Optical auto-negotiation (OAN) baseline

Matt Brown, Alphawave Semi
Kent Lusted, Intel
Gary Nicholl, Cisco
Ali Ghiasi, Ghiasi Quantum
Ulf Parkholm, Ericsson

Supporters

Mike Dudek, Marvell

Introduction

- There has been a series of contributions proposing methods and reasons to specify optical auto-negotiation for optical PHYs.
- An accompanying presentation provides some background on OAN.
  - https://www.ieee802.org/3/dj/public/24_05/brown_3dj_03_2405.pdf
- Another presentation proposes a baseline for optical link training which is a different topic.
  - https://www.ieee802.org/3/dj/public/24_05/ghiasi_3dj_01_2405.pdf
- This contribution gathers proposals from several previous contributions with the intent of proposing a baseline specification for OAN.

Previous presentations proposing specifications for OAN

- 2023 November Plenary Meeting
  - https://www.ieee802.org/3/dj/public/23_1128/brown_3dj_01a_2311.pdf
    Matt Brown, et al
- 2024 January Interim Meeting
  - https://www.ieee802.org/3/dj/public/24_01/brown_3dj_02_2401.pdf
    Matt Brown, et al
- 2024 March Plenary Meeting
  - https://www.ieee802.org/3/dj/public/24_03/brown_3dj_03_2403.pdf
    Matt Brown, et al

Motivation

- The primary motivation at this time for OAN is to provide a means to select between a FEC mode that includes Inner FEC (AKA FECi) and another that does not include Inner FEC (AKA FECo).
- These FEC modes are represented by one of two PHY types:
  - For Inner FEC (FECi)…
    - 200GBASE-FR1, 400GBASE-DR2-2, 800GBASE-DR4-2, 1.6TBASE-DR8-2, or 800GBASE-FR4
  - For no Inner FEC (FECo)…
    - 200GBASE-DR1, 400GBASE-DR2, 800GBASE-DR4, 1.6TBASE-DR8, or 800GBASE-FR4-500

