

Auto-negotiation not needed for FEC or for front-panel ports

Piers Dawe

Avago Technologies

Supporters

David Cunningham Avago Technologies

Background

- In addition to the normal MAC-to-MAC communication function, Backplane Ethernet has three additional functions
 - Auto-negotiation
 - Training
 - Parallel Detection
- Optical PMDs have never had these features
- Do not wish to burden a host that might support e.g.
 - 40GBASE-SR4 and 40GBASE-CR4 (and future 40GBASE-LR2) in the same socket for small un-retimed modules
 - or 100GBASE-SR10 and 100GBASE-CR10 in the same socket for small un-retimed modules
 - or 40GBASE-SR4, 40GBASE-LR4, and 40GBASE-CR4 in the same (XLAUI) socket
 - or 100GBASE-SR10, 100GBASE-LR4, 100GBASE-ER4 and 100GBASE-CR10 in the same (CAUI) socket

Auto-negotiation 1/2

- Clause 37 AN has a bad reputation so 802.3ap wrote a new Clause 73 AN
- Uses differential Manchester encoding at 3.2 ns (312.5 MBd), a common factor of the 1000BASE-X, 10GBASE-X and 10GBASE-R (lane) signalling rates
- Quite heavy-duty signalling scheme with frame formats, state machines and so on
- Used to advertise "technology ability" (1000BASE-KX and/or 10GBASE-KX4 and/or 10GBASE-KR) and FEC and Pause ability
- The hierarchy of "technologies" is predefined
- Highest "technology" available at both ends is chosen

Auto-negotiation 2/2

This heavy-duty protocol does just three things

1. Choosing signalling speed

- Usually not relevant for most optical PMDs which have incompatible power levels
- There is only one(?) front-panel choice to be made: 10GBASE-CX4 vs. 40GBASE-CR4
- Not necessary; Fibre Channel use a much simpler scheme, similar to "Parallel detection", to choose between e.g. 2GFC, 4GFC, 8GFC. See later

2. Choosing whether to use FEC

- Interworking between FEC-enabled ports and ports without FEC can be done by "Parallel detection" or the Fibre Channel method
 - FEC encoding should be applied always, for CR n
 - Deciding whether to just detect FEC errors (low power, low latency, preserves MTTFPA) or detect and correct (provides excellent BER and link monitoring on worse link) can be done autonomously by the receiver. See other presentation.

3. Advertising Pause ability

- Do not see the point of this; not a PMA/PMD feature at all; could be done after the link is up with Slow Protocol frames (same signalling method that Pause uses)
- Most optical ports don't advertise Pause at all(?)
- **Requires a CDR that can work at 1/33 of usual signalling rate**; ordinary CDRs won't
- **Auto-negotiation is complicated and onerous**
 - **Unnecessarily and should not be used on front-panel ports**
 - **Should not be compulsory even on backplane ports**

Auto-negotiation 3/3

- diminico_02_0708 slide 12 says
 - "Auto-Neg allows backward compatibility with legacy 10 GbE CX4 PHYs"
 - But legacy CX4 PHYs don't have Clause 73 AN.
Don't want to have to support Clause 37 AN as well!
- diminico_02_0708 slide 12 also says
 - "Clause 73 provides parallel detection function for compatibility with legacy PHYs that do not support Auto-Negotiation
 - New 40 GbE PHY can use parallel detection for auto-detection of legacy CX4 devices"

Training

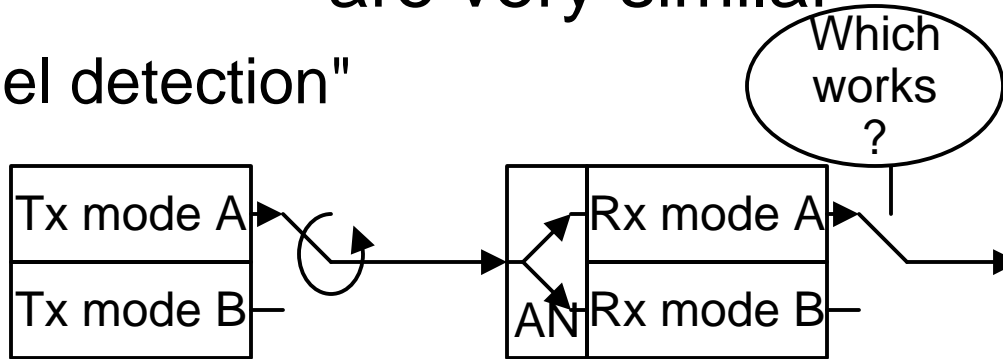
- Provided to allow receiver equalizer to train itself on a signal so distorted that the receiver might not learn what to do on a regular scrambled signal
- Defined in 72.6.10.2
- Uses a full-rate (10.3125 GBd) PRBS11 and a differential Manchester encoded handshaking at 1/4 the Baud rate
- Allows a receiver to ask its peer transmitter to change its emphasis (and amplitude?) during the training phase, before link starts up
- 10GBASE-LRM does not use such a thing
- CR n channels (cables) expected to be more consistent than backplane channels; should be able to set emphasis for a long cable and it will be fine with a short cable
- Why not just transmit the best signal to start with?
 - If wished, could handshake using Slow Protocol frames to turn the transmitter down after link has started
- **No evidence seen that training is needed for CR n**

