1000BASE-T Low-Power Idle

Presented by Adam Healey
Contributors: Noel Butler, Niall Fitzgerald, Brian Murray, Oisin O'Cuanachain, Jacobo Riesco

Energy Efficient Ethernet

IEEE P802.3az Task Force
Munich, Germany
May 13, 2008

Supporters

- Hugh Barrass, Cisco
- Jim Barnette, Vitesse
- Brad Booth, AMCC
- Mandeep Chadha, Vitesse
- Joseph Chou, Realtek
- Robert Hays, Intel
- Adam Healey, LSI
- Brian Murray, LSI
- Dimitry Taich, Teranetics
- Mario Träber, Infineon
- Aviad Wertheimer, Intel

Scope and purpose

- Illustrate changes to IEEE 802.3 Clause 40 required to implement concepts described in healey_01_0308
- Provide detailed timing diagrams illustrating the operation of the proposed state diagrams
- Estimate the worst-case time to wake up from low-power idle

GMII signaling: Transmit

- Table 35-1

TX_EN	TX_ER	TXD[7:0]	Description	PLS_DATA.request parameter
0	0	00 through FF	Normal inter-frame	TRANSMIT_COMPLETE
0	1	00	Reserved	—
0	**1**	**01**	**EEE low power idle**	**LP_IDLE**
0	1	02 through 0E	Reserved	—
0	1	0F	Carrier extend	EXTEND (eight bits)
…	…	…	…	…

GMII signaling: Receive

- Table 35-2

RX_DV	RX_ER	RXD[7:0]	Description	PLS_DATA.indication parameter
0	0	00 through FF	Normal inter-frame	No applicable parameter
0	1	00	Normal inter-frame	No applicable parameter
0	**1**	**01**	**EEE low power idle**	**LP_IDLE**
0	1	02 through 0D	Reserved	—
0	1	0E	False carrier indication	No applicable parameter
…	…	…	…	…

PCS encoding of LP_IDLE

- **40.3.1.3.4 Generation of bits Sd_n[8:0]**

  The bit Sd_n[3] is used to scramble the GMII data bit TXD_n[3] during data mode and to encode loc_lpi_req otherwise. It is defined as

$$Sd_n[3] = \begin{cases} Sc_n[3] \wedge TXD_n[3], & if\ (tx\_enable_{n-2} = 1) \\ Sc_n[3] \wedge 1, & else\ if\ (loc\_lpi\_req = TRUE) \\ Sc_n[3], & else \end{cases}$$

