## 78.4 Data Link Layer Capabilities

Additional capabilities and settings are supported using frames based on the IEEE 802.3 Organizationally Specific TLVs defined in Annex G of IEEE Std 802.1AB protocol (LLDP). Devices that require additional sleep times prior to being able to accept data on their receive paths may use the Data Link Layer capabilities defined in this section to negotiate for extended system wake up times from the transmitting link partner. This mechanism may allow for more or less aggressive energy saving modes.

The Data Link Layer capabilities shall be implemented for devices operating at link rates equal to or greater than 10 Gbps and may be implemented for all other devices.

***Editor's Notes: To be removed prior to publication***
*IEEE P802.3bc Task Force is moving the IEEE 802.3 Organizationally Specific TLVs from Annex F of IEEE Std 802.1ABREV protocol (LLDP) to Clause 79.*

Implementation that use the Data Link Layer capabilities shall comply with all mandatory parts of IEEE Std 802.1AB; shall support the EEE Type, Length, Value (TLV) defined in 78.4.1; timing requirement in 78.4.3; and shall support the control state diagrams defined in 78.4.4.

The Data Link Layer capabilities are described from a unidirectional perspective on the link between transmitting and receiving link partners. For duplex EEE links that implement the Data Link Layer capabilities, each link partner shall implement the TLV, control and state diagrams for a transmitter as well as a receiver.

.

***Editor's Notes: To be removed prior to publication***
*Cross reference to Clause 30 will be added when Clause 30 is completed.*

### 78.4.1 EEE TLV

.

***Editor's Note: To be removed prior to publication***
*The P802.3az Task Force expects that Annex F.3 will soon be transferred to 802.3 via the IEEE P802.3bc Task Force. Subsequently, the changes to the TLV captured herein will be converted to detailed editorial changes against the new 802.3 material*

The EEE TLV is used to perform the EEE Data Link Layer capabilities. Figure 78–3 shows the format of this TLV.

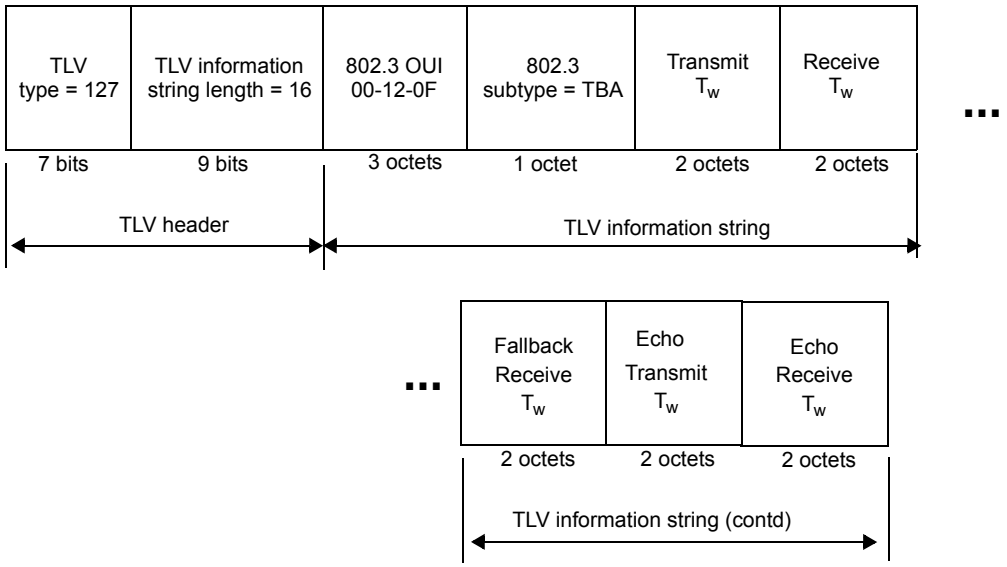


**Figure 78–3—EEE TLV format**

### 78.4.1.1 Transmit $T_w$

Transmit Tw_sys (2 octets wide) is the time (expressed in microseconds) that the transmitting link partner will wait before it starts transmitting data after leaving the Low Power Idle mode. This is a function of the transmit system design and may be constrained, for example, by the transmit path buffering. The default value for Transmit $T_{w\_sys}$ is the $T_{w\_phy}$ defined for the PHY that is in use for the link. The Transmitting link partner expects that the Receiving link partner will be able to accept data after the time delay Transmit Tw_sys (expressed in microseconds).