Parallel detection

- See 73.7.4.1
- Seems to involve a receiver listening to a signal and trying to match it to the kinds of signal it can receive (we assume its transmitter can transmit the same kinds)
- This is the first step towards the Fibre Channel link startup method, where both transmitter and receiver try different kinds

"Parallel detection" and "Link Speed Negotiation" are very similar

- "Parallel detection"

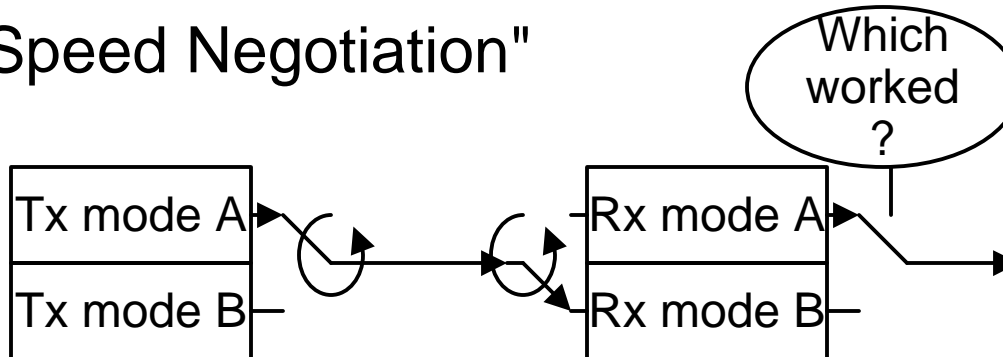


(I think)

- Are the start-up timings defined?

See next slide for details

- "Link Speed Negotiation"



AAAABBBBAAAABBBBAAAABBBB

ABABABABABABABABABABABAB

- Can negotiate coding, not just speed. Can be physically same Tx, Rx
- Transmitter cycles through its options slowly, receiver quickly
- Cycle rates (range) and total sequence time to be specified (FC has values for 8B/10B coding)

Fibre Channel way

- See FC-FS-2 Clause 26, Link Speed Negotiation*
- Transmitter cycles slowly through up to three or four transmit modes
 - In FC these would be signalling rates, e.g. 2G, 4G, 8G
- Receiver cycles more quickly through its receive modes
- Timings are defined so that at least four receive modes (if available) are tried against every transmit phase
 - We would choose timing limits appropriate to our situation
- Transmitter cycles round three? times
- At the end, each side transmits in the highest-priority mode that it successfully received
- Can be used for choosing signalling speed
 - 10GBASE-CX4 vs. 40GBASE-CR4
- Or for choosing whether to start with FEC on and off
 - CDR remains in lock all the time, receive PCS/FEC tries to parse the signal different ways e.g. by looking for lane markers
- **Provides good long-term structure for standard** for possible features in the future
 - e.g. distinguishing between a single 10GBASE-R lane and one in a 40GBASE-R4 or 100GBASE-R10 group
- **A port with only one mode doesn't need to do anything special:** it transmits the only way it knows how and receives the only way it knows how
 - Only requirement is that its signal detect, lock detect and other behaviours do not go mad when receiving a cycling signal
 - Can recommend timing limits to make this simple
- * <http://www.t11.org/ftp/t11/pub/fc/fs-2/06-085v3.pdf>

Conclusion

- For front panel ports (-SR n , LR n , ER4, CR n)
- Auto-negotiation is a sledgehammer to crack a nut
 - Do not use AN on front panel ports
 - Have it no higher than optional for 40GBASE-KR4
- The value of Training is unproven
 - Establish if it has value
 - If it increases cable length by e.g. just 10%, don't use it
 - If it has value for some, make it optional for CR n , not used for optical
- "Parallel detection" is a simple form of Fibre Channel's Link Speed Negotiation
- Use optional Link Speed Negotiation method to allow interoperation with legacy PHY types
- Use Link Speed Negotiation method to allow interoperation with non-FEC ports
- Non-FEC ports need none of these; avoid proliferating complexity