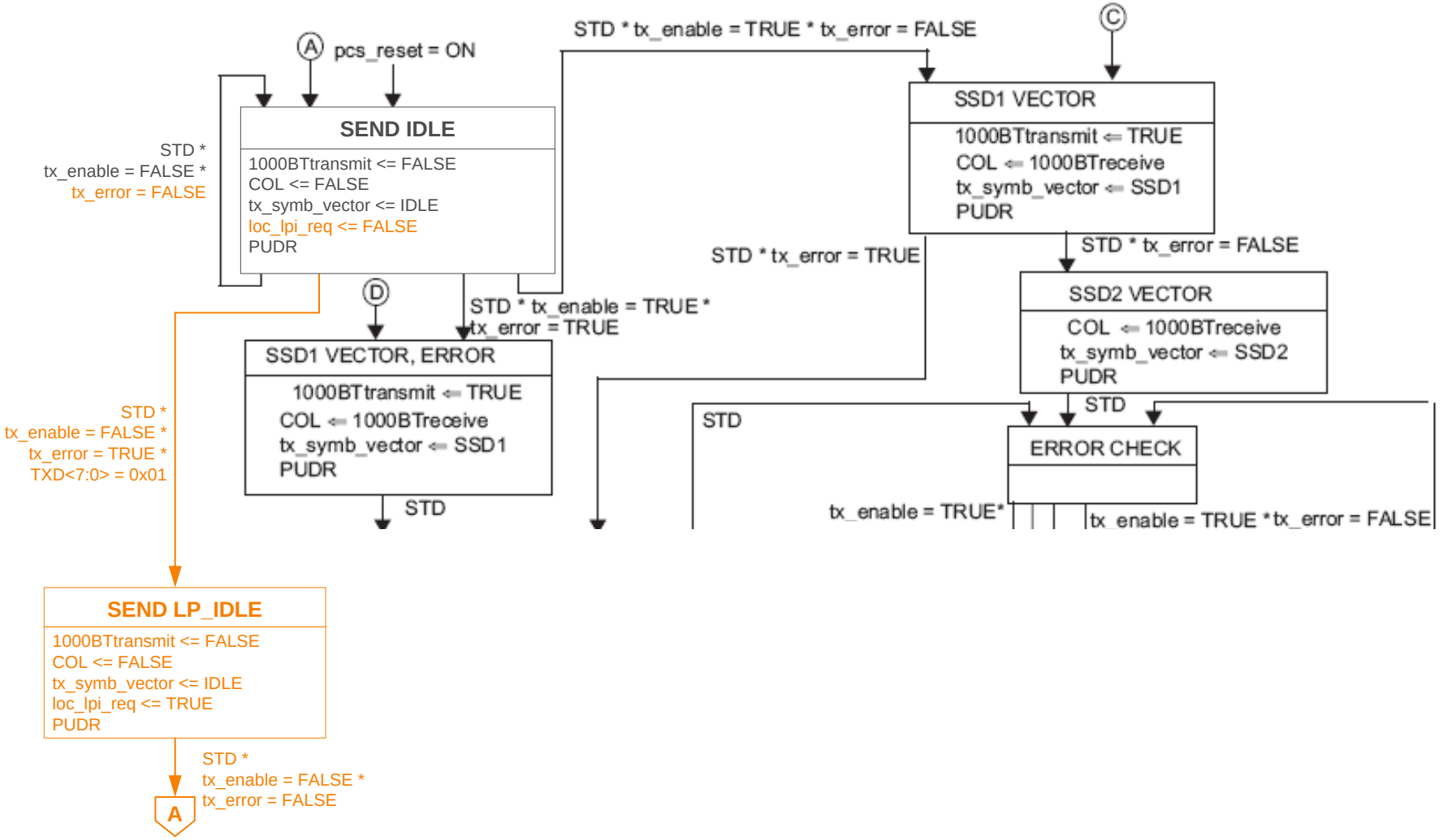
New variables

- lpi_mode
  - Set to ON when the PHY is in low-power idle mode and set to OFF otherwise.
- loc_lpi_req
  - Set to TRUE when the transmit MAC indicates it is going to sleep, and requesting low-power idle mode, via the GMII, and set to FALSE otherwise.
- rem_lpi_req
  - This is the link partner's version of loc_lpi_req.
- pma_rx_indicate
  - Indicate that transmission from the link partner has been received at the MDI. This signal will be set to FALSE when the PHY detects that the link partner is sending zero (the link partner tx_mode = SEND_Z) and will be set to TRUE otherwise.

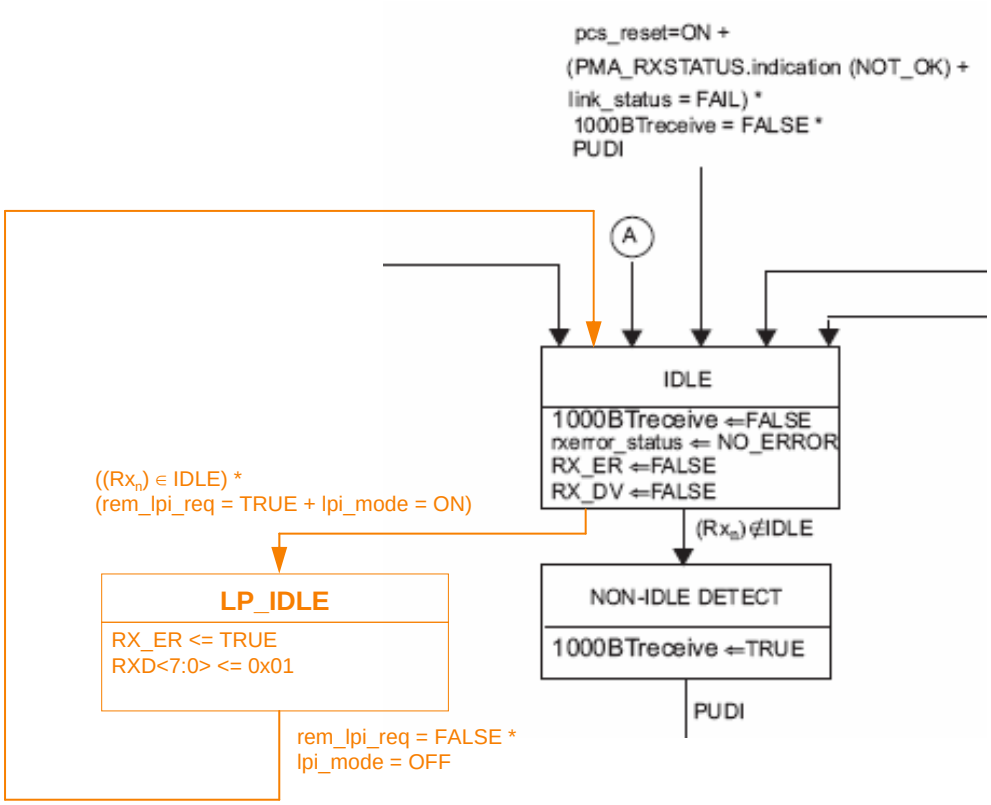
New timers

- lpi_timer
  - Generic timer function used for the various durations t_s, t_q, t_w, and t_u.
- lpi_wt_timer
  - Time the PHY transmits before it transitions to the WAKE_SILENT state following initiation of wake.
- lpi_wz_timer
  - Time the MASTER PHY remains in the WAKE_SILENT state (transmitting ZERO) before transitioning to the WAKE_TRAINING state.
- lpi_min_timer
  - Minimum time, t_m, the PHY waits in the WAKE_TRAINING state before going back into ACTIVE mode.

