8B10B Transmission Code as a Link Protocol Building Block

- 8B10B transmission code provides the following functions:
  - Improves transmission characteristics
  - Enables bit-level clock recovery
  - Improves error detection
  - Separates data symbols from control symbols
  - Derives bit and word synchronization
- Data bytes are encoded into 10-bit Data Characters.
- 12 Special Characters are defined for special signaling.
- One or more Data and/or Special Transmission Characters may be grouped into Transmission Words.
- Special Transmission Words called Ordered Sets are defined.
- Ordered Sets are flexible building blocks which may be used for in-band and or out-of-band protocol functions.

Transmission Word and Ordered Set Usage

- An 8B10B-based link protocol may specify the use of Transmission Words of various lengths:
  - Fibre Channel uses only 40-bit Transmission Words
  - A pre-cursor to Fibre Channel, SBCON, uses 10, 20 and 30-bit Transmission Words
- The use of single-length transmission words somewhat simplifies link protocol component design, especially at gigabit speeds.
- Ordered Sets should be chosen for good spectral characteristics and coding distances to provide maximum immunity against link errors.
  - This is important because ordered sets are generally not included in the Frame Check Sequence (FCS).

Ordered Sets

- Fibre Channel Ordered Sets all start with the K28.5 special character followed by three Data Characters.
  - The Data Characters indicate what control information is being sent. The following types of ordered sets are defined:
    - Frame Delimiters to mark frame boundaries;
    - Primitive Signals used to signal events;
    - Primitive Sequences used to signal states.

K28.5 Special Character

- The K28.5 Special Character is chosen as the first character of all Ordered Sets for the following reasons:
  - Bits abcdeif make up a Comma. A Comma is a singular bit pattern which in the absence of transmission errors cannot appear in any other location of a Transmission Character and cannot be generated across the boundaries of any two adjacent Transmission Characters.
  - The Comma can be used to easily find and verify character and word boundaries of the received bit stream.
  - Bits ghj of the encoded character present the maximum number of transitions, simplifying receiver acquisition of bit synchronization.

MAC and PHY Layer Procedures

- Several existing and newly proposed MAC and PHY level procedures may or may not be satisfied by using 8B10B-based solutions. These include:
  - Open Fibre Control (OFC) - Required for shortwave lasers to meet Class 1 laser safety standards
  - Link Bit and Word Synchronization - Enables reliable decoding of Transmission Words
  - Auto-Negotiation
  - Link Recovery - Provides link error notification to the other end of the link and resets the link to a known state
  - Idle and Frame Delimiters
  - Pad character
  - Jam and Fill signals

MAC and PHY Layer Procedure Requirements

Procedure	Bit Sync Required	Character Recognition Required
Open Fibre Control		
Bit and Word Synchronization	√	
Auto-Negotiation	√	√1
Link Recovery	√	√
Idle, Jam, Fill	√	√
Frame Delimiters	√	√
Pad Character	√	√

√1 - **May depend on media variant**

Open Fibre Control

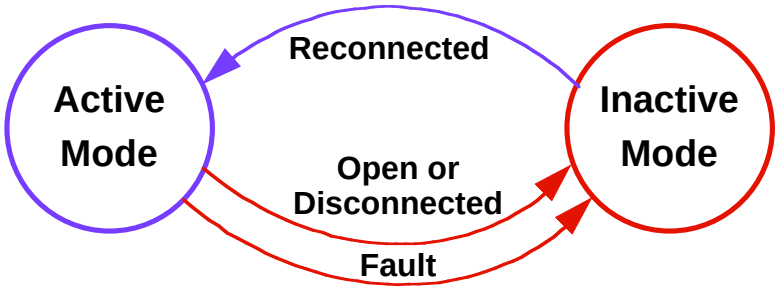
- The Ethernet PHY and MAC layers and 8B10B-based fiber-optic links require several standard (and optional) procedures to be successfully executed before MAC layer frames may be exchanged.
  - All procedures require that the PHY layers of two communicating stations transmit and receive signals which are recognized by the other end of the link.
  - Compliance with Safety regulations is required whenever a Class 1 laser transmitter is operating in order to prevent (human eye) exposure to laser radiation.
  - OFC protocol operation does not require that either bit, character or word synchronization be achieved or that characters be recognized.
- Where applicable, OFC must be the first procedure executed over a Gigabit Ethernet link.

