BROADCOM®

Supporting material for comment #460.

Jeff Slavick

IEEE Interim May 2024

Supporters

- Arthur Marris
- Kent Lusted
- Adee Ran

BROADCOM

Agenda

- Review of comment text
- Justification
- Implementation details

Comment

Cl **73** *SC* **73** *P* **83** *L* **1** # 460

Slavick, Jeff Broadcom

Comment Type **T** *Comment Status* **X**

We are now using a Next Page to advertise IEEE defined PHYs. However the order of when Next Pages are introduced, defined and then used is a bit out of order. So re-arranging the order in which AN is specified would help readers to better understand what how Next Pages are defined, how to use them and when to use them.

SuggestedRemedy

Presentation will be provided.

Justification

- Prior to 802.3dj the use of Next Pages to determine IEEE defined PHY rates was not necessary.
  - Negotiating to 802.3dj PHY rates requires Message code 2 Next Page exchange in both directions to advertise the new PHY rates.

- Next page definition is added at the end of the clause
  - After Rx process, Renegotiation and Priority resolution and just before the MDIO mapping tables and variable definitions.
  - How next pages are utilized is kind of loosely explained

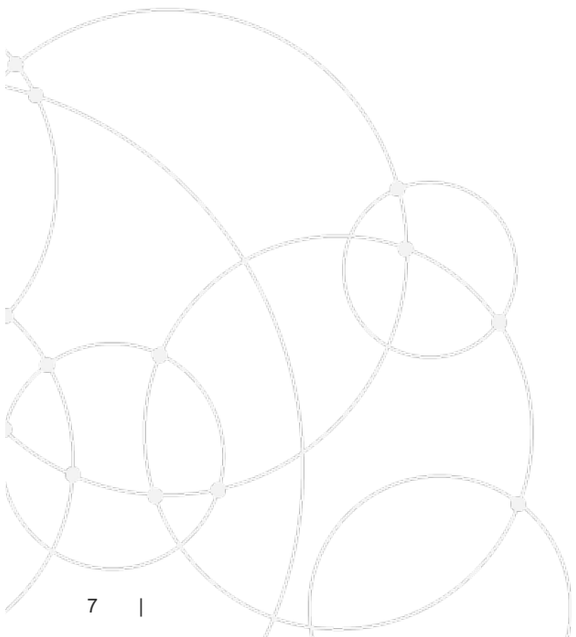
- Nomenclature of the various pages and codewords is not consistent
  - DME Page v. AN Page v. link codeword

The modifications

- Re-order the contents of Clause 73 to bring the description of Next Pages contain earlier in the clause.
  - Add some clarify sentences for how and when they're used
- Be consistent on use of DME Page, AN Page and Link Codeword
  - DME Page is 53 DME cells (4 Frame markers, 1 Random bit, 48 AN page)
  - AN Page is 48 bits of information being sent
  - link codeword <type> defines the field locations of 48 bits of information
    - Types are: Base Page, Message code Next Page, Unformatted Next Page
- Define the various link codewords using similar approach for all 3 types
- No changes to how AN works or is defined to operate
  - Just making it more clear on how the process works as we're making the use of Next Pages necessary to negotiate all possible PHY types.

The modifications – cont.

- The following slides detail the changes to 73.3 through 73.7.7 and 73A.1a
- Underlined text is new.
- Crossed out text is existing text that is removed



Detailed Implementation - 0

73.3 Functional specifications

The Auto-Negotiation function provides a mechanism to control connection of a single MDI to a single PHY type, where more than one PHY type may exist. A management interface provides control and status of Auto-Negotiation, but the presence of a management agent is not required.

The Auto-Negotiation function shall provide the following:

a) Auto-Negotiation transmit
b) Auto-Negotiation receive
c) Auto-Negotiation arbitration

These functions shall comply with the state diagrams from Figure 73–9 through Figure 73–11. The Auto-Negotiation functions shall interact with the technology-dependent PHYs through the Technology-Dependent interface (see 73.9). Technology-Dependent PHYs are those supported by the Auto-Negotiation process (see Table 73–4 and Table 73A-1a).

