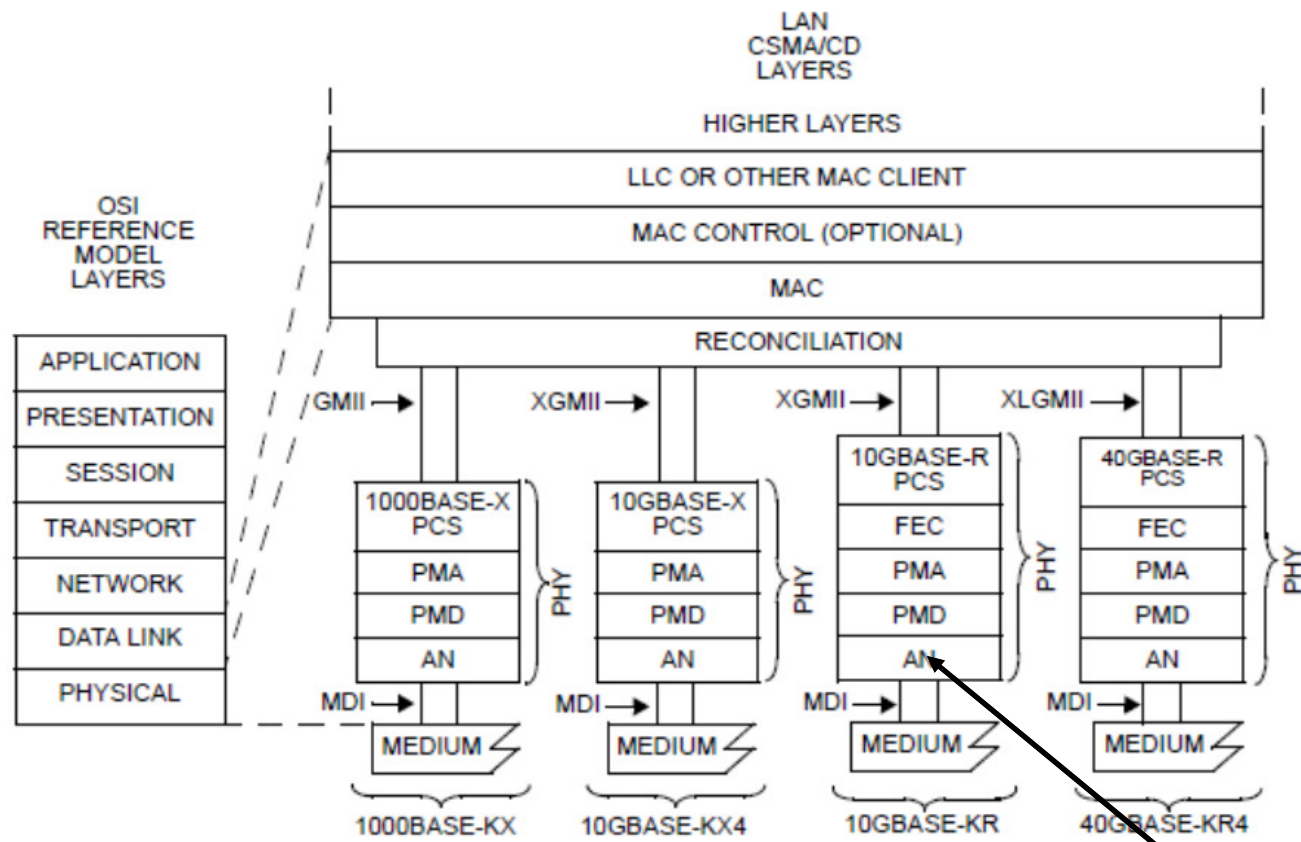


IEEE 802.3 25G Ethernet SG –
Auto-Negotiation Considerations
A thought-starter on AN.

Eric Baden (Broadcom), presenting
Cedrik Begin (Cisco),
Brad Booth (Microsoft),
Yong Kim (Broadcom),
Gary Nicholl (Cisco),