Open Fibre Control Procedure

- Two paired fibers form one closed system.
  - The receiver of one station is connected to the transmitter of the other station.

An interruption in light means potential exposure to laser radiation.

  - A safety interlock detects the optical link interruption and shuts down the laser immediately.
  - The circuitry attempts to reconnect the link.
  - The OFC system has two states: active and inactive.



Bit and Word Synchronization

- Once OFC has determined that the link is safe to use, the PHY can concentrate on information contained in the bit stream.
- The link is not considered usable by the MAC layer until Link Initialization is complete.
  - To perform Link Initialization, a station transmits a prescribed Primitive Sequence and simultaneously attempts to acquire bit and transmission word synchronization from the received signal.
  - The Primitive Sequence transmitted varies according to whether the link is initializing, has lost synchronization while initializing, or has completed initialization.
  - Synchronization is achieved when the receiver identifies the same transmission word boundary on the received bit stream as that established by the transmitter at the other end of the link.

Synchronization States

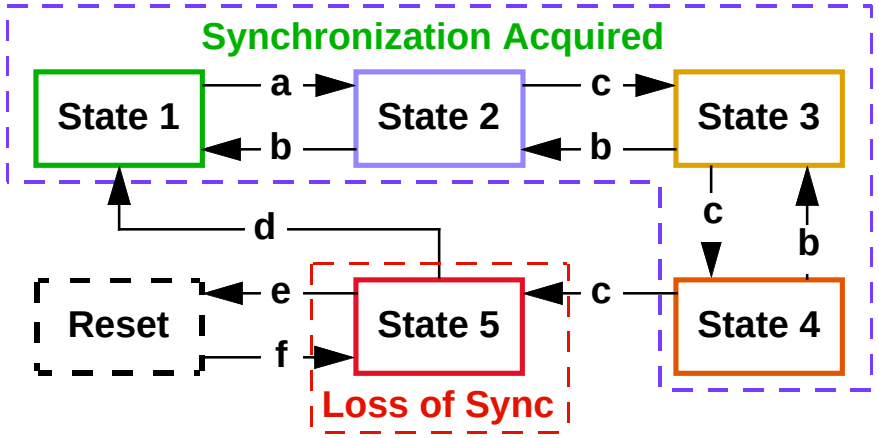
- The detection of Invalid Transmission Words is an indication that a receiver is out of synchronization. An invalid Transmission Word is recognized by the receiver when one of the following conditions is detected:
  - Code Violation is detected within a Transmission Word.
  - Special Character detected in the second or subsequent (if applicable) character in a Transmission Word (see PAD for exception).
  - Ordered Set with improper Beginning Running Disparity received.
- The clock recovery system in a receiver has sufficient "inertia" to prevent a single transmission error from causing it to go out of synchronization.
  - A Loss of Synchronization Procedure defines the method by which the receiver changes from the Synchronization Acquired state to the Loss of Synchronization State.

Running Disparity

- 8B10B code recognizes the idea of a Running Disparity (the difference between the number of 1's and 0's transmitted).
  - The sender keeps the running disparity around zero, the receiver checks the sender.
    - After powering on or exiting diagnostic mode, the transmitter assumes the negative value for its initial Running Disparity.
    - Upon transmission of any Transmission Character, the transmitter calculates a new value for its Running Disparity.
    - After powering on or exiting a special mode, the receiver assumes either a positive or negative initial Running Disparity.
    - Upon reception of any Transmission Character, the receiver determines whether the character is valid and calculates a new value for its RD.
    - The primary benefit of having all Ordered Sets end with negative RD is that it allows Idles to be removed and added from an encoded bit stream, by an inter-station unit, one word at a time without altering the Beginning RD of the immediately preceding Transmission Word.