When the MDI supports multiple lanes, then lane 0 of the MDI shall be used for Auto-Negotiation and for connection of any single-lane PHYs (e.g., 1000BASE-KX or 10GBASE-KR).

Detailed Implementation - 1

73.4 Transmit function ~~requirements~~

The Transmit function provides the ability to transmit DME pages (see 73.5.2). The first DME page exchanged by the local device and its link partner after Power-On, link restart, or renegotiation is referred to as the Base Page and contains the ~~base~~ link codeword Base Page defined in Figure 73–6 (see 73.6.2).

The local device may modify the link codeword (see 73.6) to disable an ability it possesses, but will not transmit an ability it does not possess. This makes possible the distinction between local abilities and advertised abilities so that multi-ability devices may Auto-Negotiate to a mode lower in priority than the highest common ability.

Detailed Implementation - 2

~~73.7.7~~ 73.4.1 Next Page function

The Next page function uses the Auto-Negotiation arbitration mechanisms to allow for the exchange of information beyond what is available in the Base Page. DME pages exchanged after the Base Page are referred to as Next Pages. Next page exchange occurs after the link codeword Base Page has been exchanged if either end of the link segment set the Next Page bit to logical one (see 73.6.1.2) indicating that it has at least one Next Page to send. There are two encodings of the link codeword that may be sent during Next Page exchanges, link codeword Message code (see 73.6.3) and link codeword Unformatted (see 73.6.4).

Multiple Next Pages, with appropriate link codewords (Message code and Unformatted), can be transmitted to send extended messages. Each series of Next Pages begins with the transmission of a link codeword Message code Next Page which defines how to interpret the contents of the link codeword and whether additional link codeword Unformatted Next Pages will follow (see 73A).

Any number of Next Pages may be sent in any order; with the exception that if a link codeword Message code 2 Next Page (see 73A.1a) is to be sent it shall be the first Next Page transmitted following the link codeword Base Page exchange.

It is recommended that the total number of Next Pages sent be kept small to minimize the link startup time.

Next page transmission ends when both ends of a link segment set their Next Page bits to logical zero, indicating that neither has anything additional to transmit. It is possible for one device to have more pages to transmit than the other device. Once a device has completed transmission of its Next Page information, it shall transmit link codeword Message code 1 Next Pages (Null message code) with the NP bit set to logical zero while its link partner continues to transmit valid Next Pages. An Auto-Negotiation able device shall recognize reception of the link codeword with Null message code as the end of its link partner's Next Page information.

Detailed Implementation - 3

~~73.6.10~~ 73.4.2 Transmit Switch function

During Auto-Negotiation and prior to entry into the AN_GOOD_CHECK state, the Transmit Switch function shall connect only the DME page generator controlled by the Transmit State Diagram to the MDI.

Upon entry into the AN_GOOD_CHECK state, the Transmit Switch function shall connect the transmit path from a single technology-dependent (highest common denominator) PHY to the MDI.

When a PHY is connected to the MDI through the Transmit Switch function, the signals at the MDI shall conform to all of the PHY's specifications within 20 ms.

Detailed Implementation – 4

73.5 DME transmission

<NO CHANGE>

BROADCOM

Detailed Implementation - 5

73.6 Link codeword encoding

The 48-bit Auto-Negotiation page can take on three different encodings referred to as link codewords. The three different encodings of link codewords are referred to as follows:

– link codeword Base Page
– link codeword Message code Next Page
– link codeword Unformatted Next Page

The mappings of the data bits (D[47:0]) of the Auto-Negotiation page to specific fields for each type of link codeword are shown in Figure 73-6, Figure 73-7 and Figure 73-8. Within a DME page, D0 shall be the first bit transmitted after the Manchester violation (see Figure 73-5).