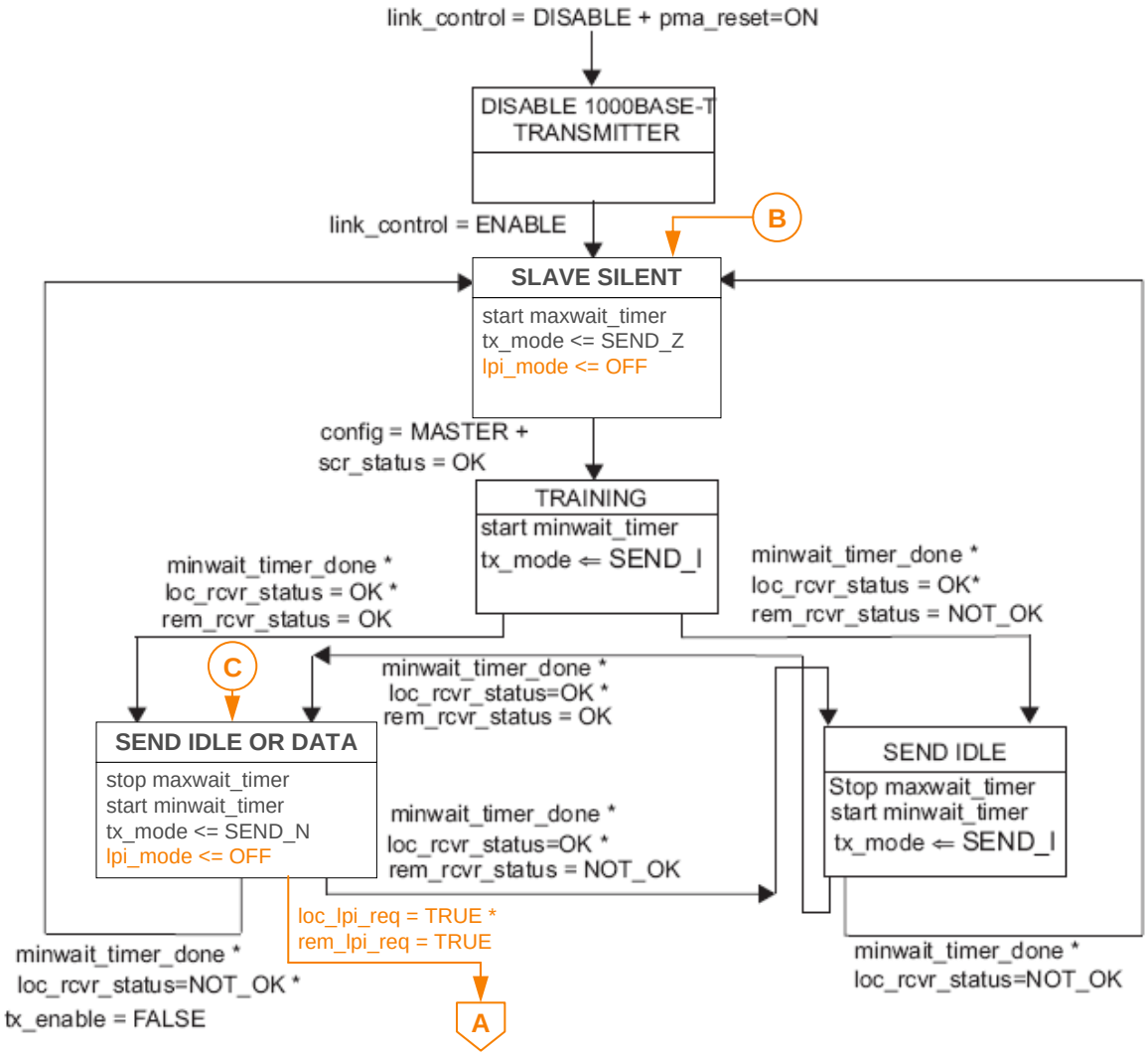
PCS Transmit state diagram, Figure 40–9



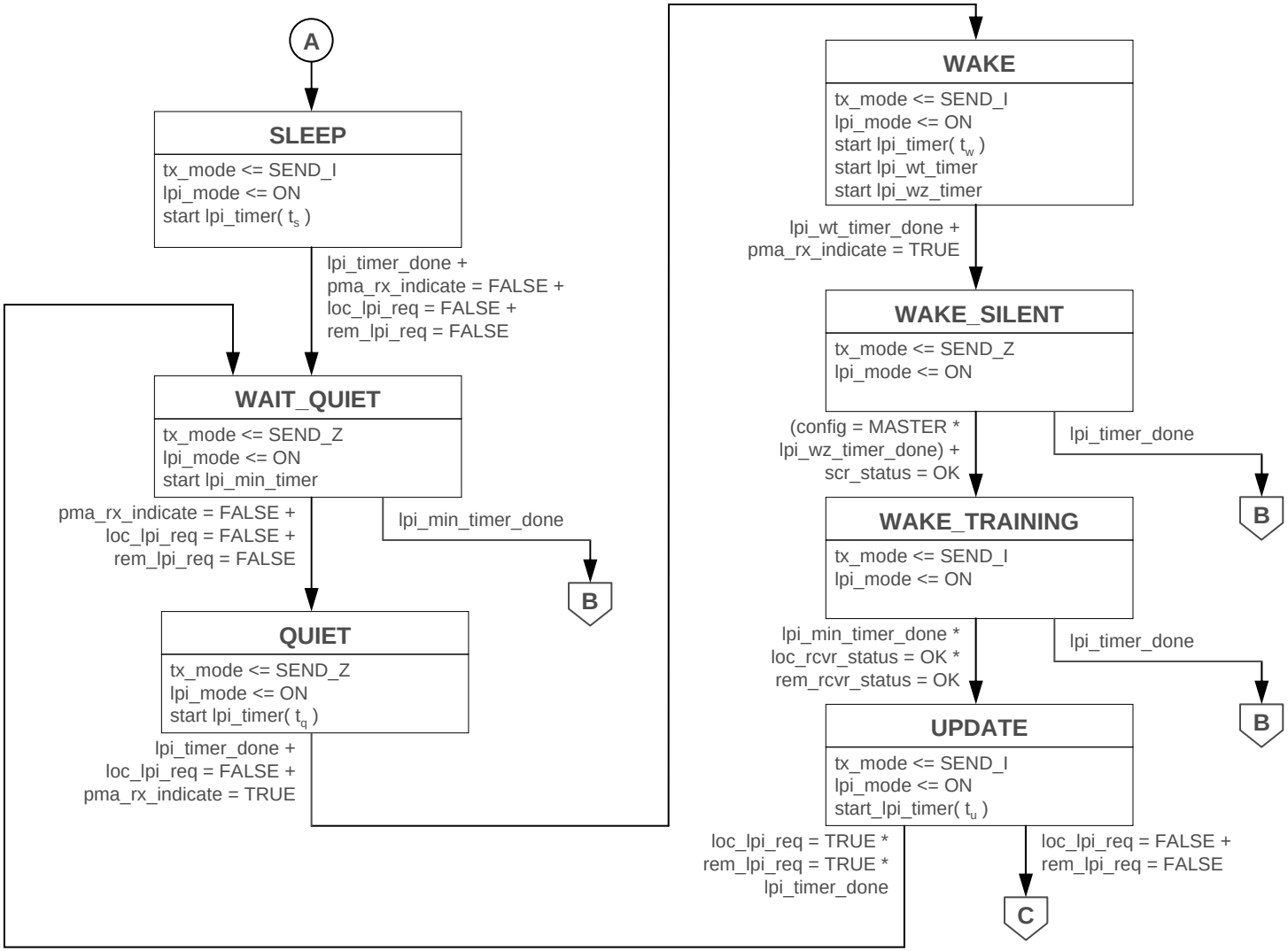
PCS Receive state diagram part a, Figure 40–10a