Loss of Synchronization Procedure States

- The following five detection states are defined as part of the loss-of-Synchronization procedure. The transitions are:
  - a) The first invalid Transmission Word is detected
  - b) An additional invalid Transmission Word is not detected in the next two or fewer consecutive Transmission Words
  - c) An additional invalid Transmission Word is detected in the next two or fewer consecutive Transmission Words
  - d) The receiver regains Synchronization
  - e) The receiver is Reset
  - f) The receiver exits a previously established Reset condition



Auto Negotiation Basics

- Auto-negotiation (AN) allows an adapter, hub or switch capable of data transfer at multiple rates to automatically use the fastest rate supported by the device at the other end.
  - AN signals the capabilities it has available, detects the technology that exists in the device it is being connected to and automatically configures to the highest common performance mode of operation.
  - There are two primary benefits to AN:
    - First, the ability to automatically connect at the appropriate speed without user intervention.
    - Second, in the event that no common technology exists, it will not make a connection.
  - AN currently supports 10Base-T, 100Base-TX, 100Base-T4 or a corresponding Full Duplex mode.

Auto Negotiation at a Gigabit

- AN support for Gigabit Ethernet may be divided according to the two distinct types of media targeted, copper and fiber.
  - It is feasible to extend 802.3U clause 28 AN support for Gigabit Ethernet copper-based variants by enabling the AN function at station power-on time.
    - This function may not be useful unless there is commonality in the physical mediums between Gigabit Ethernet and legacy networks (e.g., unless a 1000Base-T4 option exists, etc.)
    - The AN Parallel Detection function may also be implemented
    - 8B10B coding would not be applicable during the AN procedure as the signaling of the AN protocol is not amenable to encoding/decoding by the transmitter/receiver, respectively.
  - AN may be specified as a separate receiver/transmitter mode of operation similar to (optional) diagnostic mode.

Gigabit Fiber-Based Auto-Negotiation

- AN support for Gigabit Ethernet fiber-based variants may be specified as follows:
  - AN protocol follows OFC protocol, if applicable.
    - OFC protocol is too slow to effectively combine with AN.
  - Non-8B10B-based Ethernet fiber-optic media types are not considered to be common technology.
    - Transmission errors will be detected and AN will not establish a connection.
  - AN protocol follows 8B10B-based Synchronization Acquisition procedure.
  - AN Arbitration may be implemented by AN Primitive Sequence procedure based on 4-character Ordered Sets.
  - The AN Base Page and Next Page data (Link Control Words) may be specified as data contained in AN Primitive Sequences.

Auto-Negotiation Primitive Sequences

Primitive Sequence	Beginning RD	Ordered Set
Link Code Word (LCW)	negative	K28.5 D2.2 D0:7 D8:15
LCW Response (LCWR)	negative	K28.5 D2.5 D0:7 D8:15

Dx:y denotes LCW data, must be converted to 8B10B Data Character

- AN Base Link Control Word encodings are specified in 802.3U subclause 28.2.1.2.
- AN Next Page encodings are specified in 802.3U subclause 28.2.3.4.1.
- The Local Device and Link Partner may execute the AN protocol simultaneously. This protocol requirement suggests the usage of a separate Response Primitive Sequence.

Auto-Negotiation Arbitration Procedure

- The Local Device begins by transmitting the LCW Primitive Sequence with the Ack bit not set.
  - Once three consecutive, matching LCW Ordered Sets are received from the Link Partner, the Local Device transmits the LCW Response Primitive Sequence with the Ack bit set to indicate that it has received the Link Partner's Link Code Word correctly.
  - Once three consecutive, matching LCW Response Ordered Sets are received from the Link Partner, the Local Device transmits the Idle Primitive Signal to complete the AN Arbitration Procedure.
- The Next Page function uses the same Auto-Negotiation Arbitration Procedure described above to exchange Next Pages.

Gigabit Fiber-Based Auto-Negotiation Futures

- 8B10B-based AN may be used to for all AN functions specified in 803.2U clause 28.
  - May be extended in the future to automatically configure to the highest common performance mode of operation in the advent of link Baud rate increases.