### 78.4.1.2 Receive $T_w$

Receive Tw_sys (2 octets wide) is the time (expressed in microseconds) that the receiving link partner is requesting the transmitting link partner to wait before it starts transmitting data following the Low Power Idle. Receive Tw_sys (2 octets wide) is the time (expressed in microseconds) that the receiving link partner is requesting the transmitting link partner to wait before it starts transmitting data following the Low Power Idle. The default value for Receive Tw_sys is the Tw_phy defined for the PHY that is in use for the link. The Receive Tw_sys value can be larger than the default, and the extra wait time may be used by the receive link partner for power saving mechanisms that require longer wake-up time than the PHY-layer definitions.

### 78.4.1.3 Fallback $T_w$

A receiving link partner may inform of the transmitter of what an alternate desired Tw_sys. Since a receiving link partner is likely to have discrete levels for savings, this provides the transmitter with additional information that it may use for a more efficient allocation. As with the Receive Receive Tw_sys, this is 2 octets wide. Systems that do not wish to implement this option default the value to be the same as that of the Receive Tw_sys.

#### 78.4.1.4 Echo Transmit and Receive $T_w$

The respective echo values are the local partner's reflection (echo) of the remote partner's respective values. When a local partner receives its echoed values from the remote partner it can determine whether or not the remote partner has received, registered and processed its most recent values. For example, if the local link partner receives echoed parameters that do not match the values in its local MIB, then the local link partner infers that the remote partner's request was based on stale information.

### 78.4.2 EEE TLV to associated management object class cross-references

The cross-references between the EEE TLV and the EEE local and remote object class attributes (30.XX.YY) are listed in Table-78–2.

**Table 78–2—Energy Efficient Ethernet TLV to EEE object class cross-references**

| TLV name | TLV variable | Clause 30 attribute |
|---|---|---|
| Energy Efficient Ethernet | Transmit $T_{w\_sys}$ | aEEELocTxTwSys |
| | Receive $T_{w\_sys}$ | aEEELocRxTwSys |
| | Echo Transmit $T_{w\_sys}$ | aEEERemTxTwSysEcho |
| | Echo Receive $T_{w\_sys}$ | aEEERemRxTwSysEcho |
| | Fallback $T_{w\_sys}$ | aEEELocFbTwSys |

### 78.4.3 Data Link Layer capabilities timing requirements

An EEE link partner shall send an LLDPDU containing an EEE TLV within 10 seconds of the Link Layer capability exchange being enabled as indicated by the variable dll_enabled.

***Editor's Notes: To be removed prior to publication***
*The adopted baseline, diab_02_0109.pdf, requires dll_enabled to be kicked off at the end of a successful auto-negotiation between EEE capable PHYs. This requires an update to the specific PHY auto-negotiation that is not in this section. Unless otherwise changed, this will be done in Draft D1.3 or a subsequent draft.*

Under normal operation, an LLDPDU containing an EEE TLV with an updated value for the "Echo Transmit Tw_sys" field shall be sent within 10 seconds of receipt of an LLDPDU containing an EEE TLV where the value of "Transmit Tw_sys" field is different from the previously communicated value.

Under normal operation, an LLDPDU containing an EEE TLV with an updated value for the "Echo Receive Tw_sys" field shall be sent within 10 seconds of receipt of an LLDPDU containing an EEE TLV where the value of "Receive Tw_sys" field is different from the previously communicated value.

### 78.4.4 Control state diagrams

The control state diagrams for an EEE transmitting link partner and an EEE receiving link partner specify the externally observable behavior of an EEE transmitting link partner and an EEE receiving link partner implementing Data Link Layer capabilities respectively. EEE transmitting link partners implementing Data Link Layer capabilities shall provide the behavior of the state diagram as shown in Figure 78–4. EEE receiv-

ing link partners implementing Data Link Layer capabilities shall provide the behavior of the state diagram as shown in Figure 78–5.

### 78.4.4.1 Conventions

The body of this subclause is comprised of state diagrams, including the associated definitions of variables,constants, and functions. Should there be a discrepancy between a state diagram and descriptive text, the state diagram prevails.

