Thoughts on optical automatic link configuration

Matt Brown, Alphawave Semi
Ali Ghiasi, Ghiasi Quantum
Mike Dudek, Marvell
Kent Lusted, Intel
Contributors

• Adam Healey, Broadcom
Supporters

• Chris Cole, Quintessent
• Tony Chan Carusone, Alphawave Semi
• Roberto Rodes, Coherent
• Ulf Parkholm, Ericsson
Introduction

- There has been interest expressed in providing a mechanism to automate the selection of PHY type and/or PHY mode.
- Auto-negotiation methodology developed for Ethernet electrical backplane and copper cable would be a good candidate as basis.
- Although we adopted no optical PHY baselines, the proposal encompasses any of the proposals discussed so far.
- Complementary presentation ghiasi_3dj_01_2311 proposes a method for optical transmitter adaptation.
Purpose

• Establish common configuration (or lack thereof) for link partners at each end of an optical fiber.

• Examples...
  • Determine common PHY type
    • both ends use PHY with inner FEC or both use PHY without inner FEC
  • Common inner FEC mode
    • both ends use inner FEC or both ends disable/bypass inner FEC
  • For inner FEC case
    • both ends select with convolution interleaver or both ends do not use
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
- Means to transition from automatic configuration to data mode
- State machines to coordinate the above features
High level state diagram

- **Restart link**
  - **Restart action complete**
  - **Send and receive data**
  - **Data exchange complete**
  - **Determine common modes**
  - **Common mode(s) determined**
  - **Start link operation**

- **No common mode exchanged**
- **Link training fails or receive input fails**
Summary of auto-negotiation defined for backplane and copper cable PHYs in Clause 73 of IEEE 802.3-2022 and IEEE Std 802.3ck-2022.
Clause 73 – AN general
Clause 73 – AN signaling

Given 3.2 ns width (T3) this is equivalent to NRZ signaling rate 312.5 MBd
Clause 73 – link codeword— encoding

73.6 Link codeword encoding


![Figure 73–6—Link codeword Base Page](image)

73.6.1 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 25A. 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>
<thead>
<tr>
<th>S4</th>
<th>S3</th>
<th>S2</th>
<th>S1</th>
<th>S0</th>
<th>Selector description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>IEEE Std 802.3</td>
</tr>
</tbody>
</table>

73.6.2 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.

73.6.3 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 table. The method of generating the nonce is left to the implementers. 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.
Clause 73 – link codeword – technology abilities field

From 802.3-2022...

73.6.4 Technology Ability Field

Technology Ability Field (A[22:0]) is a 23-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.

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.


The fields A[22:16] are reserved for future use. Reserved fields shall be set as zero and ignored on receive.

From 802.3ck-2022 (additions)...
Clause 73 AN – Resolution of FEC type

From 802.3ck-2022...

73.6.5 FEC capability

Change 73.6.5 as follows:

FEC (F4,F2, F3, F0, F1) is encoded in bits D4+D3+D47 of the base link codeword. The four-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 F2 and F3 are used for resolving FEC operation for 25G PHYs, while bits F0 and F1 are used for 10 Gb/s per lane operation. Bits F0 and F1 are not used for 25G PHYs.

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).

Insert 73.6.5.a before 73.6.5.1 as follows:

73.6.5.1 a FEC resolution for 100GBASE-P PHYs that include RS-FEC-Int

For 100GBASE-P PHYs that 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.

From 802.3-2022...

73.6.5.1 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.

73.6.5.2 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.
Clause 73 AN – Resolution of PHY type