Link Recovery Procedures

- Link Recovery uses Primitive Sequences based on Ordered Sets used to control the status of individual links. These sequences provide link error notification to the other end of the link and reset the link to a known state.
- Not Operational (NOS)
  - The transmitting station is initiating the Link Failure protocol;
  - The transmitting station has detected a failure condition such as:
    - Loss-of-Synchronization Failure
    - Loss-of-Signal Failure
- Offline (OLS)
  - The transmitting station is:
    - Going Offline;
    - Initiating the Link Initialization protocol;
    - In the NOS Receive state, receiving and recognizing NOS.

Link Reset and Link Reset Response

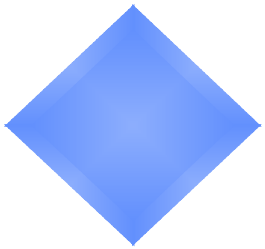
- Link Reset (LR)
  - The transmitting station is initiating the Link Reset protocol or recovering from a Link Timeout.
    - Used in FC to remove an extraneous Class 1 connection
    - May not be useful for Gigabit Ethernet
- Link Reset Response (LRR)
  - The transmitting station is in the LRR Receive state, receiving and recognizing LR.

Reasoning Behind Primitive Sequences & Link Recovery

- Under certain conditions, the use of frames for communication is either unreliable or inappropriate.
  - For example, if the error rate on a link is much higher than expected, frames are likely to fail the CRC test.
    - Primitive Sequences should be used for communication under these circumstances.
- Each Primitive Sequence consists of the continuous repetition of a particular Ordered Set until some event, defined for a particular sequence occurs.
  - For example, a state change by a receiver.
- Events that terminate Sequence transmission include:
  - The receipt of a Primitive Sequence in response;
  - Expiration of a defined time period.

Reasoning Behind Primitive Sequences & Link Recovery

- Primitive Sequences can be used to signal states and transitions that either cannot be adequately signaled using frames or need to be signaled while transmission of a frame is not possible.
  - For example - in the presence of a high link-error rate.
- Link Recovery protocols includes three link procedures which are nested and organized from the most serious to least serious link action.
  - Link Failure procedure
  - Link Initialization procedure
  - Link Reset procedure
- Bit and Word Synchronization and any of the Link Recovery procedures may be executed simultaneously.



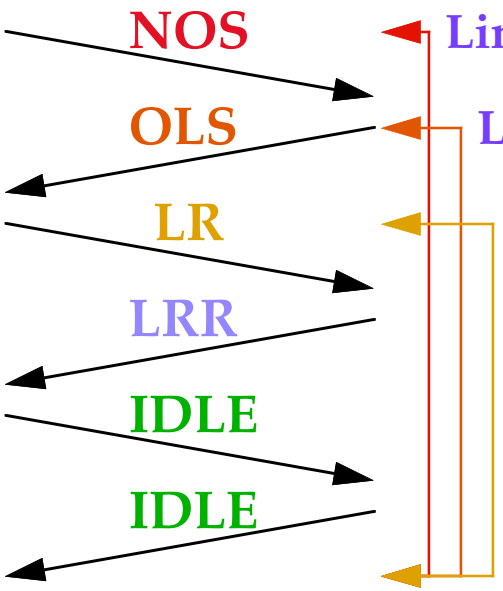
Link Initialization Procedure

- **Link Initialization**
  - Link initialization is required after a station has been powered-on, has been internally reset, or has been Offline.
    - While in the Offline State, NOS Reception or Link Failure conditions which are detected are not recorded as Link Failure events.
      - A station is Online when it is in the Active State.
      - A station performs the Online to Offline Protocol to enter the Offline State.
        a) The station transmits OLS for a minimum time of 5 ms.
        b) After transmitting OLS for 5 ms, the station is Offline and may enter a non-normal mode, turn off its transmitter, power down, or transmit any signal (excluding Primitive Sequences) without errors being detected by the other attached station.
      - To exit the Offline State, a station performs the Link Initialization procedure.