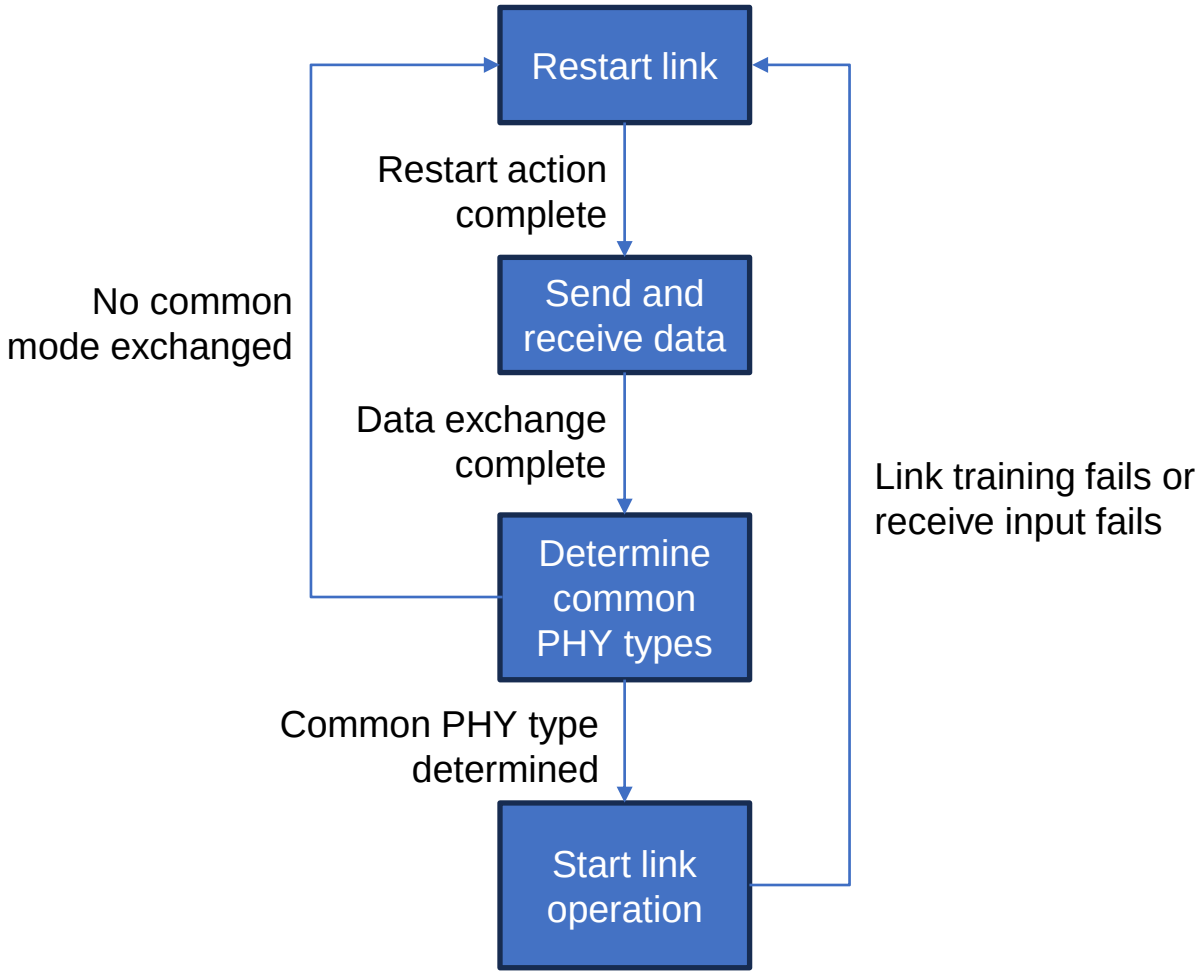
Starting point

- This motivation (previous slide) is essentially the same that inspired specification of auto-negotiation for electrical (CR, KR) PMDs defined in Clause 73.
  - brown_3dj_01a_2311 proposed that this could be repurposed for OAN
- The Clause 73 AN provides all of the features needed for OAN plus extra features to support future needs.
- Reuse of Clause 73 AN will enable significant leverage of mature designs used for electrical AN.
- No need to reinvent the wheel.

Useful features

- Means to exchange information between link partners
  - e.g., signaling and data structure
- Means to initiate automatic configuration
- Means to select which technologies to permit
- Means to select a common mode of operation
  - Also detect technology mismatch
- Means to transition from automatic configuration to data mode
- State machines to coordinate the above features