All link codeword encodings contain the Ack and NP fields described in 73.6.1. The encoding for the Base Page is described in 73.6.2. The encodings for Message code and Unformatted Next Pages are described in 73.6.3 and 73.6.4 respectively.

The mapping of all possible technologies available for backplane and copper cable assembly by Clause 73 Auto-Negotiation are allocated over two link codeword encodings. The mappings are described in Table 73-4 and 73A-1a and are transmitted in the link codeword Base Page (see 73.6.2) and link codeword Message code 2 Next Page (see 73.6.3 and 73A.1a).

Detailed Implementation - 6

73.6.1 Common link codeword fields

The Ack and NP fields are present in all link codeword encodings and are used to handshake the flow of the Auto-Negotiation process between the two partners.

73.6.1.1 Acknowledge

Acknowledge (Ack) is used by the Auto-Negotiation function to indicate that a device has successfully received its link partner's link codeword. The Acknowledge Bit is encoded in bit D14 of the link codeword. If no Next Page information is to be sent, this bit shall be set to logical one in the link codeword after the reception of at least three consecutive and consistent DME pages (ignoring the Acknowledge bit value). If Next Page information is to be sent, this bit shall be set to logical one after the device has successfully received at least three consecutive and matching DME pages (ignoring the Acknowledge bit value), and will remain set until the Next Page information has been loaded into the AN XNP transmit register (registers 7.22, 7.23, 7.24). In order to save the current received link codeword, it has to be read from the AN LP XNP ability register (register 7.25, 7.26, 7.27) before the Next Page of transmit information is loaded into the AN XNP transmit register. After the COMPLETE ACKNOWLEDGE state has been entered, the link codeword will be transmitted at least six times.

73.6.1.2 Next Page

Next Page (NP) is encoded in bit D15 of the link codeword. Support of Next Pages is mandatory. If the device does not have any Next Pages to send, the NP bit shall be set to logical zero. If a device wishes to engage in Next Page exchange, it shall set the NP bit to logical one. If a device has no Next Pages to send and its link partner has set the NP bit to logical one, it shall transmit Next Pages with Null message codes and the NP bit set to logical zero while its link partner transmits valid Next Pages. Next page exchanges will occur if either the device or its link partner sets the Next Page bit to logical one. The Next Page function is defined in ~~73.7.7~~ 73.4.1.

Detailed Implementation – 7

73.6.2 Link codeword Base Page

The ~~base~~ link codeword ~~(~~Base Page~~)~~ transmitted within a DME page shall convey the encoding shown in Figure 73–6. ~~The Auto-Negotiation function supports additional pages using the Next Page~~ function. ~~Encoding for the link codeword(s) used in the Next Page exchange are defined in 73.7.7~~.

Figure 73–6—Link codeword Base Page

D[4:0] contains the Selector Field. D[9:5] contains the Echoed Nonce field. D[12:10] contains capability bits to advertise capabilities not related to the PHY. C[1:0] is used to advertise pause capability. The remaining capability bit C[2] is reserved. D[15:13] contains the RF, Ack, and NP bits. ~~These bits shall function as specified in 28.2.1.2~~. D[20:16] contains the Transmitted Nonce field. D[42:21] contains the Technology Ability Field. D[47:43] contains FEC capability~~ (see 73.6.5)~~.

Detailed Implementation - 8

73.6.2.1 Selector Field

Selector Field (S[4:0]) is a five-bit wide field, encoding 32 possible messages. Selector Field encoding definitions are shown in Annex 28A. Combinations not specified are reserved for future use. Reserved combinations of the Selector Field shall not be transmitted.

The Selector Field for IEEE Std 802.3 is shown in Table 73–3.

Table 73–3—Selector Field Encoding

73.6.2.2 Echoed Nonce Field

Echoed Nonce Field (E[4:0]) is a 5-bit wide field containing the nonce received from the link partner. When Acknowledge is set to logical zero, the bits in this field shall contain logical zeros. When Acknowledge is set to logical one, the bits in this field shall contain the value received in the Transmitted Nonce Field from the link partner.