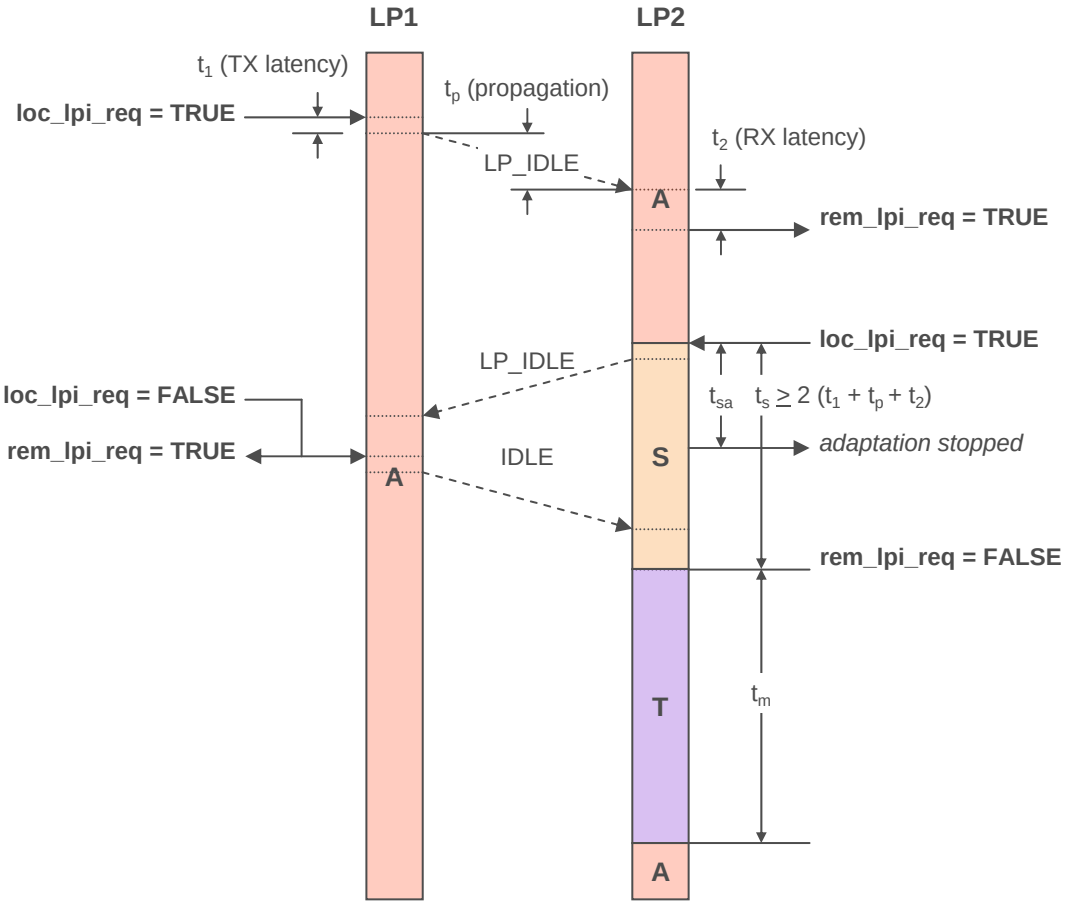
PHY Control state diagram, Figure 40–15

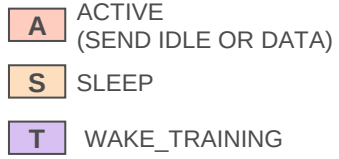


Low-power idle state diagram (PHY Control extension)