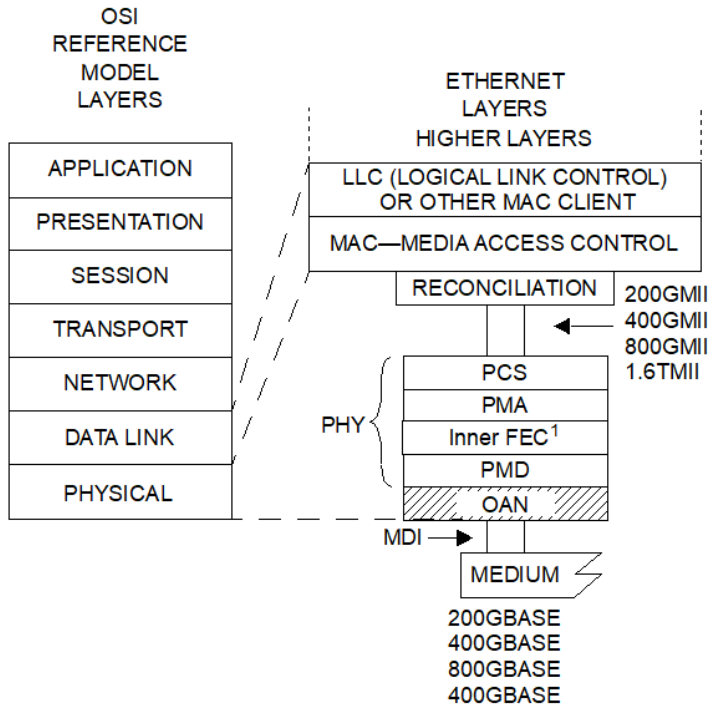
General high level state diagram



Optical OAN specification summary

- Use Clause 73 as a starting point with considerations as follows.
- Include new optical auto-negotiation (OAN) sublayer below the PMD.
- Specify OAN is optional to implement and optional to use.
- Imply that OAN is co-resident with the PMD on the module, i.e., not on the host.
- Use signaling as defined in 73.5 except...
  - specify OMA requirements rather than electrical peak to peak swing
  - Specify other optical parameters as necessary, but inclusive of any anticipated PMD specifications
  - On transmitter, transmit AN over the lane that is deemed as lane 0.
  - On receiver, monitor for AN on all lanes, then use lane on which AN signal is detected. (allows for mismatch fiber connect)
- Use page structure (base and next pages) and delineation based on those defined in 73, except…
  - Specify new allocation of PHY types and capabilities to base page rather than specification in 73.6.
  - Specify new PHY prioritization table.
- Use state machines defined in Clause 73

Optical Auto-Negotiation (OAN) sublayer



200GMII = 200 Gb/s MEDIA INDEPENDENT INTERFACE¶
400GMII = 400 Gb/s MEDIA INDEPENDENT INTERFACE¶
800GMII = 800 Gb/s MEDIA INDEPENDENT INTERFACE¶
1.6TMII = 1.6 Tb/s MEDIA INDEPENDENT INTERFACE¶
AN = AUTO-NEGOTIATION¶
FEC = FORWARD ERROR CORRECTION§

MDI = MEDIUM DEPENDENT INTERFACE¶
PCS = PHYSICAL CODING SUBLAYER¶
PHY = PHYSICAL LAYER DEVICE¶
PMA = PHYSICAL MEDIUM ATTACHMENT¶
PMD = PHYSICAL MEDIUM DEPENDENT¶¶
NOTE 1—CONDITIONAL BASED ON PHY TYPE§

Figure 500–1—OAN function relationship to the ISO/IEC Open Systems Interconnection (OSI) reference model and IEEE 802.3 Ethernet model§

- OAN specified in a standalone clause or annex.
  - In the figure to the left, it is assumed that it is specified in Clause 500.
  - Based upon, but not the same as Clause 73.
- OAN sublayer is positioned below the PMD.
- Implementation of OAN is optional.
- Use of OAN is optional.
- The PHY is configured only once AN has determined a common technology (highest common denominator or HCD).

Functions and signaling

- OAN has three functions
  - OAN transmit – send advertised information to the link partner
  - OAN receive – receive advertised information from the link partner
  - OAN arbitration – determines common advertised technology

Clause 73 – link codeword structure

73.6 Link codeword encoding

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. In a DME page, D0 shall be the first bit transmitted.

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[43:21] contains the Technology Ability Field. D[47:44] contains FEC capability (see 73.6.5).