73.6.2.3 Transmitted Nonce Field

Transmitted Nonce Field (T[4:0]) is a 5-bit wide field containing a random or pseudo-random number. A new value shall be generated for each entry to the Ability Detect state. The method of generating the nonce is left to the implementer. The transmitted nonce should have a uniform distribution in the range from 0 to $2^5 - 1$. The method used to generate the value should be designed to minimize correlation to the values generated by other devices.

Detailed Implementation - 9

73.6.2.4 Technology Ability Field

Technology Ability Field (A[21:0]) is a 22-bit wide field containing information indicating supported technologies specific to the selector field value when used with the Auto-Negotiation for backplane and copper cable assembly. These bits are mapped to individual technologies such that abilities are advertised in parallel for a single selector field value. The Technology Ability Field encoding for the IEEE 802.3 selector with Auto-Negotiation for backplane and copper cable assembly is described in Table 73–4.

Table 73–4—Technology Ability Field encoding

Multiple technologies may be advertised in the link codeword. A device shall support the data service ability for a technology it advertises. It is the responsibility of the Arbitration function to determine the common mode of operation shared by a link partner and to resolve multiple common modes.

NOTE—Previous editions of this standard prohibited simultaneous advertisement of PHYs that support operation over electrical backplanes with PHYs that support operation over copper cable assemblies.

25GBASE-KR-S abilities are a subset of 25GBASE-KR abilities, and likewise 25GBASE-CR-S abilities are a subset of 25GBASE-CR abilities. To allow interoperation between 25GBASE-KR-S and 25GBASE-KR PHY types, and between 25GBASE-CR-S and 25GBASE-CR PHY types, a device that supports 25GBASE-KR or 25GBASE-CR should advertise both A9 and A10 ability bits during auto-negotiation.

The fields A[21:20] are reserved for future use. Reserved fields shall be sent as zero and ignored on receive.

Detailed Implementation - 10

73.6.2.5 FEC capability
FEC capability (F4, F2, F3, F0, F1) is encoded in bits D43:D47 ~~of the base link codeword~~. The FEC bits are used as follows:

a) F0 is 10 Gb/s per lane FEC ability
b) F1 is 10 Gb/s per lane FEC requested
c) F2 is 25G RS-FEC requested
d) F3 is 25G BASE-R FEC requested
e) F4 is 100GBASE-P RS-FEC-Int requested

Bits F0 and F1 are used for 10 Gb/s per lane operation PHYs. F2 and F3 are used for resolving FEC operation for 25G PHYs. F4 is used by 100GBASE-P PHYs where RS-FEC-Int (see Clause 161) is an alternative to the default RS-FEC (see Clause 91).

73.6.5.1 FEC resolution for 100GBASE-P PHYs which include RS-FEC-Int
For 100GBASE-P PHYs which include RS-FEC-Int (see Clause 161) in addition to RS-FEC (see Clause 91), the F4 field is used to negotiate which FEC sublayer is to be used. If either PHY requests RS-FEC-Int operation then the RS-FEC-Int sublayer is enabled and the RS-FEC sublayer is disabled. Otherwise, the RS-FEC-Int sublayer is disabled and the RS-FEC sublayer is enabled.

73.6.2.5.2 FEC resolution for 25G PHYs
For 25G PHYs if neither PHY requests FEC operation in bits F2 or F3 then FEC is not enabled.
For 25GBASE-KR and 25GBASE-CR PHYs if either PHY requests RS-FEC then RS-FEC operation is enabled, otherwise if either PHY requests BASE-R FEC then BASE-R operation is enabled.

For 25GBASE-KR-S and 25GBASE-CR-S PHYs, if either PHY requests RS-FEC or BASE-R FEC then BASE-R operation is enabled. This is because 25GBASE-KR-S and 25GBASE-CR-S PHYs do not support RS-FEC operation.

