BASE-U EEE proposal

Rubén Pérez-Aranda, KDPOF
Alejandra Gilarranz, KDPOF

EEE Introduction

EEE introduction (based on clause 78.1)

- EEE (Energy Efficient Ethernet) provides a protocol to **coordinate transitions** to or from a lower level of power consumption, **without changing the link status** and **without dropping or corrupting frames**
- **LPI** (Low Power Idle) is indicated during periods of low data link utilization to allow both sides of the link **saving power**
- The **transition times** into and out of the lower level of power consumption are kept small enough to be transparent to upper layer protocols and applications.
- Currently in IEEE Std 802.3, except for BASE-T, for PHYs with an operating speed of 25Gb/s or greater that implement the optional EEE capability, **two modes of operation** may be supported: **deep sleep** and **fast wake**.

EEE introduction (based on clause 78.1)

- **Deep sleep** refers to the mode for which the **transmitter ceases transmission during LPI**.
  - It is the only mechanism defined for 10 Gb/s or below.
  - **Deep sleep** support is **optional** for PHYs with an operating speed of **25Gb/s or greater** that implement EEE (there are PHY exceptions that do not support deep sleep).
  - When **LPI assertion** is detected on the xMII, the PHY signals **sleep** to its link partner to indicate that the local transmitter is entering LPI mode.
  - PHY transmitter goes **quiet after sleep** is signaled.
  - The **transmit** function of the local PHY is **enabled periodically** to transmit **refresh** signals that are used by the link partner to update adaptive filters and timing circuits in order **to maintain link integrity**.
  - This **quiet-refresh cycle** continues **until** the reception of the **normal interframe** encoding on the xMII. A **wake** signal is sent by the PHY for a predefined period of time before entering the normal operating state.
  - When the **receiver** detects the **sleep** signal, the local PHY indicates “Assert LPI” on the xMII and the local receiver can disable some functionality to reduce power consumption.
  - When the **receiver** detects the **wake** signal, it prepares its circuitry for normal operation and transition from the “Assert LPI” encoding to the normal interframe encoding on the xMII.
  - After a system specified recovery time, the link supports the nominal operational data rate.

EEE introduction (based on clause 78.1)

- **Fast wake** refers to the mode for which the **transmitter continues to transmit signals during LPI** (definition of C/ 1.4.255) so that the receiver can resume operation with a **shorter wake time**.
  - 78.1.3.3.1: For transmit, other than the PCS encoding LPI, there is no difference between fast wake and normal operation.

	EEE mode	
Criteria	**Deep sleep**	**Fast wake**
TX signals sleep to indicate LPI	Yes	No
TX goes quiet after sleep	Yes	No
TX transmits quiet-refresh cycle during LPI	Yes	No
TX signals wake before normal mode	Yes	No
TX continuously transmits signals during LPI	No	Yes
BASE-T PHYs specify it	Yes	No
1000BASE-H PHYs specify it	Yes	No
BASE-SR PHYs specify it (< 25 Gb/s)	No	No
BASE-SR PHYs specify it (≥ 25 Gb/s)	No	Yes
Complexity	Higher	Lower
Wake-up time	Long	Very short
Maximum latency due to LPI	High	Very low
Potential energy saving in TX	Big	Very small
Potential energy saving in RX	Big	Intermediate

EEE introduction (based on clause 78.1)