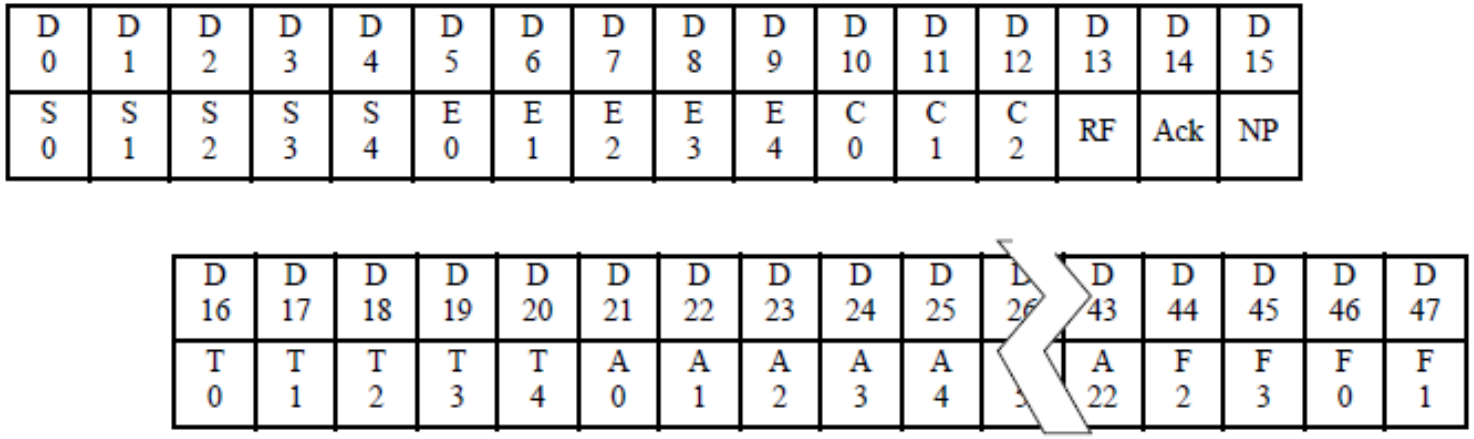


Figure 73–6—Link codeword Base Page

Link codeword fields

- Define the link codeword fields based on Clause 73 as follows:
  - The base page provides 48 bits, denoted D0 to D47 or D[0:47] of information to be exchanged.
    - Selector field S[0:4] from D[0:4] – no change
    - Echoed nonce field E[0:4] from D[5:9] – no change
    - Pause capability field C[0:2] from D[10:12] – no change
    - RF/ACK/NP fields from D(13:15) – no change
    - Transmitted nonce field T[0:4] from D[16:20] – no change
    - Technology advertisement A[0:26] from D[21:47] – this is changed
      - Define A[0:26] with technologies mapped on next slide.

Link codeword advertisement field A[0:27]

- Allocate A[0:27] to technologies as shown in the table to the right.
- A value of 1 indicates that the PHY is advertising the corresponding technology.
- It is expected that if advertising a technology, the implementation can support that PHY type.
- Although module may support a large set of PHY types, AN may advertise only a subset.
  - e.g., A particular PHY type can be forced by advertising only that PHY type.

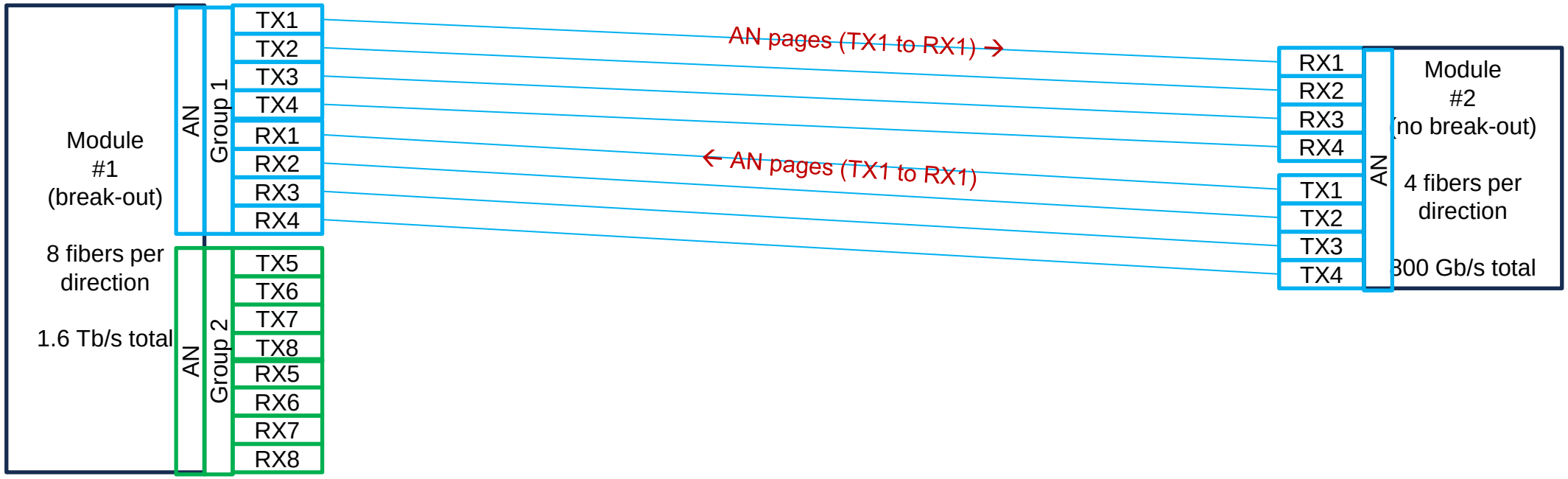
Bit	Technology	Capability
A0	1.6TBASE-DR8-2	1.6 Tb/s, 2 km, parallel
A1	1.6TBASE-DR8	1.6 Tb/s, 500 m, parallel
A2	800GBASE-FR4	800 Gb/s, 2 km, duplex
A3	800GBASE-DR4-2	800 Gb/s, 2 km, parallel
A4	800GBASE-FR4-500m	800 Gb/s, 500 m, duplex
A5	800GBASE-DR4	800 Gb/s, 500 m
A6	400GBASE-DR2-2	400 Gb/s, 2 km, parallel
A7	400GBASE-DR2	400 Gb/s, 500 m
A8	200GBASE-FR1	200 Gb/s, 2 km, parallel
A9	200GBASE-DR1	200 Gb/s, 500 m
…		
A27	Reserved	