The notation used in the state diagrams follows the conventions of state diagrams as described in 21.5.

### 78.4.4.2 Constants

LOCAL INITIAL TX VALUE
: Integer (2 octets wide) representing the initial $T_{w\_sys}$ (expressed in microseconds) that the local link partner's transmitter is capable of supporting. This is the value of Transmit $T_{w\_sys}$ than the local system advertises upon initialization.

LOCAL INITIAL RX VALUE
: Integer (2 octets wide) representing the initial $T_{w\_sys}$ (expressed in microseconds) that the local link partner's receiver wants to request from the remote link partner's transmitter. This is the value of Receive $T_{w\_sys}$ that the local system advertises upon initialization.

PHY WAKE VALUE
: Integer (2 octets wide) representing the $T_{w\_phy}$ defined for the PHY that is in use for the link

### 78.4.4.3 Variables

LocTxSystemValue
: Integer that indicates the value of $T_{w\_sys}$ that the local system can support. This value is updated by the EEE Transmitter L2 state diagram. This variable maps into the aEEELocTxTwSys attribute.

LocTxSystemValueEcho
: Integer that indicates the value Transmit $T_{w\_sys}$ echoed back by the remote system. This value maps from the aEEELocTxTwSysEcho attribute.

LocRxSystemValue
: Integer that indicates the value of $T_{w\_sys}$ that the local system requests from the remote system. This value is updated by the EEE Receiver L2 state diagram. This variable maps into the aEEE-LocRxTwSys attribute.

LocRxSystemValueEcho
: Integer that indicates the value of Receive $T_{w\_sys}$ echoed back by the remote system. This value maps from the aEEELocRxTwSysEcho attribute.

LocFbSystemValue
: Integer that indicates the value of fallback $T_{w\_sys}$ that the local system requests from the remote system. This value is updated by the local system software.

RemTxSystemValue
: Integer that indicates the value of $T_{w\_sys}$ that the remote system can support. This value maps from the aEEERemTxTwSys attribute.

RemTxSystemValueEcho

Integer that indicates the remote system's Transmit $T_{w\_sys}$ that was used by the local system to compute the $T_{w\_sys}$ that it wants to request from the remote system. This value maps into the aEEERemTxTwStsEcho attribute.

RemRxSystemValue

Integer that indicates the value of $T_{w\_sys}$ that the remote system requests from the local system. This value maps from the aEEERemRxTwSys attribute.

RemRxSystemValueEcho

Integer that indicates the remote systems Receive $T_{w\_sys}$ that was used by the local system to compute the $T_{w\_sys}$ that it can support. This value maps into the aEEERemRxTwSysEcho attribute.

LocResolvedTxSystemValue

Integer that indicates the current $T_{w\_sys}$ supported by the local system.

LocResolvedRxSystemValue

Integer that indicates the current $T_{w\_sys}$ supported by the remote system.

TempTxVar

Temporary integer used to store the value of $T_{w\_sys}$.

TempRxVar

Temporary integer used to store the value of $T_{w\_sys}$.

local_system_change

An implementation specific control variable that indicates that the local system wants to change either the Transmit $T_{w\_sys}$ or the Receive $T_{w\_sys}$.

dll_ready

This variable indicates that the system is ready to send/receive LLDPDU containing EEE TLV. This variable is updated by the local system software.

A summary cross-references between the EEE object class attributes and the transmit and receive control state diagrams, including the direction of the mapping, is provided in Table-78–3.

**Table 78–3—Attribute to state diagram variable cross-reference**

| Object | Attribute | Mapping | State diagram variable |
|---|---|---|---|
| oEEE managed object class | aEEELocTxTwSys | <= | LocTxSystemValue |
| | aEEELocRxTwSys | <= | LocRxSystemValue |
| | aEEELocTxTwSysEcho | => | LocTxSystemValueEcho |
| | aEEELocRxTwSysEcho | => | LocRxSystemValueEcho |
| | aEEELocFbTwSys | <= | LocFbSystemValue |
| | aEEERemTxTwSys | => | RemTxSystemValue |
| | aEEERemRxTwSys | => | RemRxSystemValue |
| | aEEERemTxTwSysEcho | <= | RemTxSystemValueEcho |
| | aEEERemRxTwSysEcho | <= | RemRxSystemValueEcho |
| | aEEEDLLReady | <= | dll_ready |

#### 78.4.4.4 Functions

examine_Tx_change

This function computes the new value of $T_{w\_sys}$ that the local system can support when there is as updated request from the remote system or if local system conditions require a change in the value of the presently supported $T_{w\_sys}$. This function returns the following variable.