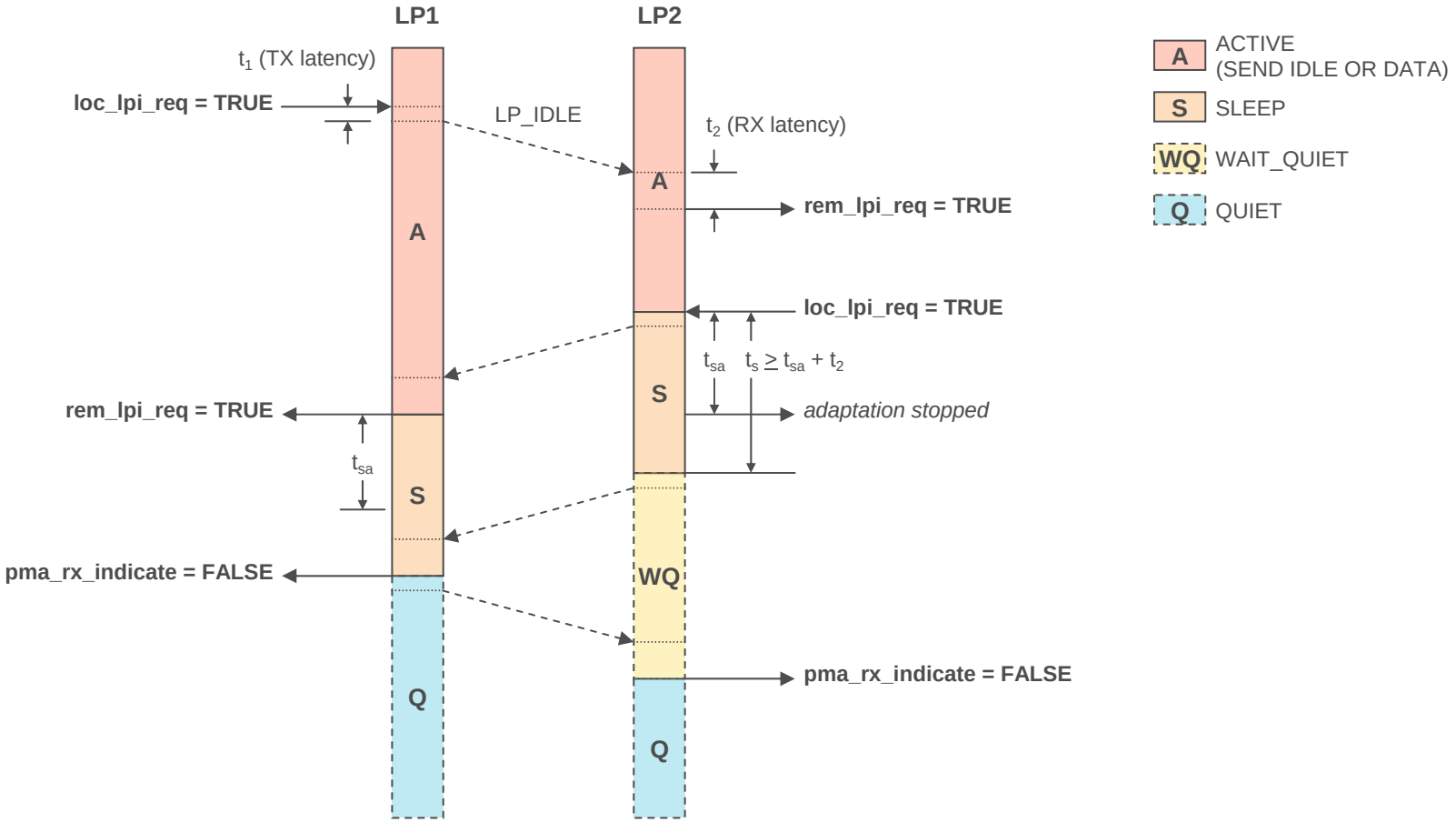


Timing diagram: Only Master or Slave sleeps

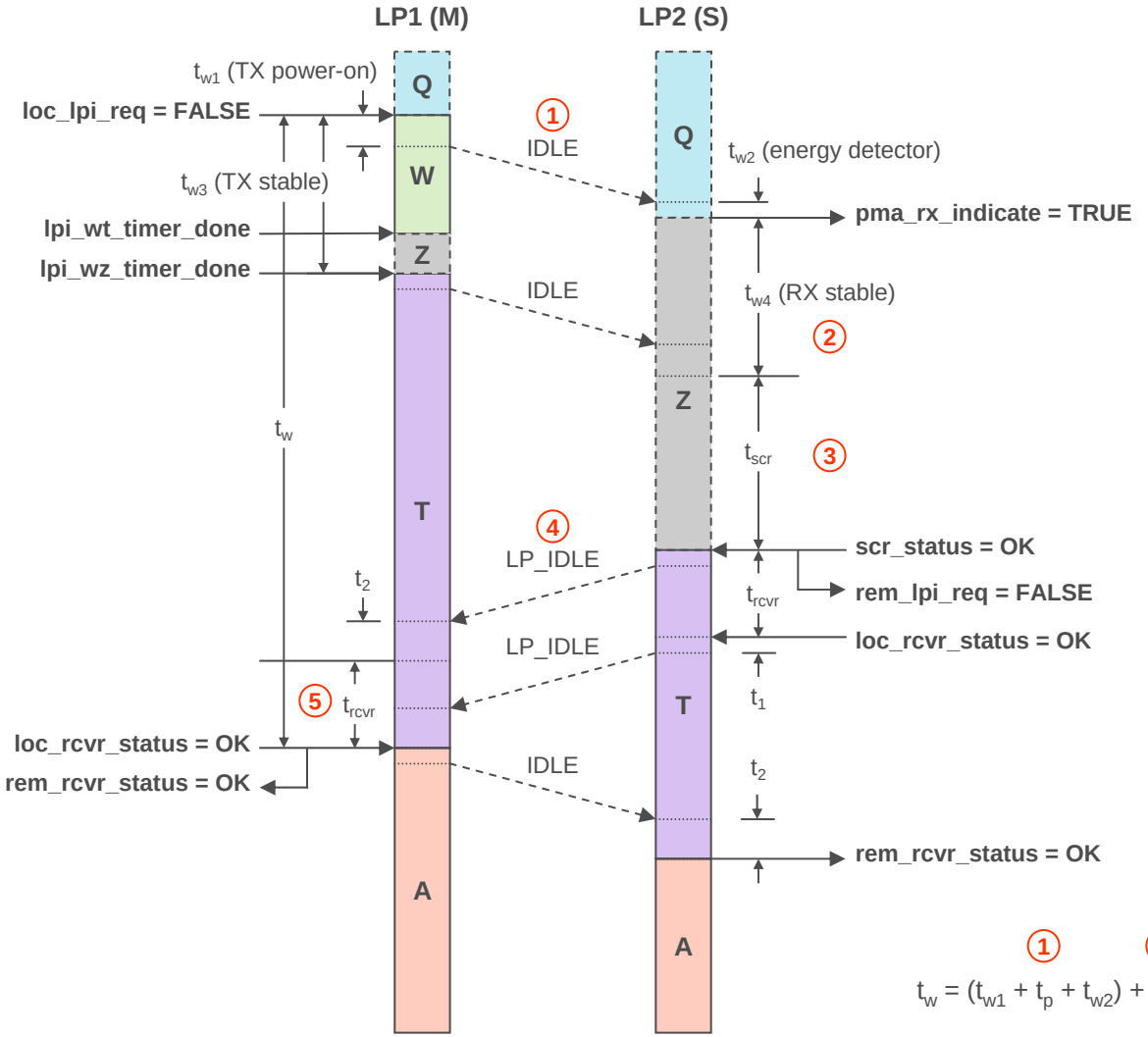




Timing diagram: Enter low-power idle



Timing diagram: Master wakes up first



A ACTIVE (SEND IDLE OR DATA)
S SLEEP
WQ WAIT_QUIET
Q QUIET
W WAKE
Z WAKE_SILENT