Table 73-5—Priority Resolution

<table>
<thead>
<tr>
<th>Priority</th>
<th>Technology</th>
<th>Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100BASE-KR4 or 100BASE-CR4</td>
<td>40 Gb/s 4 lane, highest priority</td>
</tr>
<tr>
<td>2</td>
<td>200BASE-KR2 or 200BASE-CR2</td>
<td>200 Gb/s 2 lane</td>
</tr>
<tr>
<td>3</td>
<td>200BASE-KR4 or 200BASE-CR4</td>
<td>200 Gb/s 4 lane, highest priority</td>
</tr>
<tr>
<td>4</td>
<td>100BASE-KR1 or 100BASE-CR1</td>
<td>100 Gb/s 1 lane</td>
</tr>
<tr>
<td>5</td>
<td>100BASE-KR2 or 100BASE-CR2</td>
<td>100 Gb/s 2 lane</td>
</tr>
<tr>
<td>6</td>
<td>100BASE-CR4</td>
<td>100 Gb/s 4 lane</td>
</tr>
<tr>
<td>7</td>
<td>100BASE-KR4</td>
<td>100 Gb/s 4 lane</td>
</tr>
<tr>
<td>8</td>
<td>100BASE-KP4</td>
<td>100 Gb/s 4 lane</td>
</tr>
<tr>
<td>9</td>
<td>100BASE-CR10</td>
<td>100 Gb/s 10 lane</td>
</tr>
<tr>
<td>10</td>
<td>50BASE-KR or 50BASE-CR</td>
<td>50 Gb/s 1 lane</td>
</tr>
<tr>
<td>11</td>
<td>40BASE-CR4</td>
<td>40 Gb/s 4 lane</td>
</tr>
<tr>
<td>12</td>
<td>40BASE-KR4</td>
<td>40 Gb/s 4 lane</td>
</tr>
<tr>
<td>13</td>
<td>25BASE-KR or 25BASE-CR</td>
<td>25 Gb/s 1 lane</td>
</tr>
<tr>
<td>15</td>
<td>10BASE-KR</td>
<td>10 Gb/s 1 lane</td>
</tr>
<tr>
<td>16</td>
<td>10BASE-KX4</td>
<td>10 Gb/s 4 lane</td>
</tr>
<tr>
<td>17</td>
<td>5BASE-KR</td>
<td>5 Gb/s 1 lane</td>
</tr>
<tr>
<td>18</td>
<td>5BASE-KX</td>
<td>5 Gb/s 1 lane</td>
</tr>
<tr>
<td>19</td>
<td>10BASE-KX</td>
<td>1 Gb/s 1 lane, lowest priority</td>
</tr>
</tbody>
</table>

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 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).

The common technology is referred to as the highest common denominator, or HCD, technology. If the local device receives a 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 end 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.
Clause 73 AN – State diagrams – transmitter and receiver
Clause 73 AN – arbitration

- Exchange advertised abilities
- "Parallel detection" for PHYs that do not support auto-negotiation
- Initiate AN (start here)
- Wait for negotiated link to start up successfully
- Normal operation
- Restart link
- Mutually acknowledge that common mode is available.
- Wait for next page if available.
Optical link auto-configuration using Clause 73 auto-negotiation as basis
Reuse of Clause 73 auto-negotiation

- Include new auto-configuration sublayer below the PMD.
- Use signaling as defined in Clause 73, except...
  - edge spacing at 75.3 ps (53.125 GBd / 4 = 13.28125 GBd) rather than 3.2 ns (312.5 MBd)
  - Same as training frame control channel proposed in ghiasi_3dj_01_2311; 1/8 PAM4 symbol rate
  - Specify OMA requirements rather than electrical peak to peak swing as well as other optical parameters as necessary, inclusive of any anticipated PMD specifications
- Signaling over a specific lane TBD for multi-lane PHYs, other lanes disabled.
- Use page structure (base and next pages) and delineation as defined in Clause 73.
- Use state machines defined in Clause 73.
- Specify new allocation of PHY types and capabilities to base page.
- Specify new PHY prioritization table.
- Specify selection criteria for other capabilities, e.g., FEC, interleaving.
- Should it be mandatory to implement and optional to use as it is for Clause 73?
- AN is co-resident with the PMD on the module.
  - Host may need to confirm PCS status to module or PCS monitor on module. Some coordination between host and module is required regardless.
Example priority resolution table assuming two inner FEC modes (FECo/FECi) per PHY