NEW_TX_VALUE

Integer that indicates the value of $T_{w\_sys}$ that the local system can support.

examine_Rx_change

This function computes the new value of $T_{w\_sys}$ that the local system wants the remote system to support. This function is called when the remote system wants to change its presently allocated $T_{w\_sys}$ or if local system conditions require a change in the value of $T_{w\_sys}$ presently supported by the remote system. This function returns the following variable.

NEW_RX_VALUE

Integer that indicates the value of $T_{w\_sys}$ that the local system wants the remote system to support.

#### 78.4.4.5 State diagrams

***Editor's Notes: To be removed prior to publication***
*The state diagrams were redrawn natively in Frame for future maintainability of the document. Hence, the appearance may have changed from the .ppt version, however, any functionality changes were limited to the adopted comment resolutions and/or motions from the March 2009 Plenary.*

Control for placing data on the medium rests with the transmitting side, hence Tw_sys is enforced by the transmitter. Thus, for a given path between a set of link partners (i.e. a transmitter and its associated receiver), the transmitting link partner shall wait for the time indicated by the Transmit Tw_sys after deasserting Low Power Idle (at the xxMII) before sending data frames. Similarly the receiving link partner shall be ready to accept data based on its echoed value of Transmit link partner's Tw_sys. This ensures that the link partners transition out of LPI mode and receive frames without loss or corruption.

The general state change procedure for transmitter is shown in Figure 78–4.