Link Rest and Link Failure Procedures

- **Link Failure**
  - The Link Failure procedure is performed after a station has detected either a Loss of Synchronization for a period of time greater than a timeout period, or Loss of Signal while not in the Offline State.
  - The Link Failure procedure is also performed after a Link Reset procedure timeout error is detected.
- **Link Reset**
  - The Link Reset procedure is performed following a link timeout.
  - In case of link errors, the state of an existing connection (through a switch), if any, may not be known with certainty. When the state of a connection is unknown, the Link Reset procedure shall be used to remove the connection and establish a known state.
    - Note: Link timeouts and Connections may not be pertinent to Gigabit Ethernet.

Link Recovery Hierarchy & Ordered Set Usage



Primitive Sequence	Beginning RD	Ordered Set
OLS	negative	K28.5 D21.1 D10.4 D21.2
NOS	negative	K28.5 D21.2 D31.5 D5.2
LR	negative	K28.5 D9.2 D31.5 D9.2
LRR	negative	K28.5 D21.1 D31.5 D9.2

- The same Link Recovery Primitive Sequences used for Fibre Channel are proposed for use with Gigabit Ethernet.

Idle Primitive Signal

- The Idle Primitive Signal indicates that the station is ready for MAC Frame transmission and reception.
  - Idles are transmitted whenever Frames, other Primitive Signals and Primitive Sequences are not being sent to insure that a receiver remains in synchronization.
  - Idles can be inserted and removed by inter-station units (e.g., repeaters, converters, etc.) to compensate for differences in clock frequencies.
- The same Idle Primitive Signal used for Fibre Channel is proposed for use with Gigabit Ethernet.

Primitive Signal	Beginning RD	Ordered Set
Idle	negative	K28.5 D21.4 D21.5 D21.5

Other Primitive Signals

- The base Fibre Channel standard supports only one other Primitive Signal, R_RDY (FC-AL supports others).
  - The R_RDY Primitive Signal indicates that the transmitting station is Ready to Receive a frame.
  - Alternate flow control schemes have been proposed for Ethernet and the 802.3 HSSG has recommended accommodation of the proposed P802.3x flow control scheme.
    - Further investigation of the P802.3x flow control scheme and its applicability to Gigabit Ethernet is warranted.
      - Current scheme defines extremely short frames.

Frame Delimiters

- A Frame Delimiter is an Ordered Set that immediately precedes or follows the contents of a frame.
  - Separate and distinct delimiters identify the start of a frame and the end of a frame.
  - Frame delimiters are recognized when a single Ordered Set is detected.
- The Ethernet and Fast Ethernet Preamble consists of a series of alternating 1s and 0s to allow network hardware to synchronize with the frame timing.
  - Bit and Word synchronization is provided by out-of-band signaling during the transmission of Primitive Signals and Primitive Sequences surrounding the transmission of MAC layer frames.
    - The Preamble is extraneous and not required for Gigabit Ethernet.

Start of Frame Delimiter (SFD)

- The Ethernet and Fast Ethernet single-byte Start of Frame Delimiter (SFD) may be replaced by an 8B10B-based Ordered Set specified such that it is easily differentiated from other defined Ordered Sets.
  - The Fibre Channel Class 3 SOF delimiter is one candidate delimiter.
    - FC Class 3 service is a connectionless and unacknowledged, not unlike Ethernet
  - The suggested alternative is to define a Start of Frame delimiter separate from all Fibre Channel SOF Delimiters in order to clearly distinguish FC frames from Gigabit Ethernet frames.

Delimiter Function	Beginning RD	Ordered Set
Start of Frame Delimiter (SDF)	negative	K28.5 D21.5 D2.1 D2.1

End of Frame Delimiter (EFD)

- Ethernet and Fast Ethernet signal the termination of a frame of data by the inclusion of an IDL condition is the bit stream.
  - The size of each frame and the location of each field is implicit.
  - All fields are fixed except for the Data field whose size is specified by the Type field, located at a fixed offset from the SDF.
  - Fibre Channel uses EOF Delimiters for the following purposes:
    - Terminate a frame;
    - Terminate a frame and a Fibre Channel Sequence;
    - Implicitly define the size of the Data field of a frame;
    - Remove Class 1 connections;
    - Indicate that the frame content is invalid.
  - None of the preceding uses are required for Gigabit Ethernet.
  - Idles may be used to reliably terminate a Gigabit Ethernet frame.
  - An EFD may be specified to satisfy ancillary requirements.