PHY transmitter operating in EEE deep sleep mode

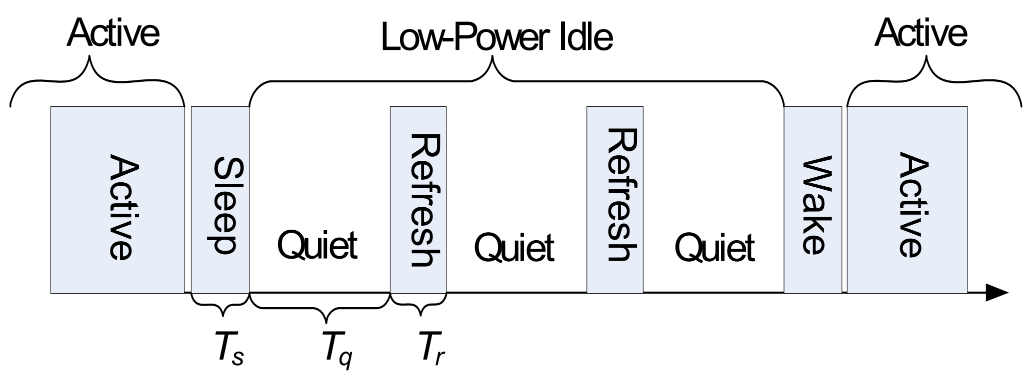
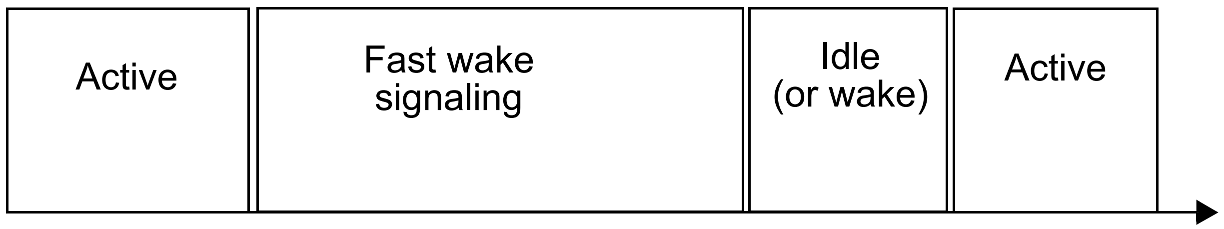


Figure 78–3—Overview of EEE LPI operation

PHY transmitter operating in EEE fast wake mode

Physical Layer signaling continues with higher layer functions suspended during fast wake signaling



NOTE—Fast wake signaling continually indicates LPI in a normally constituted data stream.

Figure 78–4—Overview of fast wake operation

EEE proposal for 802.3cz

EEE proposal for 802.3cz

- This EEE proposal keeps the PHY transmitting signals during LPI to make possible receiver alignment without dealing with long transition times of photonics (it would be the deep sleep case)
- Transmitter side is expected to be low power consumption in normal operation. In addition, RS-FEC encoder function can be disabled during LPI
- Proposal is designed for allowing **big power saving in the receiver** with **very short wake times** that may be beneficial in some use cases, e.g. camera sensors connectivity, where LPI mode can be used during front/back porches time slots