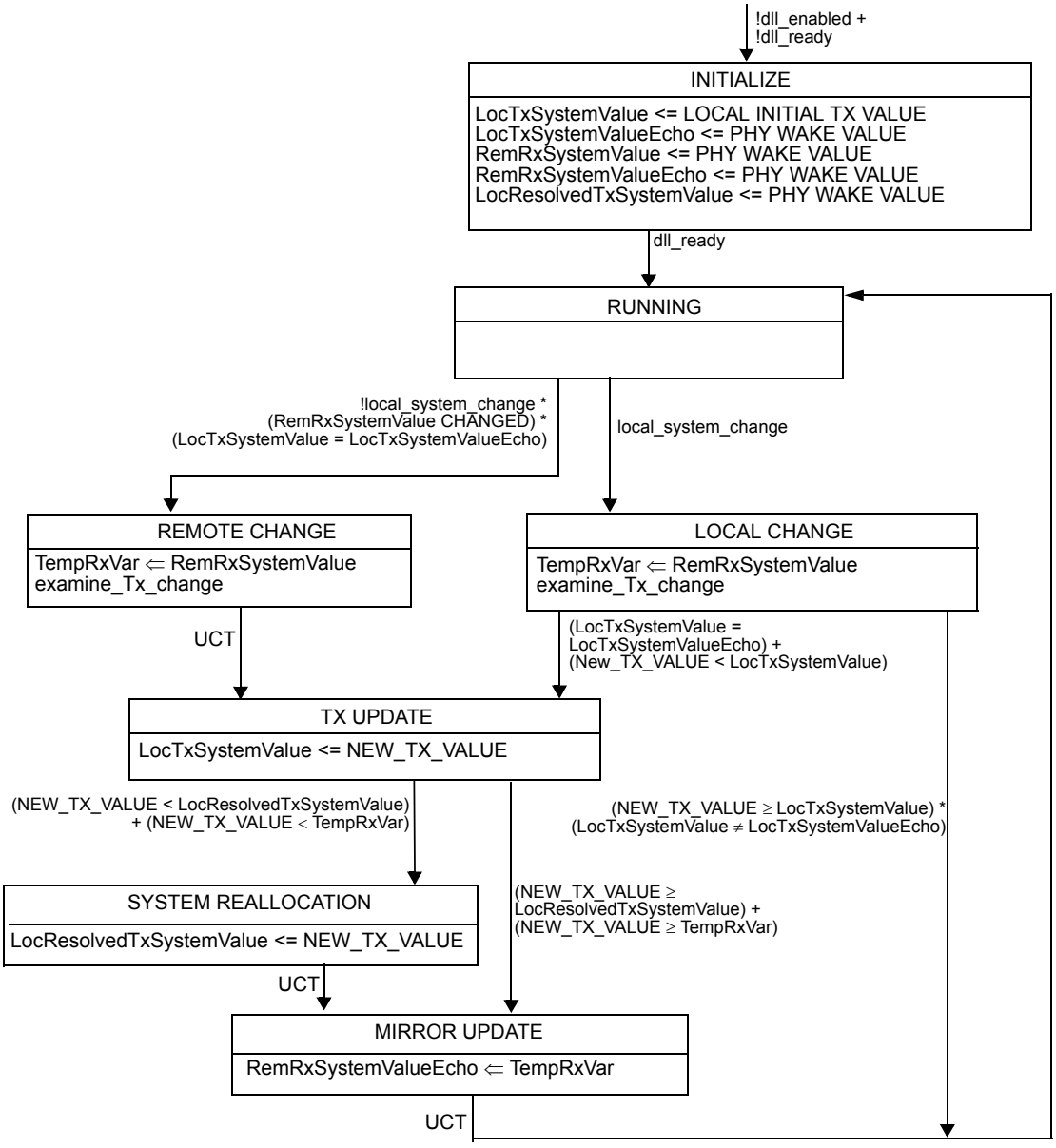


**Figure 78–4—EEE Layer DLL Transmitter State Diagram**

The general state change procedure for receiver is shown in Figure 78–5.