CL73 AN REVIEW - LAYERs



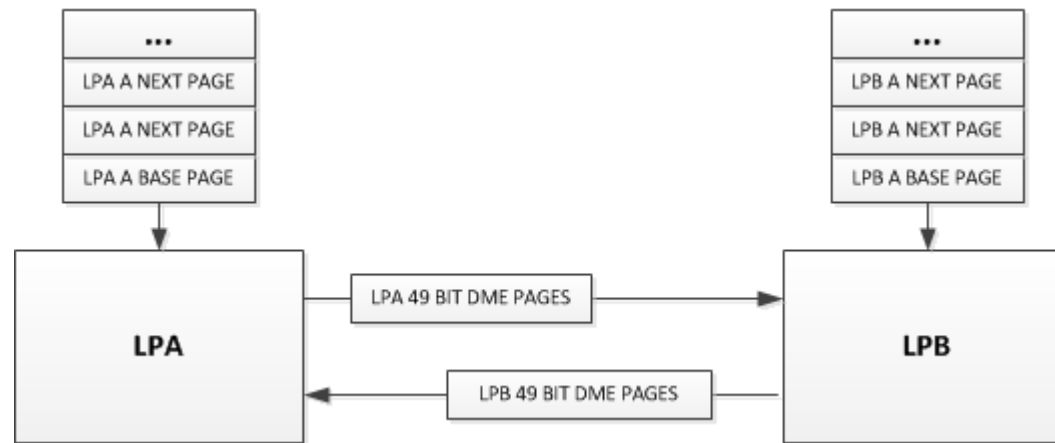
IEEE CL73 AN

CL73 AN REVIEW - FLOW

- AN is a per PORT function.
- To bring up a PORT, the following 'FLOW is followed:
 1. The Link Partners (LPs) exchange (or advertise) abilities and other information.
 2. The LPs resolve to a common Speed (HCD), FEC selection, etc., based on the information exchanged.
 3. Link Training (CL72) is performed for CR/KR Speeds.
 4. The PCS is brought up at the HCD, with the optional (CL 74) FEC selected.

CL73 AN REVIEW – PAGES

- Link Partners (LPs) exchange abilities and other information by sending CL73 PAGES.



- CL73 PAGES are DME (Differential Manchester Encoded) encoded (~1/33 - 10G rate). Transitions are slow enough to work even before training, and fast enough to not have AC coupling issues.
- All** LPs send the **BASE** Page.
- Next** Pages are **optional**.

CL73 – Link Codeword Base Page (Fig 73-6)

D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
S0	S1	S2	S3	S4	E0	E1	E2	E3	E4	C0	C1	C2	RF	Ack	NP

Selector S<4:0>
 00001 == IEEE 802.3

Echoed Nonce E<4:0>
 0 if ACK == 0, else Echoed T<4:0>

Pause Capable C<1:0>
 C2 reserved

RF = Remote Fault
 Ack = Acknowledge
 NP = New Page

D16	D17	D18	D19	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31
T0	T1	T2	T3	T4	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
					1G KX	10G KX4	10G KR	40G KR4	40G CR4	100G CR10	100G KP4	100G KR4	100G CR4		

Tx Nonce T<4:0>
 [P-]Random # per Ability Detect

D32	D33	D34	D35	D36	D37	D38	D39	D40	D41	D42	D43	D44	D45	D46	D47
A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24	F0	F1

F0 = [CL74] FEC Ability
 F1 = [CL74] FEC Request

25G Ethernet Study Group – Auto-Negotiation

CL73 AN REVIEW – BASE PAGE

- The BASE Page format is as follows.
 - Speed Abilities: Twenty-five bits allocated. Nine bits are used.
 - [CL74] FEC Abilities: Two bits allocated. Two bits are used. (CL73.6)
 - D46, F0, Bit 0: FEC Ability (FEC is available)
 - D47, F1, : FEC Request (FEC is requested)
 - If both LPs advertise the FEC Ability, and EITHER LP requests the FEC, it is enabled.**
 - NOTE: The FEC Abilities only apply to ports with 10G Links.
 - 10GKR, 40GKR4, 100GCR10
 - RS FEC is mandatory for 100GKR4.
 - Other Abilities (Pause, Fault, ...)
 - The BASE Page may indicate a Next Page follows.

D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
S0	S1	S2	S3	S4	E0	E1	E2	E3	E4	C0	C1	C2	RF	Ack	NP
Selector S<4:0>				Echoed Nonce E<4:0>					Pause Capable C<1:0>			RF = Remote Fault Ack = Acknowledge NP = New Page			
D16	D17	D18	D19	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31
T0	T1	T2	T3	T4	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
					1G KX	10G KX4	10G KR	40G KR4	40G CR4	100G CR10	100G KP4	100G KR4	100G CR4		
Tx Nonce T<4:0>															
D32	D33	D34	D35	D36	D37	D38	D39	D40	D41	D42	D43	D44	D45	D46	D47
A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24	F0	F1
F0 = [CL74] FEC Ability F1 = [CL74] FEC Request															

Table 73-4—Technology Ability Field encoding

Bit	Technology
A0	1000BASE-KX
A1	10GBASE-KX4
A2	10GBASE-KR
A3	40GBASE-KR4
A4	40GBASE-CR4
A5	100GBASE-CR10
A6	100GBASE-KP4
A7	100GBASE-KR4
A8	100GBASE-CR4
A6 through A24	Reserved for future technology