EEE proposal for 802.3cz

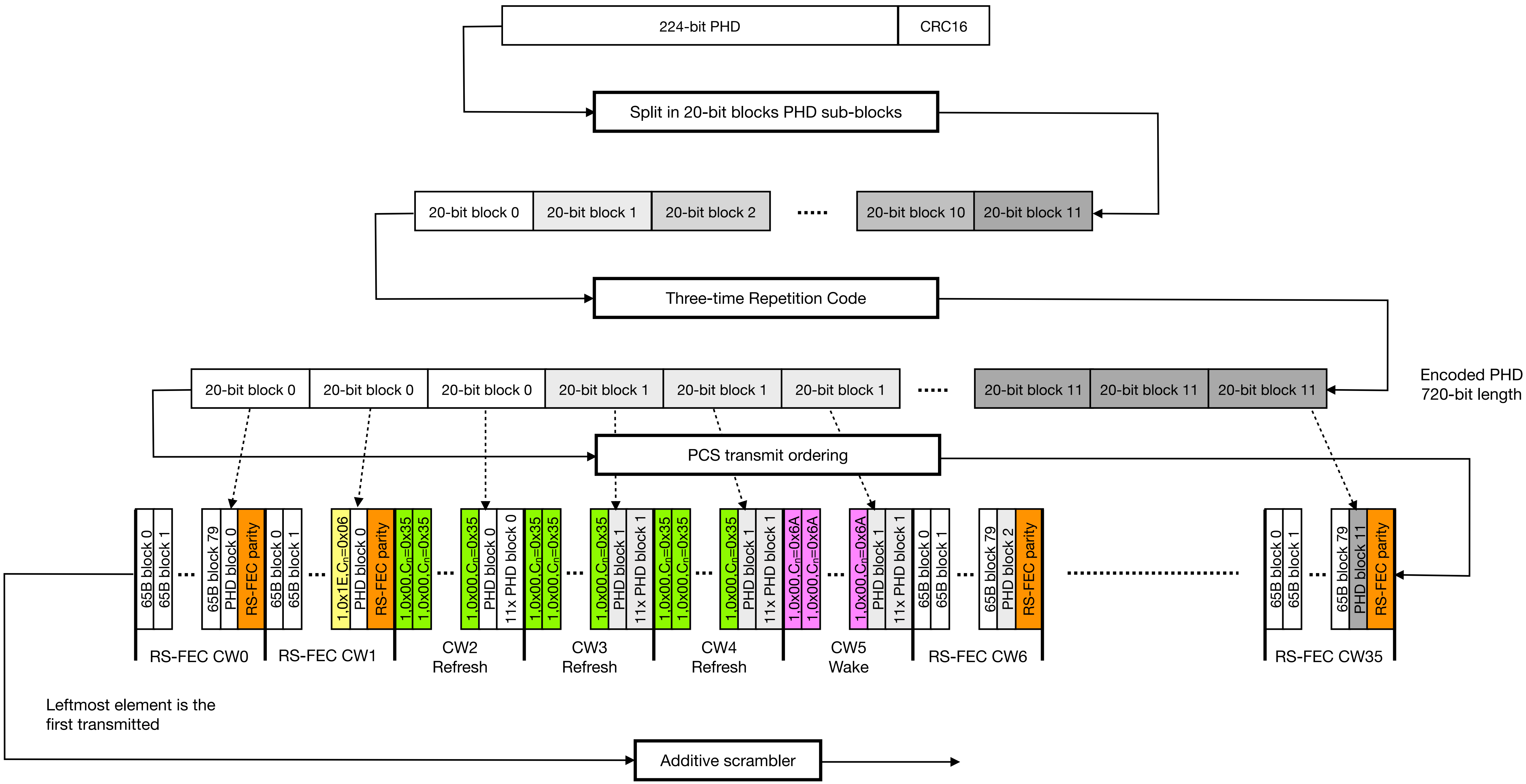
- This EEE proposal applies to the PCS and PMA sublayers of 2.5, 5, 10, 25 and 50GBASE-U
- This is a EEE fast wake proposal:
    - PHY transmitter remains transmitting signals during LPI (same symbol rate and modulation of normal mode)
    - Therefore, the opto-electrical communications channel is as in normal mode and it can be easily tracked
    - The data generated by the PCS sublayer is modified with respect to transparent LPI encoding of normal operation in order to allow power saving, robust OAM side communication channel and robust wake signal detection in the receiver
- The PCS transmit function shall enter LPI mode when the last 65-bit block (i.e. the 65-bit block #79) of a RS-FEC codeword encodes 8 7-bit control codes with value /LI/
    - This last 65-bit block of a RS-FEC codeword has to be composed of: data/control header bit = 1, block type field octet = 0x1E and C_0, C_1, … C_7 = 0x06.
- Then, the PCS transmit function shall enter LPI mode aligned to the next RS-FEC codeword boundary after transmission of the RS-FEC codeword with last 65-bit block encoding 8 /LI/
- When LPI mode, the PCS transmit function shall generate Refresh codewords, replacing normal RS-FEC codewords, each Refresh codeword composed of:
    - 80 equal 65-bit blocks, each one encoding: data/control header bit = 1, block type field octet = 0x00 and C_0, C_1, … C_7 = **0x35**. The content of this part in the Refresh codeword will be static.
    - 12-time sequentially repetition of the 20-bit PHD sub-block that would be transmitted in each codeword in normal mode. The content of this part in the Refresh codeword will be dynamic depending on the 20-bit PHD sub-block content.
- As consequence, Refresh codewords are (80 x 65) + (12 x 20) = 5440 bits long, and RS-FEC parity is not transmitted

EEE proposal for 802.3cz

- If the last 65-bit block of a RS-FEC codeword replaced with Refresh codeword encodes 8 control codes with value /I/, then a Wake codeword shall be transmitted aligned with the next RS-FEC codeword boundary
- The Wake codeword shall be composed of:
  - 80 equal 65-bit blocks, each one encoding data/control header bit = 1, block type field octet = 0x00 and C_0, C_1, … C_7 = **0x6A**. The content of this part in the Wake codeword will be static**.**
  - 12-time sequentially repetition of the 20-bit PHD sub-block that would be transmitted in each codeword in normal mode. The content of this part in the Refresh codeword will be dynamic depending on the 20-bit PHD sub-block content.
- The Wake and Refresh codewords are only different in the 8 7-bit control code values
- When one Wake codeword has been transmitted, normal operation shall be resumed by the PCS transmit function carrying out xMII encoding according to PCS 64B/65B transmit state diagram, PHD encoding and RS-FEC encoding, aligned with the next RS-FEC codeword after Wake codeword transmission