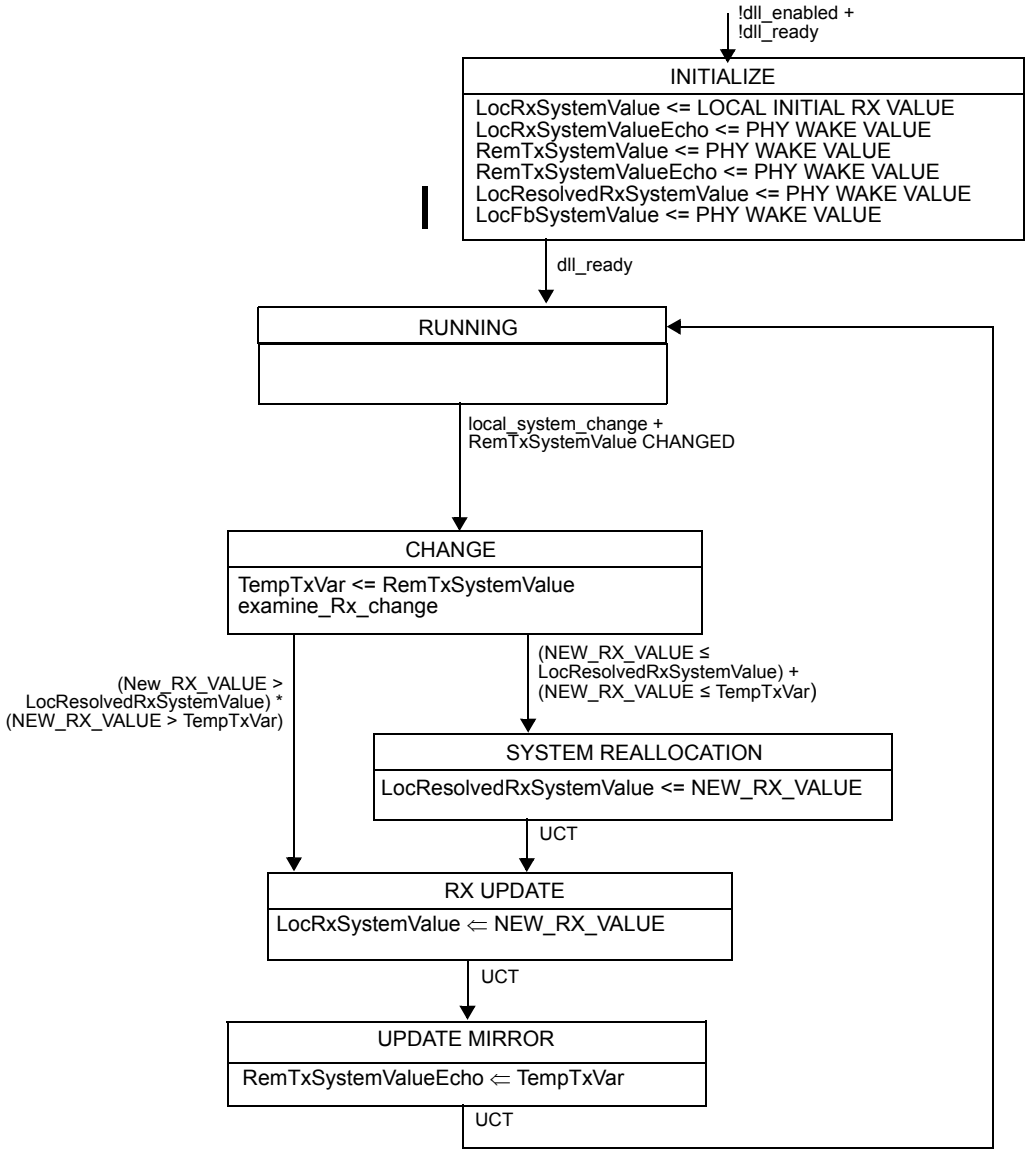


**Figure 78–5—EEE DLL Receiver State Diagram**

### 78.4.5 State change procedure across a link

The transmitting and receiving link partners utilize the LLDP mechanism to advertise their various attributes to the other entity.

The initial Tw_sys defaults governing the EEE operation of the link default to the wake values required by the PHYs. This provides for EEE operation and functionality on initialization and prior to the exchange and processing of the TLVs.

The receiving link partner may request a new $T_{w\_sys}$ value through the aEEELocRxTwSys (30.XX.YY.ZZ) attribute in the EEE object class. The request appears to the transmitting link partner as a change to the aEEERemRxTwSys (30.XX.YY.ZZ) attribute in the EEE object class. The transmitting link partner responds to its receiving partner's request through the aEEELocTxTwSys (30.XX.YY.ZZ) attribute in the EEE object class. The transmitting link partner also copies the value of the aEEERemRxTwSys (30.XX.YY.ZZ) attribute in the EEE object class to the aEEERemRxTwSysEcho (30.9.1.1.19) attribute in the EEE object class.

The transmitting link partner may advertise new value of $T_{w\_sys}$ through the aEEELocTxTwSys (30.XX.YY.ZZ) attribute in the EEE object class. This appears to the receiving link partner as a change to the aEEERemRxTwSys (30.XX.YY.ZZ) attribute in the EEE object class. The receiving link partner responds to a transmitter's request through the aEEELocRxTwSys (30.XX.YY.ZZ) attribute in the EEE object class. The receiving link partner also copies the value of the aEEERemRxTwSys (30.XX.YY.ZZ) attribute in the EEE object class to the aEEERemTxTwSysEcho (30.XX.YY.ZZ) attribute in the EEE object class. This appears to the transmitting link partner as a change to the aEEERemTxTwSysEcho (30.XX.YY.ZZ) attribute in the EEE object class.

The state diagrams describe the behavior above.

#### 78.4.5.1 Transmitting link partner's state change procedure across a link

A transmitting link partner is said to be in sync with the receiving link partner if presently advertised value of Transmit $T_{w\_sys}$ and the corresponding echoed value are equal.

During normal operation the transmitting link partner is in the RUNNING state. If the transmitting link partner wants to initiate a change to the presently resolved value of $T_{w\_sys}$, the local_system_change is asserted and the transmitting link partner enters the LOCAL CHANGE state where NEW_TX_VALUE is computed. If the new value is smaller than the presently advertised value of $T_{w\_sys}$ or if the transmitting link partner is in sync with the receiving link partner, then it enters TX UPDATE state. Otherwise it returns to the RUNNING state.

If the transmitting link partner machine sees a change in the $T_{w\_sys}$ requested by the receiving link partner it recognizes the request only if it is in sync with the transmitting link partner. The transmitting link partner examines the request by entering the REMOTE CHANGE state where a NEW TX VALUE is computed and it then enters the TX UPDATE state.