T WAKE_TRAINING

$$t_w = (t_{w1} + t_p + t_{w2}) + t_{w4} + t_{scr} + (t_1 + t_p + t_2) + t_{rcvr}$$

(1) (2) (3) (4) (5)

Wake-up time analysis: Master wakes up first

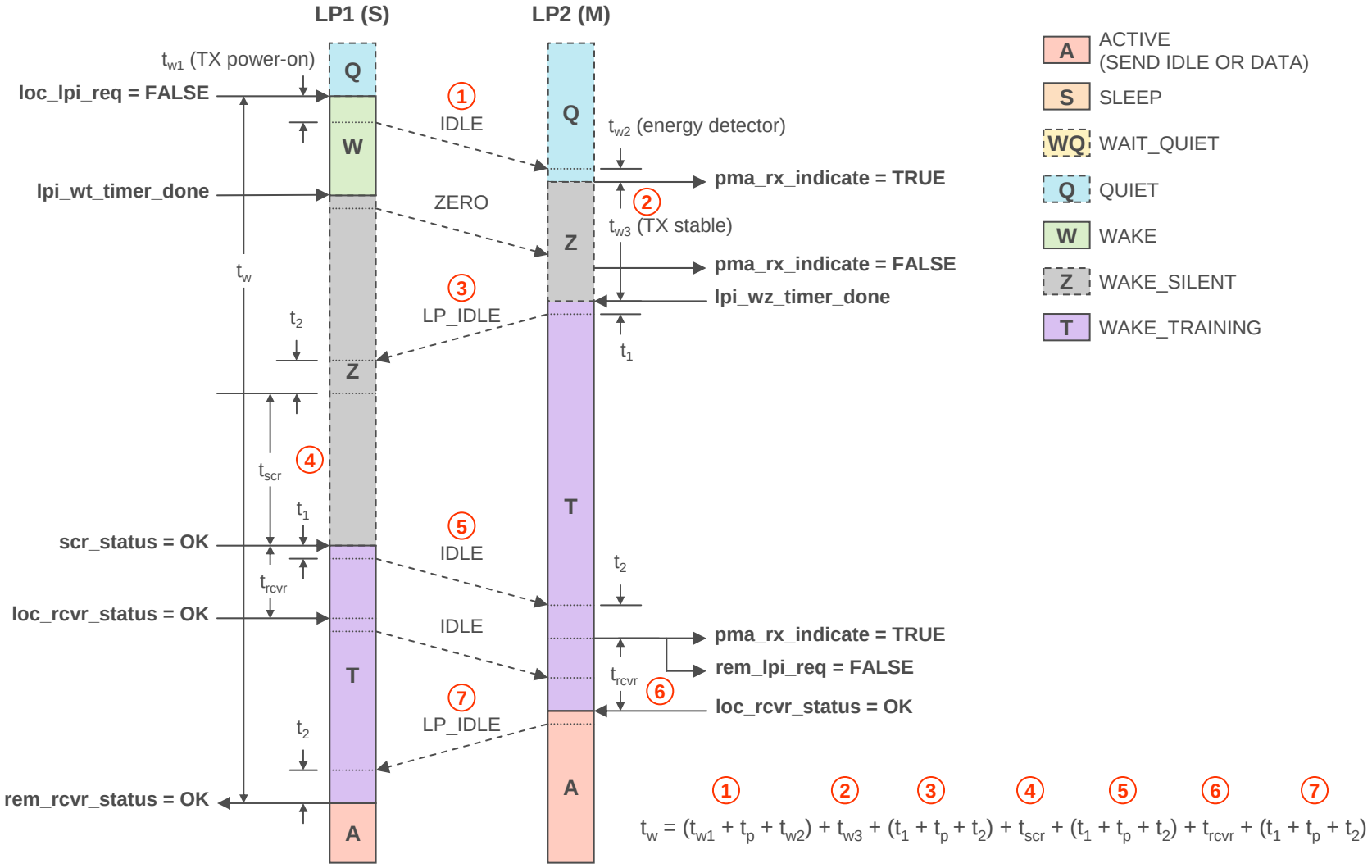
$$t_w = (t_{w1} + t_p + t_{w2}) + t_{w4} + t_{scr} + (t_1 + t_p + t_2) + t_{rcvr}$$