- The PCS receive function shall indicate “Assert LPI” on the xMII during the reception of the Refresh codewords and normal interframe during the reception of the Wake codewords
- PCS receive function shall resume normal operation (RS-FEC codewords decoding, PHD decoding and 64B/65B decoding) aligned with the start of reception of the first RS-FEC codeword after the Wake codeword

EEE proposal for 802.3cz: example



Transmit block, composed by 36 RS codewords, 195840 bits

Robust encoding considerations

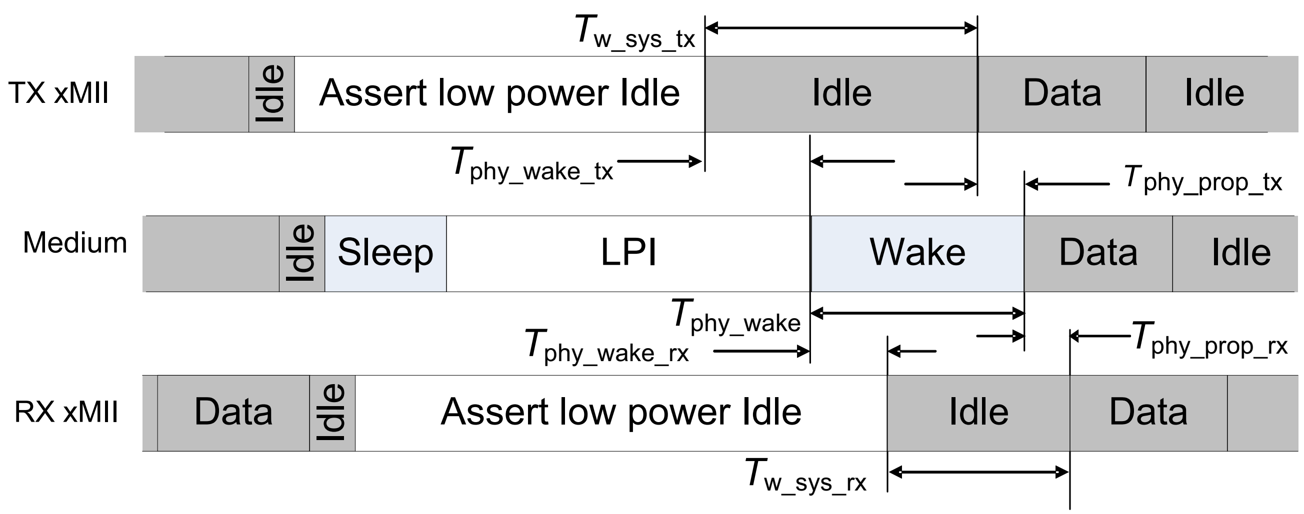
- Block type and control codes
  - Block type 0x00 has been selected because it is the one with minimum Hamming distance of 4 with all the used block types 0x1E, 0x2D, 0x33, 0x66, 0x55, 0x78, 0x4B, 0x87, 0x99, 0xAA, 0xB4, 0xCC, 0xD2, 0xE1, 0xFF
    - 50GMII uses a subset of block types of XGMI and 25GMII
    - LPI can be detected by the receiver based on detection of the control-data header and the block type field of the 65-bit blocks belonging to the Refresh codewords
  - 7-bit control codes 0x35 (Refresh) and 0x6A (Wake) have been selected to have a minimum Hamming distance of 4 with control codes 0x00 /I/, 0x06 /LI/ and 0x1E /E/ and Hamming distance of 6 between them
    - Wake can be detected by the receiver based on detection of the 7-bit control codes of the 65-bit blocks belonging to the Wake codewords
- PHD decoding
  - In normal operation, the $BER_{PHD} < 2 \cdot 10^{-24}$ for $RFER < 5 \cdot 10^{-10}$ ($BER_{PAYLOAD} < 10^{-12}$) after RS-FEC decoding and TRC decoding
  - In LPI mode, there is a concatenation of mRC (m-time repetition code) with the TRC
    - For m = 11: $BER_{PHD} < 7.4 \cdot 10^{-41}$, after mRC and TRC decoding
    - For m = 9: $BER_{PHD} < 2.0 \cdot 10^{-34}$
    - For m = 7: $BER_{PHD} < 5.6 \cdot 10^{-28}$
    - For m = 5: $BER_{PHD} < 1.7 \cdot 10^{-21}$
    - For m = 3: $BER_{PHD} < 6.7 \cdot 10^{-15}$
  - Selection of m value for the 20-bit PHD sub-blocks decoding depends on receiver implementation