Upon entering the TX UPDATE state, the transmitter updates the advertised value of Transmit $T_{w\_sys}$ with NEW_TX_VALUE. If the NEW_TX_VALUE is lesser than either the resolved $T_{w\_sys}$ value or the value requested by the receiving link partner then it enters the SYSTEM REALLOCATION state where it updates the value of resolved $T_{w\_sys}$ with NEW_TX_VALUE. Irrespective of whether the transmitting link partner enters the SYSTEM REALLOCATION state from the TX UPDATE state; it ultimately returns to the RUNNING state through the UPDATE MIRROR state where it updates the echo for the Receive $T_{w\_sys}$.

#### 78.4.5.2 Receiving link partner's state change procedure across a link

A receiving link partner is said to be in sync with the transmitting link partner if the presently requested value of Receive Tw_sys and the corresponding echoed value are equal.

During normal operation the receiving link partner is in the RUNNING state. If the receiving link partner wants to request a change to the presently resolved value of $T_{w\_sys}$, the local_system_change is asserted. When local_system_change is asserted or when the receiving link partner sees a change in the $T_{w\_sys}$ advertised by the transmitting link partner, it enters the CHANGE state where NEW_RX_VALUE is computed. If NEW_RX_VALUE is lesser than the presently resolved value of $T_{w\_sys}$ or the presently advertised value by the transmitting link partner, it enters SYSTEM REALLOCATION state where it updates the resolved value of Tw_sys to NEW_RX_VALUE. Irrespective of whether the receiving link partner enters the SYSTEM REALLOCATION state, it ultimately gets to the RX UPDATE state.

In the RX UPDATE state, it updates the presently requested value to NEW_RX_VALUE, then it updates the echo for the Transmit Tw_sys in the UPDATE MIRROR state and finally goes back to the RUNNING state.

## 78.5 Communication link access latency

In full duplex mode, predictable operation of the MAC ControlPAUSE operation (Clause 31, Annex 31B) also demands that there be an upper bound on the propagation delays through the network. This implies that MAC, MAC Control sublayer, and PHY implementors must conform to certain delay maxima, and that network planners and administrators conform to constraints regarding the cable topology and concatenation of devices.

***Editor's Notes: To be removed prior to publication***
*This sub-clause will be changed based on shinkage ad-hoc committee report. This committee will use framework in law_01_0109.pdf as adopted by the task force during January interim meeting. In this presentation David identified cases where there could be shrinkage in the system wake time. He introduced new variables, including one for a default system wake time as well as parameters to specify shrinkage*

In addition, EEE operational mode adds latency to be considered by network designer. When at Low Power Idle mode, PHY device is not available immediately for data transmission request. System has to wake it up by sending normal idle code on the MAC interface. Following IDLE code reception on the MAC interface, PHY starts waking up process. The maximal PHY recovery time $T_{w\_phy}$ is defined for each PHY. Table 78–4 summarizes maximal $T_{w\_phy}$ for supported protocols along with three additional key parameters ($T_s$, $T_q$, and $T_r$). This should assist systems designer while considering Low Power Idle modes effect on the overall operation..

***Editor's Notes: To be removed prior to publication***
*Values in Table 78-2 are based on the draft 1.1 and should be treated as temporary. Some of the parameters were taken from the presentation given by Task force members. Editor encourages detailed review and comments on this table values*

**Table 78–4—Summary of the key EEE parameters for supported PHYs**

| Protocol | $T_{w\_phy}$ μsec | | $T_s$ μsec | | $T_q$ μsec | | $T_r$ μsec | |
|---|---|---|---|---|---|---|---|---|
| | **min** | **max** | **min** | **max** | **min** | **max** | **min** | **max** |
| 10GBASE-KR | 11.0 | 16.9 | 4.5 | 5.5 | 1,530 | 1,870 | 15.2 | 18.5 |
| 10GBASE-KX4 | 8.0 | 18.0 | 18.0 | 22.0 | 2,250 | 2,750 | 18.0 | 22.0 |
| 1000BASE-KX | 10.0 | 20.0 | 18.0 | 22.0 | 2,250 | 2,750 | 18.0 | 22.0 |
| 10GBASE-T | 4.16 | 7.36 | 2.88 | 3.2 | 39.7 | 39.68 | 1.28 | 1.28 |
| 1000BASE-T | 16.0 | 16.0 | 182.0 | 202.0 | 20,000 | 24,000 | 198.0 | 218.2 |
| 100BASE-TX | 30.0 | 24,000 | 100 | 100 | 20,000 | 20,000 | 100 | 100 |