Priority resolution table

- If the two devices, one at each end of a fiber, advertise multiple common technologies, then a mechanism to reconcile is required.
- A prioritization table is defined to determine which of a set of common technologies to select.
- A priority table, to the right, is proposed as a means to determine the highest common denominator (HCD).
- The proposed order of priority based upon the following criteria, in order of prioritization:
  - Ethernet rate: higher rate = higher priority
  - Reach: longer reach = higher priority
  - Data rate per lane: higher data rate = higher priority
  - Number of fibers (inverse): fewer fibers = higher priority
    - e.g., FR4 higher priority than DR4-2

Priority	Technology	Capability
1 highest	1.6TBASE-DR8-2	1.6 Tb/s, 2 km, parallel
2	1.6TBASE-DR8	1.6 Tb/s, 500 m, parallel
3	800GBASE-FR4	800 Gb/s, 2 km, duplex
4	800GBASE-DR4-2	800 Gb/s, 2 km, parallel
5	800GBASE-FR4-500m	800 Gb/s, 500 m, duplex
6	800GBASE-DR4	800 Gb/s, 500 m
7	400GBASE-DR2-2	400 Gb/s, 2 km, parallel
8	400GBASE-DR2	400 Gb/s, 500 m
9	200GBASE-FR1	200 Gb/s, 2 km, parallel
10 lowest	200GBASE-DR1	200 Gb/s, 500 m

Example of technology resolution



Scenario	Module #1 advertises on TX1 (TX "lane 0")	Module #2 advertises on TX1 (TX "lane 0")	Common advertised capabilities	HCD selected
1	800GBASE-DR4-2 (A3) 800GBASE-DR4 (A5)	800GBASE-DR4-2 (A3) 800GBASE-DR4 (A5)	800GBASE-DR4-2 800GBASE-DR4	800GBASE-DR4-2 (based on priority table)
1	800GBASE-DR4-2 (A3)	800GBASE-DR4-2 (A3) 800GBASE-DR4 (A5)	800GBASE-DR4-2	800GBASE-DR4-2 (i.e., module #1 forces FECi)
2	800GBASE-DR4 (A5)	800GBASE-DR4-2 (A3) 800GBASE-DR4 (A5)	800GBASE-DR4	800GBASE-DR4 (i.e., module #1 forces FECo)
3	800GBASE-DR4-2 (A3)	800GBASE-DR4 (A5)	No common capability	MISMATCH ERROR

Use of “next pages”

- “next pages” provide extra bits to expand beyond the base page.
- Any number of “next pages” may be used.
- These can be used for various reasons:
  - Not enough bits on base page; use for forward compatibility.
  - Use for vendor specific purpose outside the scope of IEEE.
- For OAN, include “next pages” as specified in 73.7.7 for forward compatibility and vendor specific use.