Detailed Implementation - 11

73.6.2.5.3 FEC resolution for 10 Gb/s per lane PHYs

For 10 Gb/s per lane operation, when the FEC ability bit F0 is set to logical one, it indicates that the PHY has FEC ability (see Clause 74). When the FEC requested F1 bit is set to logical one, it indicates a request to enable FEC on the link.

Since the local device and the link partner may have set the FEC capability bits differently, the priority resolution function is used to enable FEC in the respective PHYs. The FEC function shall be enabled on the link if 10GBASE-KR, 40GBASE-KR4, 40GBASE-CR4, or 100GBASE-CR10 is the HCD technology (see 73.7.6), both devices advertise FEC ability on the F0 bits, and at least one device requests FEC on the F1 bits; otherwise FEC shall not be enabled.

73.6.2.5.4 FEC control variables

The variable an_baser_fec_control indicates that BASE-R FEC operation has been negotiated. If the value is false, then BASE-R FEC has not been negotiated. If the value is true, then BASE-R FEC has been negotiated.

The variable an_rs_fec_control indicates that RS-FEC operation has been negotiated. If the value is false, then RS-FEC has not been negotiated. If the value is true, then RS-FEC has been negotiated.

The variable an_rs_fec_int_negotiated_control indicates that RS-FEC-Int operation has been negotiated. If the value is false, then RS-FEC-Int has not been negotiated. If the value is true, then RS-FEC-Int has been negotiated.

The mapping of these variables to MDIO register bits is defined in Table 73–6.

If mr_autoneg_enable (see 73.10.1) is false, the FEC function is controlled by implementation-dependent means.

Detailed Implementation - 12

73.6.2.6 Pause Ability

Pause (C0:C1) is encoded in bits D11:D10 ~~of the base link codeword~~. The two-bit Pause is encoded as follows:

a) C0 is the same as PAUSE as defined in Annex 28B

b) C1 is the same as ASM_DIR as defined in Annex 28B

The Pause encoding is defined in Clause 28B.2, Table 28B–2. The PAUSE bit indicates that the device is capable of providing the symmetric PAUSE functions as defined in Annex 31B. The ASM_DIR bit indicates that asymmetric PAUSE is supported. The value of the PAUSE bit when the ASM_DIR bit is set indicates the direction the PAUSE frames are supported for flow across the link. Asymmetric PAUSE configuration results in independent enabling of the PAUSE receive and PAUSE transmit functions as defined by Annex 31B. See 28B.3 regarding PAUSE configuration resolution.

73.6.2.7 Remote Fault

Remote Fault (RF) is encoded in bit D13 ~~of the base link codeword~~. The default value is logical zero. The Remote Fault bit provides a standard transport mechanism for the transmission of simple fault information. When the RF bit in the AN advertisement register (register 7.16.13) is set to logical one, the RF bit in the transmitted base link codeword is set to logical one. When the RF bit in the received base link codeword is set to logical one, the Remote Fault bit in the AN LP Base Page ability register (register 7.19.13) will be set to logical one, if the management function is present.

Detailed Implementation - 13

73.6.3 Link codeword Message code Next Page encoding

The link codeword Message code Next Page transmitted within a DME page shall convey the encoding shown in Figure 73–7.

Figure 73–7—Link codeword Message code Next Page

D[10:0] contains the Message Code Field, D[15:11] contains the T, Ack2, MP, Ack and NP bits. D[47:16] contains the Unformatted Code Field.

73.6.3.1 Message Code Field

Message Code Field (M[10:0]) is an eleven bit wide field, encoding 2048 possible messages. Message Code Field definitions are shown in Annex 73A. Combinations not specified are reserved for future use. Reserved combinations of the Message Code Field shall not be transmitted.

73.6.3.2 Toggle

Identical to that specified in 28.2.3.4.7

73.6.3.3 Acknowledge 2

Identical to that specified in 28.2.3.4.6

73.6.3.4 Message page

Identical to that specified in 28.2.3.4.5

73.6.3.5 Unformatted code field

The Unformatted code field (U31:0) is a 32-bit generic container. The definition for each bit depends upon the Message code Field of the link codeword. See 73A for details.