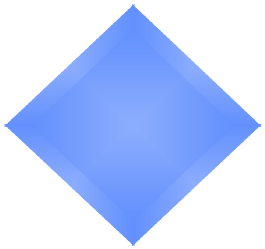
Ancillary Requirements for an End of Frame Delimiter (EFD)

- For Gigabit Ethernet an EFD may be used to satisfy the following ancillary requirements:
  - Terminate a frame and implicitly define the size of the Data field of a frame;
    - The MAC Type field could be eliminated. However, this is undesirable as the 'Use 802.3/Ethernet frame format' Gigabit Ethernet Project Objective is not met and the Type field must be reconstructed if the frame is forwarded to other Ethernet variants.
  - Indicate that the frame content is invalid.
- If specified, both positive and negative Beginning Running Disparity versions of all EFDs must be specified since the FCS field immediately preceding the EFD may end with either positive or negative RD.

End of Frame Delimiter (EFD) Ordered Sets

- If Specified, Different EFDs than those used for Fibre Channel are proposed for use with Gigabit Ethernet in order to clearly distinguish frames.
- If Idles are used to terminate a Gigabit Ethernet frame, a positive Beginning RD Idle must be defined.

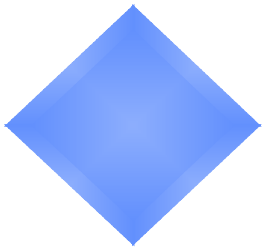
Delimiter or Primitive Signal	Beginning RD	Ordered Set
EFD Normal (EFDN)	negative	K28.5 D21.5 D5.0 D5.0
EFD Normal (EFDN)	positive	K28.5 D21.4 D5.0 D5.0
EFD Abort (EFDA)	negative	K28.5 D21.5 D5.4 D5.4
EFD Abort (EFDA)	positive	K28.5 D21.4 D5.4 D5.4
Idle	positive	K28.5 D21.5 D21.5 D21.5



JAM

- Ethernet and Fast Ethernet JAM signal is 4 octets of b'10101010' transmitted by a station when a collision is detected.
  - JAM serves as a Collision End of Frame Delimiter.
  - The JAM signal may be specified as a Primitive Signal including both positive and negative Beginning Running Disparity versions since the data immediately preceding the JAM may end with either positive or negative RD.

Primitive Signal	Beginning RD	Ordered Set
JAM	negative	K28.5 D21.5 D9.0 D9.0
JAM	positive	K28.5 D21.4 D9.0 D9.0



PAD

- A single PAD character is required to achieve even byte count for frames with odd-byte length.
  - Proposal: Accept Andreas Bechtolsheim (Granite Systems) which Recommends using K23.7 for PAD character.
  - Invalid Transmission Word rules must be altered to allow PAD character.

Extension (FILL)

- Required to extend minimum frame size for Gigabit Ethernet in order to achieve useful topologies.
  - FILL Primitive Signal is specified;
  - Collision window is extended to include FILL;
  - FCS is not altered by FILL;
  - FILL is not passed to LLC;
  - FILL follows FCS;
    - If IDLE is used to normally terminate frames, FILL precedes IDLE.
    - If EFD is used to normally terminate frames, FILL follows EFD.

Primitive Signal	Beginning RD	Ordered Set
FILL	negative	K28.5 D21.5 D9.4 D9.4
FILL (req'd only if EFD not used)	positive	K28.5 D21.4 D9.4 D9.4

Mutual Benefits

- Supporting the proposed four character Ordered Sets in Gigabit Ethernet provides the following benefits:
  - MAC frames are transported intact and may easily be forwarded to Fast Ethernet and Ethernet stations.
  - Gigabit Ethernet and Fibre Channel can use the same links and link hardware.
    - A Host "mode" change allows a single host port to support both protocols.
    - A Receiving station may easily distinguish between Gigabit Ethernet and Fibre Channel frames.
    - Auto Negotiation, OFC, Link Recovery protocols increase link reliability and may be common for both protocols.
    - Common components to the MAC/FC-2 layers reduce costs for both protocols.