(1) (2) (3) (4) (5)

Label	Max.	Units
(1)	1,650	BT
(2)	5,000	BT
(3)	3,000	BT
(4)	878	BT
(5)	1,000	BT
t_w	**11,528**	**BT**

- Times chosen to allow very low power during quiescent mode
- The wake up time includes two cable propagation delays
- Maximum envelope frame size, with preamble and minimum inter-frame gap, is 16,160 BT

Timing diagram: Slave wakes up first



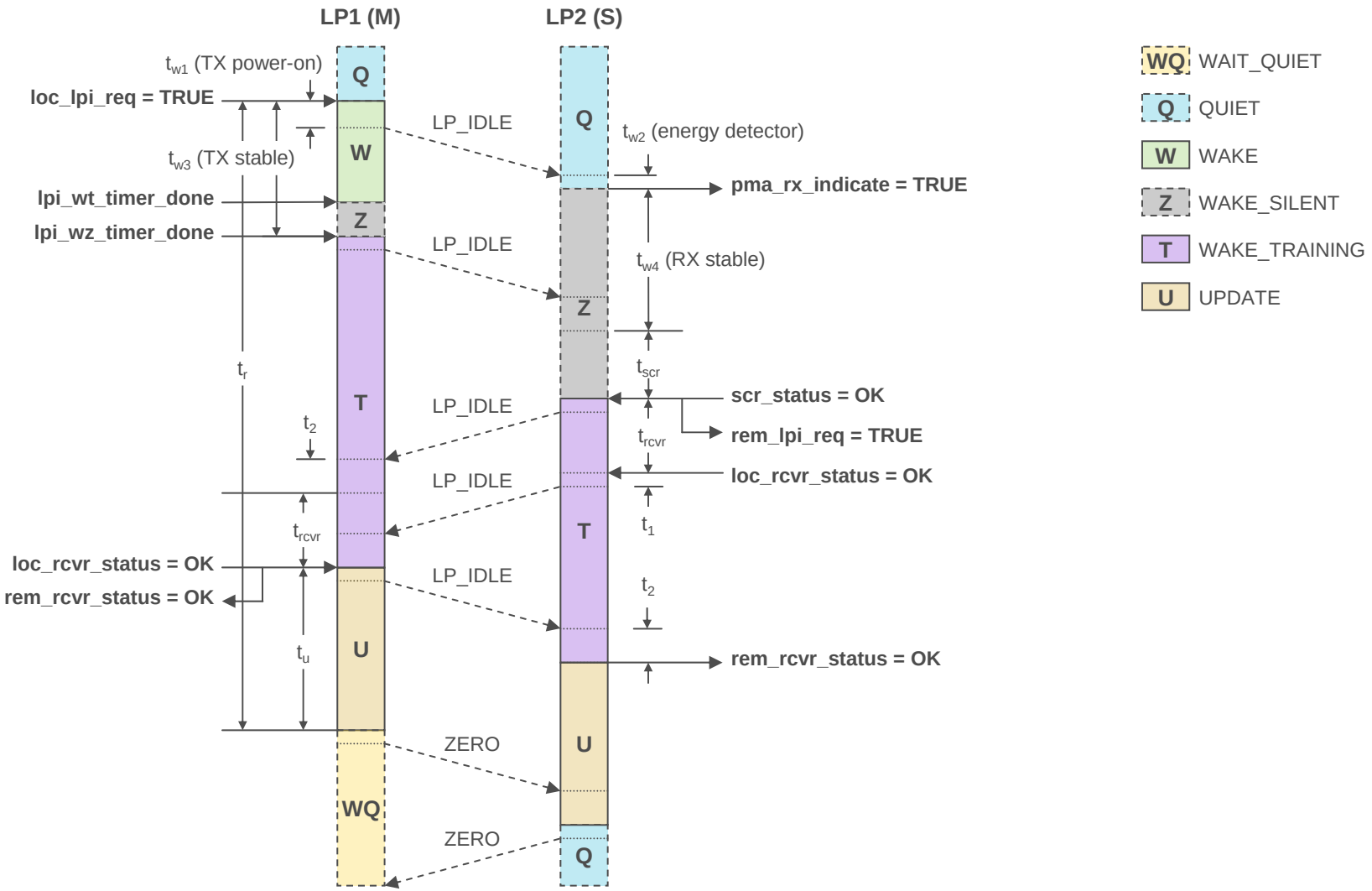
Wake-up time analysis: Slave wakes up first

$$t_w = \underset{①}{(t_{w1} + t_p + t_{w2})} + \underset{②}{t_{w3}} + \underset{③}{(t_1 + t_p + t_2)} + \underset{④}{t_{scr}} + \underset{⑤}{(t_1 + t_p + t_2)} + \underset{⑥}{t_{rcvr}} + \underset{⑦}{(t_1 + t_p + t_2)}$$