Detailed Implementation - 14

73.6.4 Link codeword Unformatted Next Page encoding

The link codeword Unformatted Next Page transmitted within a DME page shall convey the encoding shown in Figure 73–8.

Figure 73–8—Link codeword Unformatted Next Page

D[10:0] and D[47:16] contains the Unformatted Code Field, D[15:11] contains the T, Ack2, MP, Ack and NP bits.

73.6.4.2 Toggle

Identical to that specified in 28.2.3.4.7

73.6.4.3 Acknowledge 2

Identical to that specified in 28.2.3.4.6

73.6.4.4 Message page

Identical to that specified in 28.2.3.4.5

73.6.4.5 Unformatted code field

The Unformatted code field (U42:0) is a 43-bit generic container. The definition for each bit depends upon the link codeword Message code Next Page that preceded the link codeword Unformatted Next Page. See 73A for details.

Detailed Implementation - 15

73.7 Receive function ~~requirements~~

<no change other than section title>

73.7.1 DME page reception

<no change>

73.7.2 Receive switch function

<no change>

73.7.3 Link codeword matching

<no change>

73.7.4 Arbitration function requirements

<no change>

73.7.5 Renegotiation function

<no change>

Detailed Implementation - 16

73.7.6 Priority Resolution function

Since a local device and a link partner may have multiple common abilities, a mechanism to resolve which mode to configure is required. The mechanism used by Auto-Negotiation is a Priority Resolution function that predefines the hierarchy of supported technologies. The single PHY enabled to connect to the MDI by Auto-Negotiation shall be the technology corresponding to the bit in the Technology Ability Field (see 73.6.2.4) and the Extended Technology Ability Field (see 73A.1a) that is common to the local device and link partner that has the highest priority as defined in Table 73–5 (listed from highest priority to lowest priority). Priority Resolution function can only select technologies from the Extended Technology Ability Field when a link codeword Message code 2 Next Page is both transmitted and received.

Table 73-5 Priority Resolution

The common technology is referred to as the highest common denominator, or HCD, technology. If the local device receives a Technology Ability Field or the Extended Technology Ability Field with a bit set that is reserved, the local device shall ignore that bit for priority resolution. Determination of the HCD technology occurs on entrance to the AN GOOD CHECK state. In the event that a technology is chosen through the parallel detection function, that technology shall be considered the highest common denominator (HCD) technology. In the event that there is no common technology, HCD shall have a value of "NULL", indicating that no PHY receives link_control=ENABLE and link_status[HCD]=FAIL.

NOTE—If both local device and link partner are Backplane Ethernet compliant PHYs, then both ends use abilities exchanged through Clause 73 Auto-Negotiation function. If the Link partner is a legacy device (or has disabled Auto-Negotiation) as indicated by the parallel detect function, then the peer 1 Gb/s devices can opt to use abilities exchanged through Clause 37. This will ensure there are no interoperability issues when connected to a Backplane Ethernet PHY.

Detailed Implementation - 17

~~73.7.7 Next Page Function~~

Detailed Implementation - 18

73A.1a Message code 2 – Extended FEC and Technology Ability Message code

The Extended FEC and Technology Ability Message code indicates additional abilities that were not accommodated in the link codeword Base Page (see 73.6.2). Extended Technology Ability bits EA0:27 map to bits D16:D43 (U0:U27) ~~in the next page~~ and Extended FEC capability bits EF0:3 map to bits D44:D47 (U28:31).

See Table 73A–1a for more details of the Extended Technology Ability Field bit definitions.

Table 73A-1a Extended Technology Ability

For message code 2 next pages, ACK2 is set to zero for transmitted next pages and ignored in received next pages.

The 32-bit unformatted code field of link codeword Message code 2 Next Page are saved in read only management registers 7.52 to 7.55 (see 45.2.7.13a and 45.2.7.13b).

Conclusion

- Adopting the modifications to Clause 73 described on the previous slides will help improve the description of AN, especially when utilizing Next pages.