<table>
<thead>
<tr>
<th>Priority</th>
<th>Technology</th>
<th>Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6TBASE-DR8-2</td>
<td>1.6 Tb/s, 2 km, parallel</td>
</tr>
<tr>
<td>2</td>
<td>1.6TBASE-DR8</td>
<td>1.6 Tb/s, 500 m, parallel</td>
</tr>
<tr>
<td>3</td>
<td>800GBASE-LR4</td>
<td>800 Gb/s, 10 km</td>
</tr>
<tr>
<td>4</td>
<td>800GBASE-FR4</td>
<td>800 Gb/s, 2 km, duplex</td>
</tr>
<tr>
<td>5</td>
<td>800GBASE-DR4-2</td>
<td>800 Gb/s, 2 km, parallel</td>
</tr>
<tr>
<td>6</td>
<td>800GBASE-DR4</td>
<td>800 Gb/s, 500 m</td>
</tr>
<tr>
<td>7</td>
<td>400GBASE-DR2-2</td>
<td>400 Gb/s, 2 km, parallel</td>
</tr>
<tr>
<td>8</td>
<td>400GBASE-DR2</td>
<td>400 Gb/s, 500 m</td>
</tr>
<tr>
<td>9</td>
<td>200GBASE-FR1</td>
<td>200 Gb/s, 2 km, parallel</td>
</tr>
<tr>
<td>10</td>
<td>200GBASE-DR1</td>
<td>200 Gb/s, 500 m</td>
</tr>
</tbody>
</table>

Notes:
- Separate FEC mode advertisement and priority handling is required.
# Example priority resolution table assuming single inner FEC mode (FECi/FECo) per PHY

<table>
<thead>
<tr>
<th>Priority</th>
<th>Technology</th>
<th>Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6TBASE-DR8-2 (FECi)</td>
<td>1.6 Tb/s, inner FEC, 2 km, parallel</td>
</tr>
<tr>
<td>3</td>
<td>1.6TBASE-DR8 (FECi)</td>
<td>1.6 Tb/s, inner FEC, 500 m, parallel</td>
</tr>
<tr>
<td>4</td>
<td>1.6TBASE-?R8 (FECo place holder)</td>
<td>1.6 Tb/s, reach TBD, parallel</td>
</tr>
<tr>
<td>5</td>
<td>800GBASE-LR4 (assume always inner FEC)</td>
<td>800 Gb/s, 10 km</td>
</tr>
<tr>
<td>6</td>
<td>800GBASE-FR4 (FECi)</td>
<td>800 Gb/s, inner FEC, 2 km, duplex</td>
</tr>
<tr>
<td>7</td>
<td>800GBASE-?R4 (FECo place holder)</td>
<td>800 Gb/s, reach TBD, duplex</td>
</tr>
<tr>
<td>8</td>
<td>800GBASE-DR4-2 (FECi)</td>
<td>800 Gb/s, inner FEC, 2 km, parallel</td>
</tr>
<tr>
<td>10</td>
<td>800GBASE-DR4 (FECi)</td>
<td>800 Gb/s, inner FEC, 500 m</td>
</tr>
<tr>
<td>11</td>
<td>800GBASE-?R4 (FECo place holder)</td>
<td>800 Gb/s, reach TBD</td>
</tr>
<tr>
<td>12</td>
<td>400GBASE-DR2-2 (FECi)</td>
<td>400 Gb/s, inner FEC, 2 km, parallel</td>
</tr>
<tr>
<td>13</td>
<td>400GBASE-DR2 (FECi)</td>
<td>400 Gb/s, inner FEC, 500 m</td>
</tr>
<tr>
<td>14</td>
<td>400GBASE-?R2 (FECo place holder)</td>
<td>400 Gb/s, reach TBD</td>
</tr>
<tr>
<td>15</td>
<td>200GBASE-FR1 (FECi)</td>
<td>200 Gb/s, inner FEC, 2 km, parallel</td>
</tr>
<tr>
<td>16</td>
<td>200GBASE-DR1 (FECi)</td>
<td>200 Gb/s, inner FEC, 500 m</td>
</tr>
<tr>
<td>17</td>
<td>200GBASE-?R1 (FECo place holder)</td>
<td>200 Gb/s, reach TBD</td>
</tr>
</tbody>
</table>

**Notes:**
- ?Rn is for PMDs associated with new objectives proposed in lusted_3dj_05_2311
Things to think about

- For multilane PMDs, which lane should be used for signaling?
- Should AN be mandatory to implement and optional to use?
- Support for negotiation between lane rates, future proof.
Summary

• Purpose of optical automatic configuration along with helpful features discussed.
• Clause 73 auto-negotiation, used for electrical links, reviewed.
• Optical automatic configuration using Clause 73 as a starting point outlined.
  • This could be a candidate for optical link automatic configuration.
Thanks