CL73 AN REVIEW – NEXT PAGEs

- LPs may optionally send additional (Next) page(s) to share more information, including vendor dependent information.
 - A Next Page may be ‘Message’ Page, or an Unformatted Page.
 - A Message Page is similar to the Base Page in that it has at least some, pre-defined, format to it.
 - For example: Message Page 5 (MP5) contains an Organizationally Unique Identifier (OUI) for vendor identification.
 - A Message Page may be followed by an Unformatted Page.
 - For example: The MP5 is followed by an Unformatted page.
 - Unformatted Pages allows vendors to send proprietary information defined by which vendor (OUI)

CL73 AN– 25G Needs

- Maintain the existing ‘Flow’ (outlined in earlier slide)
- Need to advertise 25G as a new speed:
- Need to advertise all FEC options for 25G:
 - NO FEC
 - CL74 FEC
 - CL91 FEC
- Minimize development efforts:
- Some observations:
 - Base Page is full, except for technology abilities
 - 25 abilities bits allocated, but 9 are used.
 - A24 (D45), next to the F(1:0) fields, is unused.
 - [CL74] FEC controls for 10G links already defined.

D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
S0	S1	S2	S3	S4	E0	E1	E2	E3	E4	C0	C1	C2	RF	Ack	NP
Selector S<4:0>					Echoed Nonce E<4:0>				Pause Capable C<1:0>			RF = Remote Fault Ack = Acknowledge NP = New Page			
D16	D17	D18	D19	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31
T0	T1	T2	T3	T4	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Tx Nonce T<4:0>					1G KX	10G KX4	10G KR	40G KR4	40G CR4	100G CR10	100G KP4	100G KR4	100G CR4		
D32	D33	D34	D35	D36	D37	D38	D39	D40	D41	D42	D43	D44	D45	D46	D47
A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24	F0	F1

F0 = [CL74] FEC Ability
F1 = [CL74] FEC Request

Table 73–4—Technology Ability Field encoding

Bit	Technology
A0	100GBASE-KX
A1	10GBASE-KX4
A2	10GBASE-KR
A3	40GBASE-KR4
A4	40GBASE-CR4
A5	100GBASE-CR10
A6	100GBASE-KP4
A7	100GBASE-KR4
A8	100GBASE-CR4
A9 through A24	Reserved for future technology

CL73 AN– 25G Considerations

- **Maintain BASE PAGE** as a sole means of advertising technology ability and FEC options.
 - Likely to preserve backward compatibility to large number of 10G implementations.
- **Possible Options** if the above is desirable.
 1. Extend FEC field into A24
 - Protect the use of CL74 FEC in current position D46 (F0), and apply to 25G use as well.
 - Redefine D47 (F1) – FEC Request, to be generic.
 - Add new FEC Ability bit, e.g. redefine { A24, F0 } to be new FEC Ability field.
 - 2 FEC Ability types, to allow for CL74, CL91.
 - Consider additional FEC ability (A23?) for future expansion.
 2. FEC type as a part of the Technology Ability Selection.
 - 25G needs two bits from Technology Ability field.
 - 25G [no FEC, CL74 FEC], and
 - 25G CL91 (Mandatory) Technology Types.
 3. Any other options? Next Page use?

D16	D17	D18	D19	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31
T0	T1	T2	T3	T4	A0 1G KX	A1 10G KX4	A2 10G KR	A3 40G KR4	A4 40G CR4	A5 100G CR10	A6 100G KP4	A7 100G KR4	A8 100G CR4	A9	A10
Tx Nonce T<4:0>															
D32	D33	D34	D35	D36	D37	D38	D39	D40	D41	D42	D43	D44	D45	D46	D47
A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24	F0	F1

F0 = [CL74] FEC Ability
 F1 = [CL74] FEC Request

Note: Protect the implicit use of CL91 RS-FEC in 100G if option 1 is considered.

25G Ethernet Study Group – Auto-Negotiation

One more thing...

Priority Resolution (CL 73.7.6)

- Consequence of adding 25G to AN.
 - Insertion of 25G into Priority table – straight forward.

Priority	Technology	Capability
1	100GBASE-CR4	100 Gb/s 4 lane, highest priority
2	100GBASE-KR4	100 Gb/s 4 lane
3	100GBASE-KP4	100 Gb/s 4 lane
4	100GBASE-CR10	100 Gb/s 10 lane
5	40GBASE-CR4	40 Gb/s 4 lane
6	40GBASE-KR4	40 Gb/s 4 lane
<u>7</u>	<u>25GBASE-??</u>	<u>25 GB/s 1 lane</u>
8 7	10GBASE-KR	10 Gb/s 1 lane
9 8	10GBASE-KX4	10 Gb/s 4 lane
10 9	1000BASE-KX	1 Gb/s 1 lane, lowest priority

Hopefully, this promotes further (lively!) discussions 😊

THANK YOU!