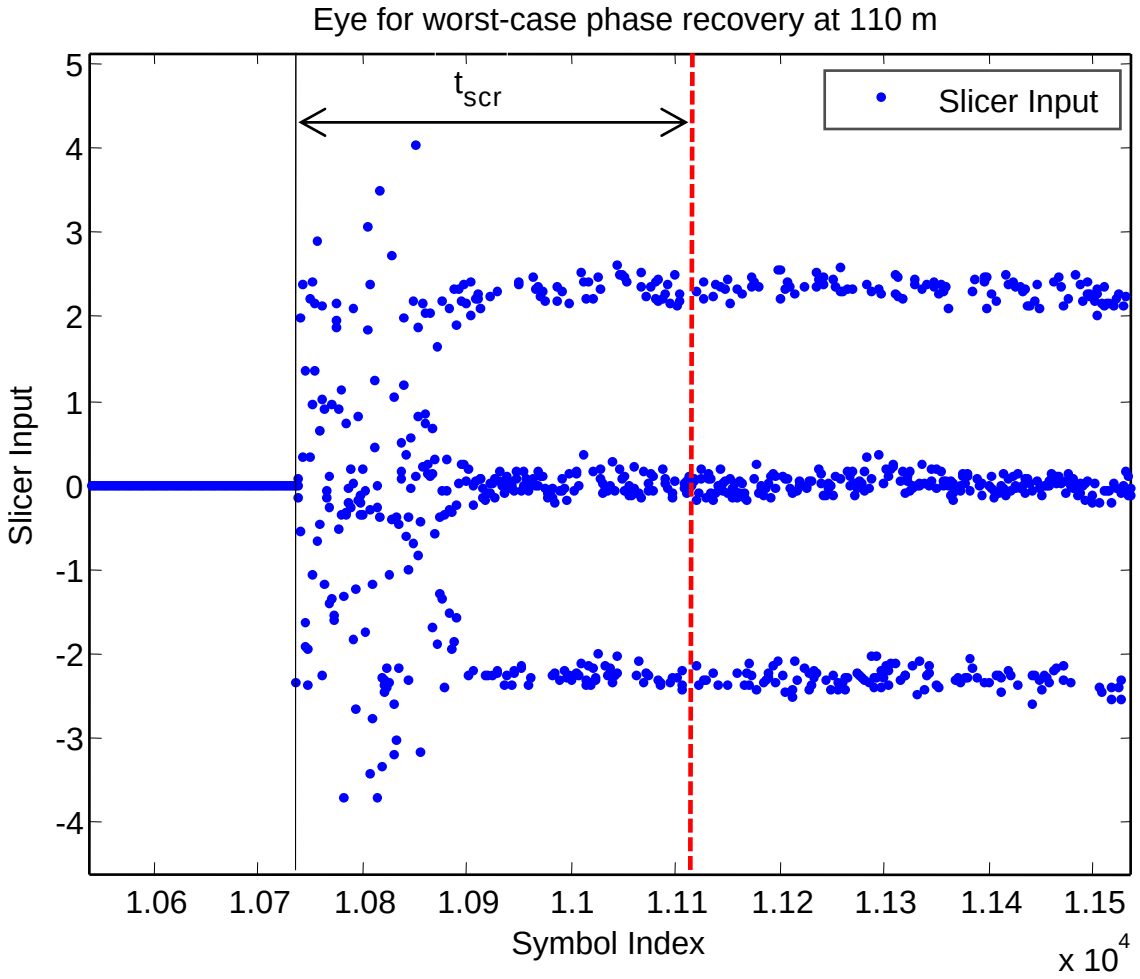
Label	Max.	Units
①	1,650	BT
②	5,000	BT
③	878	BT
④	3,000	BT
⑤	878	BT
⑥	1,000	BT
⑦	878	BT
t_w	**13,284**	**BT**

- Times chosen to allow very low power during quiescent mode
- The wake up time includes four cable propagation delays
- Maximum envelope frame size, with preamble and minimum inter-frame gap, is 16,160 BT

Timing diagram: Refresh



Timing recovery simulation results



1000BASE-T timing parameters (current view)

Label	Parameter	Min.	Typ.	Max.	Units
t_s	Minimum time PHY transmits before going quiet	TBD			ms
t_m	Minimum time PHY waits in the WAKE_TRAINING state before returning to normal operation	1.8			µs
t_q	Time PHY remains quiet before sending refresh		100		ms
t_u	Time PHY transmits during refresh to enable timing and coefficient update		1		ms
t_{wt}	Time PHY transmits to initiate wake-up in the link partner		TBD		ms
t_{w3}	Transmitter stabilization time			5.0	µs
t_w	Maximum time to transition from low-power idle to normal operation			13.3	µs

Summary

- 1000BASE-T low-power idle may be implemented with modest modifications of the existing Clause 40 state diagrams and a new state diagram defining behavior in the low-power mode

- The worst-case time to wake up from low-power idle is estimated to be less than the time to transmit a maximum length envelope frame

- Auto-negotiation of Energy Efficient Ethernet capability, and timing parameters still need to be described
  - Identify which parameters should be negotiated, and which should be fixed
  - Define the range, resolution, and encoding of timing parameters