Power saving considerations

- Receiver power saving depends on the implementation, but in general:
  - After detected LPI, while receiving Refresh codewords, the receiver only needs to sample, equalize and detect a small portion of symbols for each codeword: only the last *n* 65-bit blocks plus the first *m* repeated 20-bit PHD sub-blocks are needed to detect Wake codeword and make robust decoding of PHD content
  - Example: if $n = 1$ and $m = 7$, the $(1 \times 65 + 7 \times 20)/5440 = 0.0376$ —> less than 4% of the symbols per codeword need to be received
  - In addition, RS-FEC decoder can be fully disabled in LPI
  - Values for *n* and *m* depend on the receiver implementation

EEE proposal for 802.3cz: timing description copied from C/ 78

Figure 78–5 illustrates the relationship between the LPI mode timing parameters and the minimum system wake time.



$T_{w\_sys\_tx}$ (min) $= T_{w\_sys\_rx}$ (min) $+ T_{phy\_shrink\_tx}$ (max) $+ T_{phy\_shrink\_rx}$ (max)

T_{w_phy} (min) $= T_{phy\_wake}$ (min) $+ T_{phy\_shrink\_tx}$

$T_{w\_sys\_res}$ (min) is greater of $T_{w\_sys\_tx}$ (min) and T_{w_phy} (min)

$T_{phy\_shrink\_tx}$ (max) $= (T_{phy\_wake\_tx}$ (max) $- T_{phy\_prop\_tx}$ (min))

$T_{phy\_shrink\_rx}$ (max) $= (T_{phy\_wake\_rx}$(max) $- T_{phy\_prop\_rx}$ (min))

where

- $T_{phy\_wake\_tx}$ = xMII start of wake to MDI start of wake delay
- $T_{phy\_prop\_tx}$ = xMII to MDI data propagation delay
- $T_{phy\_wake\_rx}$ = MDI start of wake to xMII start of wake delay
- $T_{phy\_prop\_rx}$ = MDI to xMII data propagation delay
- T_{phy_wake} = Minimum wake duration required by PHY

Figure 78–5—LPI mode timing parameters and their relationship to minimum system wake time

- T_{w_phy}: Parameter employed by the system that corresponds to the behavior of the PHY. It is the period of time between reception of an IDLE signal on the xMII and when the first data codewords are permitted on the xMII. The wake time of a compliant PHY does not exceed T_{w_phy} (min).
- $T_{w\_sys\_tx}$: Parameter employed by the system that corresponds to its requirements. It is the longest period of time the system has to wait between a request to transmit and its readiness to transmit.
- $T_{w\_sys\_rx}$: Parameter employed by the system that corresponds to its requirements. It is the minimum time required by the system between a request to wake and its readiness to receive data.

EEE proposal for 802.3cz: timing proposal

- The worst case of the time between an normal idle signal in the xMII and when the first data codewords are permitted on the xMII is equal to the time needed to transmit 2 codewords (1 Refresh + 1 Wake)
- The minimum time required by the system between a request to wake and its readiness to receive data is equal to the time equivalent for transmitting a Wake codeword

	Data-rate (Gb/s)				
Timing parameters	**2.5**	**5**	**10**	**25**	**50**
$T_{w\_sys\_tx}$ (min) (μs)	4.10	2.05	1.03	0.41	0.21
T_{w_phy} (min) (μs)	4.10	2.05	1.03	0.41	0.21
$T_{phy\_shrink\_tx}$ (min) (μs)	0	0	0	0	0
$T_{phy\_shrink\_rx}$ (min) (μs)	0	0	0	0	0
$T_{w\_sys\_rx}$ (min) (μs)	2.05	1.03	0.52	0.21	0.11

Thank you!