Clause 73 – AN signaling

Table 73–1—DME electrical characteristics

Parameter	Value	Units
Transmit differential peak-to-peak output voltage	600 to 1200	mV
Receive differential peak-to-peak input voltage	200 to 1200	mV

Table 73–2— DME page timing summary

	Parameter	Min.	Typ.	Max.	Units
T1	Transition position spacing (period)	3.2 –0.01%	3.2	3.2 +0.01%	ns
T2	Clock transition to clock transition	6.2	6.4	6.6	ns
T3	Clock transition to data transition (data = 1)	3.0	3.2	3.4	ns
T4	Transitions in a DME page	51	—	100	—
T5	DME page width	338.8	339.2	339.6	ns
T6	DME Manchester violation delimiter width	12.6	12.8	13.0	ns

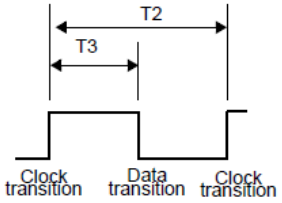


Figure 73–4—DME page transition timing

Given 3.2 ns width (T3) this is equivalent to NRZ signaling rate 312.5 MBd

The encoding of data using DME bits in an DME page is illustrated in Figure 73–3.

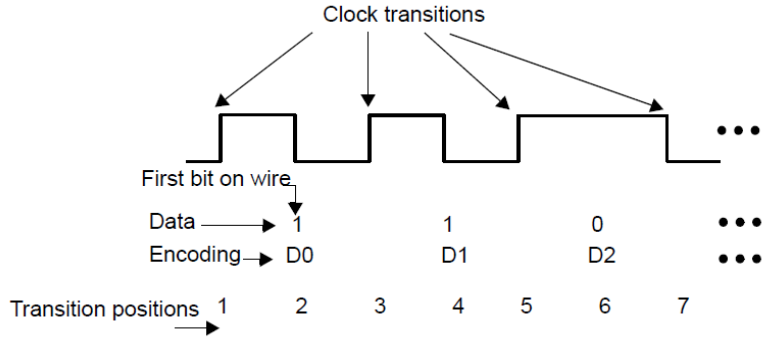


Figure 73–3—Data bit encoding within DME pages

73.5.3.1 Manchester violation delimiter

A violation is signaled as shown in Figure 73–5.

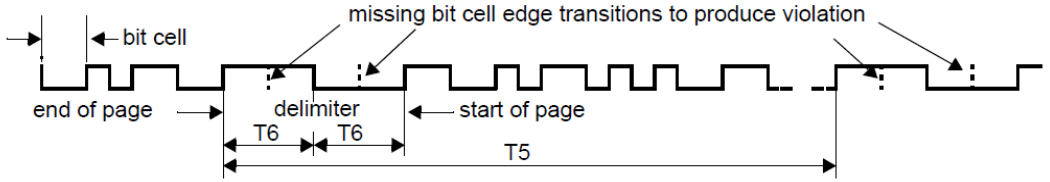


Figure 73–5—Manchester violation

OAN Signal Characteristics

- The timing requirement are the same as defined in Clause 73.
- The signal will be defined by optical rather than electrical characteristics.
  - The table below is proposed as a starting point.
  - The range of OMA should be specified to be inclusive of the ranges natural for the expected PMD types.
  - Are there other optical parameters that must be constrained?

