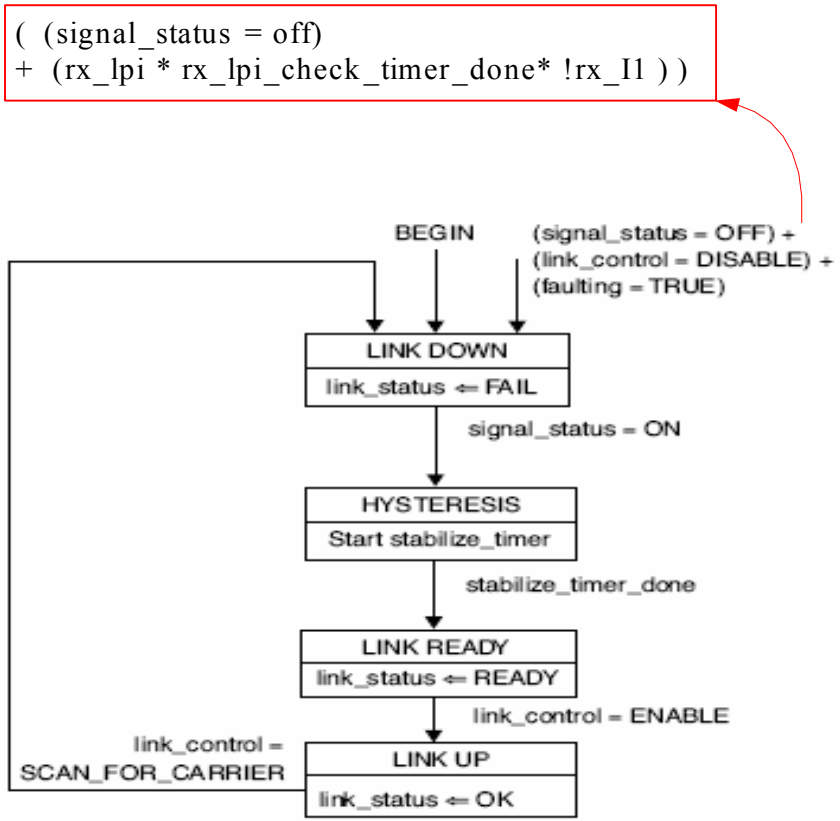


Figure 24–11—Receive state diagram

100Base-TX Standard Modification
Link Monitor FSM

- rx_lps_check_timer > T3+ T4
- When RX is in LPI, if RX does not receive I1

  in rx_lpi_check_timer, it will be re-link.



NOTE—The variables link_control and link_status are designated as link_control_[TX] and link_status_[TX], respectively, by the Auto-Negotiation Arbitration state diagram (Figure 28–16).

Figure 24–15—Link Monitor state diagram