Parameter	Value	Unit
Transmit OMA	TBD to TBD	dBm
Receive OMA	TBD to TBD	dBm

Clause 73 AN – State diagrams – transmitter and receiver

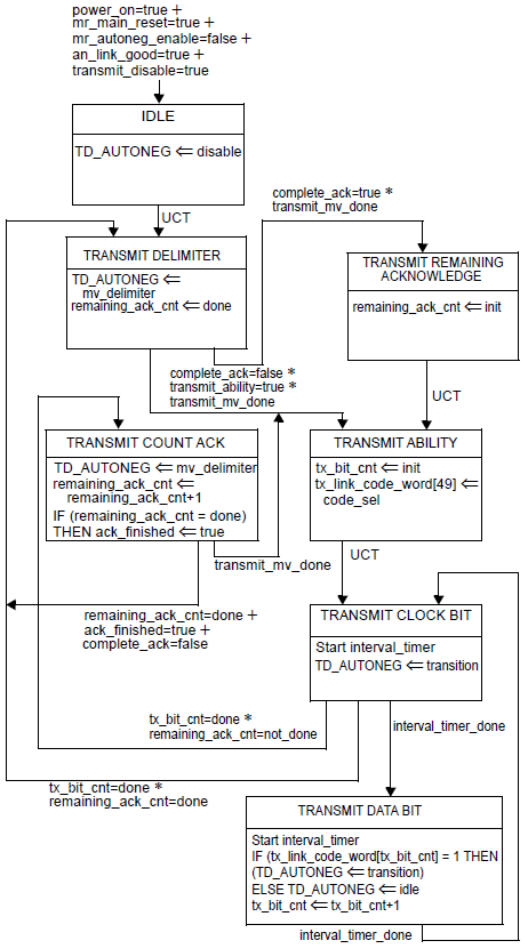


Figure 73–9—Transmit state diagram

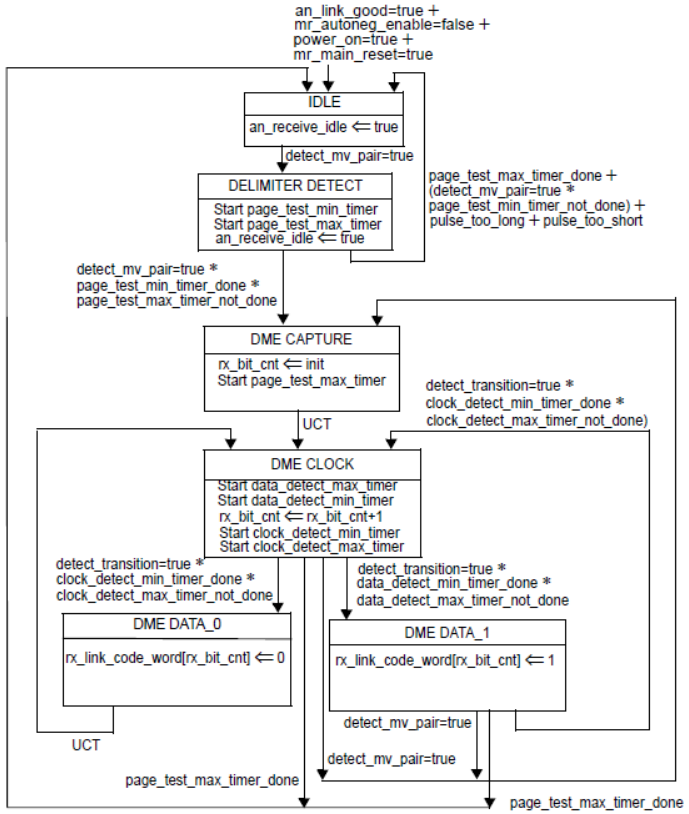


Figure 73–10—Receive state diagram

Clause 73 AN – arbitration

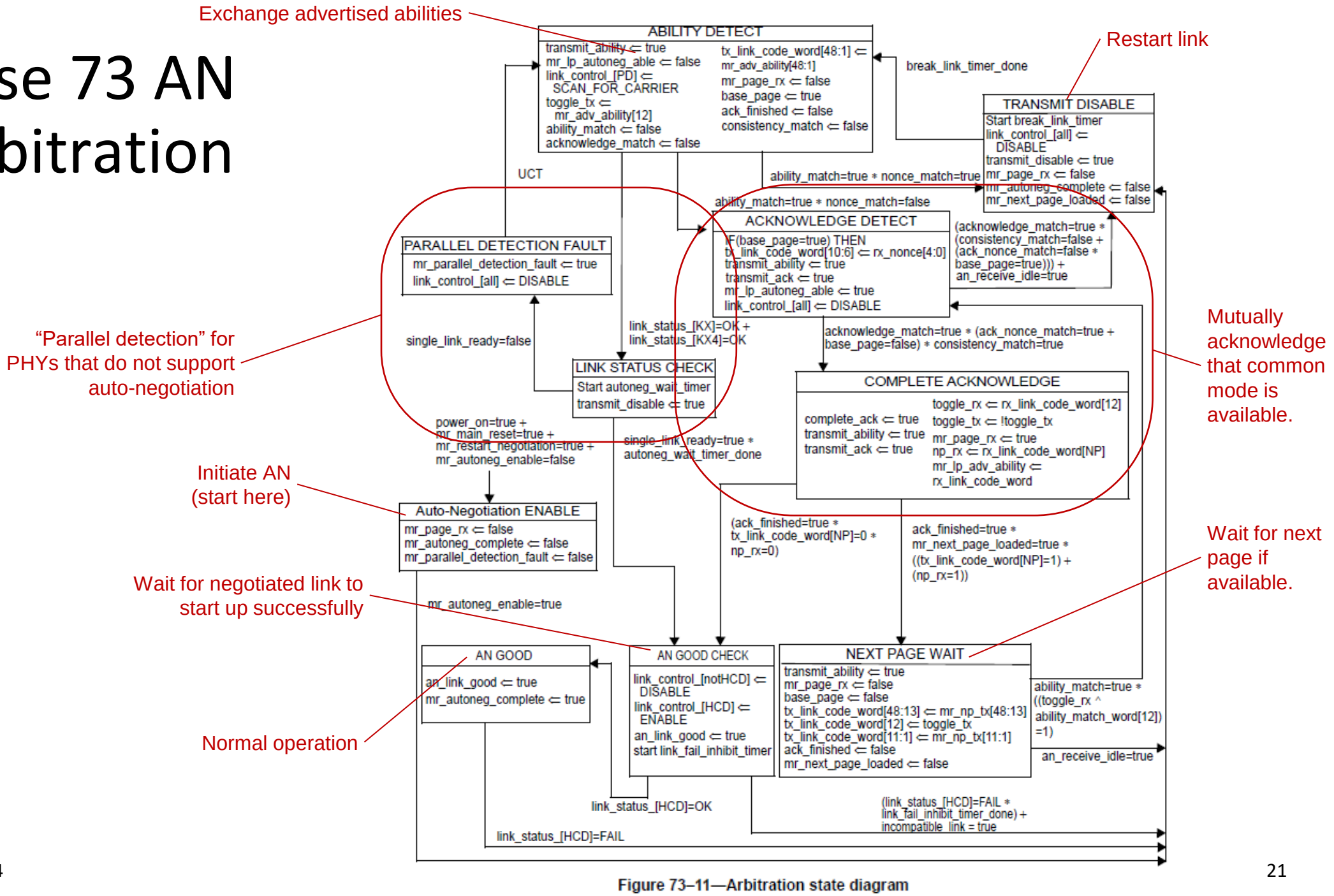


Figure 73–11—Arbitration state diagram

State Diagrams

- The transmit and receive state diagrams manage the pattern generation and detection
  - For OAN use the state diagrams in Figure 73-9 and Figure 73-10 without modification.
- The arbitration state diagram manages AN as a whole including initializing and restarting the link, interpretation of the pages.
  - For OAN use the state diagram in Figure 73-11 except
    - link_fail_inhibit_timer value is TBD.
    - link_status is TBD
      - Might be the PCS status provided by the host or might be a parameter from the segment by segment link training (or both).

Example AN High-Level Process

- Host picks Ethernet rate: e.g., 800GE (same number of fibers for both PHY types)
- Host configures the AUI or AUIs for 800G
- Host configure the module optics:
  - Configures for OAN
  - Configures advertised PHY types (e.g., 800GBASE-DR4 and 800GBASE-DR4-2)
  - Initiates module PHY operation
- Module actions after host passes the baton
  - Sends AN pages
  - Receive AN pages
  - Determine HCD
  - If HCD, then module configures itself for that HCD PHY type; otherwise AN restarts
  - Starts the optical PMD operation
  - Going to AN_good requires some confirmation that the link is working (TBD)
    - could be provided by host or by link training
  - On fault AN state machine restarts
    - Could be provided by host or by local Inner FEC state (if Inner FEC PHY detects loss of inner FEC synchronization)

Summary

- This contribution proposes a specification for optical auto-negotiation of PHY types with future expandability.
- This proposal may be adopted as a starting point for optical auto-negotiation for the following PHY types:
  - 200GBASE-FR1 and 200GBASE-DR1
  - 400GBASE-DR2-2 and 400GBASE-DR2
  - 800GBASE-DR4-2 and 800GBASE-DR4
  - 1.6TBASE-DR8-2 and 1.6TBASE-DR8
  - 800GBASE-FR4 and 800GBASE-FR4-500

Thanks