"Fast Ethernet" Gets Plug-and-Play

Ian Crayford
I/O & Network Products Division
Advanced Micro Devices,
One AMD Place,
Sunnyvale, California. 94088-3453

ABSTRACT

The IEEE 802.3 "Fast Ethernet" standard implements an "Auto-Negotiation" mechanism which allows any existing or future Ethernet LAN using an RJ-45 connector and unshielded twisted pair cable to simply "plug and play".

The mechanism is designed to ensure interoperability between the variety of standards based Ethernet networks, including exiting 10 Mb/s devices, and new 100 Mb/s half duplex and full duplex capable devices. Automatic configuration occurs without user intervention, and will negotiate the optimal mode at which the devices at both ends of a link can commonly operate. In the event the devices share no common mode, the Auto-Negotiation scheme ensures that operation of the rest of the network is not disrupted, and allows the absence of a shared operational mode to be reported via network management.

This paper will examine the key aspects of the technology surrounding the Auto-Negotiation scheme, as well as specific examples of the use of the Auto-Negotiation approach and how it solves potential interoperability problems.

INTRODUCTION

Why Fast Ethernet?

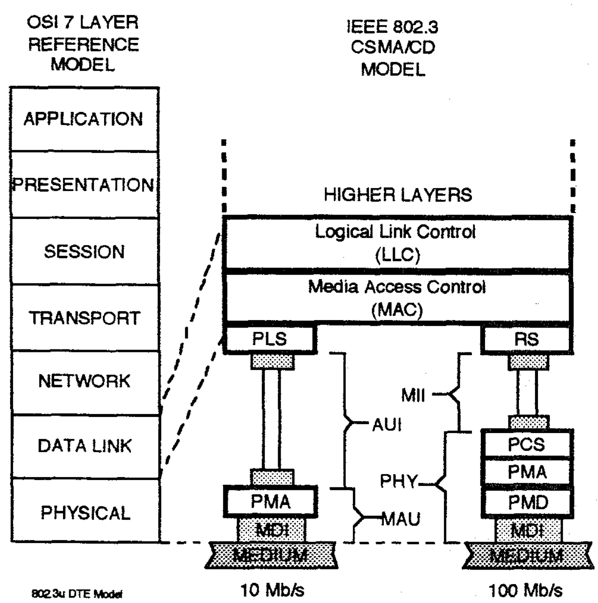
The Fast Ethernet specifications were developed within the IEEE 802.3u 100BASE-T Working Group by a large number of individuals from many companies. Numerous factors led to the development of the series of specifications which make up the 100BASE-T document, which became an official IEEE Standard in June, 1995[1]. These factors included the recognition of both technology advancements and market developments. Key factors which led to the development of 100BASE-T included:

- 10BASE-T[2] had become the overwhelmingly dominant LAN technology deployed worldwide.
- Rapid migration to 10BASE-T was the result of it's simple, low cost, star wired topology. 10BASE-T networks were far easier to manage and modify than the coaxial bus topologies of the earlier Ethernet (802.3, 10BASE5[3]) and Cheapernet (802.3, 10BASE2[4]) versions.
- Network growth, in terms of number of users and network reliant applications, had caused saturation of some 10 Mb/s Ethernet networks, and more were likely to follow.
- Emerging applications, with richness of data types and reliance on connectivity were likely to drive additional bandwidth requirements.
- The use of segmentation and switching to reduce the number of stations sharing an Ethernet and relieve bandwidth bottlenecks was obtaining growing acceptance.
- Improvements in silicon and signal processing technologies allowed higher speed options to be considered.

Fast Ethernet Overview

100BASE-T retains the same Ethernet CSMA/CD Media Access Control (MAC) protocol, effectively scaling the data rate from the original 10 Mb/s to 100 Mb/s. The topology is a repeater based star, with two repeater types supported (defined as Class I and Class II repeaters).

Figure 1 shows the layer model for a 10 or 100 Mb/s end-station (Data Terminal Equipment, or DTE).



AUI = Attachment Unit Interface (Option in 10 Mb/s, not in 100 Mb/s)
MAU = Medium Attachment Unit
MDI = Medium Dependent Interface
MII = Media Independent Interface (Option in 10 Mb/s and 100 Mb/s)
PCS = Physical Coding Sublayer
PHY = Physical Layer Device
PLS = Physical Layer Signaling
PMA = Physical Medium Attachment
PMD = Physical Medium Dependent (100BASE-X only, not 100BASE-T4)
RS = Reconciliation Sublayer

Figure 1 - 10/100 Mb/s DTE Layer Model

100BASE-T defines a new sublayer under the MAC, called the Reconciliation Sublayer (RS). The RS essentially maps the behavior of the MAC to the electrical signals of the Media Independent Interface (MII). The MII is a logical equivalent to the Attachment Unit Interface (AUI) of the original 10 Mb/s Ethernet. The AUI allows different media types to be supported under the original Ethernet MAC and Manchester encoder/decoder (the Physical Layer Signaling, or PLS sublayer). Unlike the AUI, which required a 6 wire (3 pair) serial interface between the DTE and the

ISBN# 0-7803-2636-9

MAU, the MII is an 18 wire signal interface. Data is moved across the MII in the form of nibbles (4 bits) for each clock cycle, so although the network data rate is 100 Mb/s, the MII operates at 25 MHz. The MII consists of the transmit and receive data nibbles, clocks, data valid and error signals. In addition, carrier (network) activity, collision detect and management interface signals are provided.

100BASE-T supports 2 Unshielded Twisted Pair (UTP) versions, defined as 100BASE-T4 and 100BASE-TX. Both operate over 100m lengths of UTP cable. A 100BASE-FX fiber option is permitted for longer distance communications.

Why Auto-Negotiation?

As 100BASE-T evolved, it was clear that two main wiring types would need to be supported, Category 3 (Cat 3) and Category 5 (Cat 5) UTP. The "category" rating refers to the electrical performance of the cable. Cat 5 is better quality cable than Cat 3, having less high frequency attenuation, lower susceptibility to external electromagnetic interference (EMI) and radio frequency interference (RFI), and reduced self emission of EMI/RFI.

Cable quality, and the number of pairs required are major issues. 10BASE-T requires only 2 pairs of Cat 3 cable to operate. To achieve 100 Mb/s data rate transmissions over UTP, and meet data integrity and regulatory emissions requirements, is highly challenging. This generally means that either better quality cable is required, or more pairs of lower quality cable.

In addition, cable quality and pair availability tend to vary geographically. For instance, in Europe, very little 4 pair cabling exists, and 2 pair UTP or Shielded Twisted Pair (STP) is common. In the U.S., 4 pair Cat 3 cabling exists extensively in the installed base, but new installations are almost exclusively 2 or 4 pair Cat 5.

With these issues in mind, two solutions for UTP evolved. 100BASE-T4 operates over 4 pairs of Cat 3 (and better) cable, while 100BASE-TX operates over 2 pairs of Cat 5 cable.

With the considerable sharing in functionality between the Ethernet MAC, RS and MII, which are designed to support both 10 Mb/s and 100 Mb/s operation, it was an obvious choice for vendors to want to support both 10 and 100 Mb/s operation with one device, especially adapter and switch applications.

A key attribute that had made 10BASE-T successful was its plug-and-play ability. A concern in developing 100BASE-T was that this attribute may be compromised, especially if the signaling solutions for different cables were incompatible with each other and 10BASE-T. It was felt that if 10 Mb/s and 100 Mb/s Ethernet components were interconnected, the effect should be minimal. This would be especially important for instance, when installing a new 100 Mb/s adapter. If it was inadvertently connected to an existing 10BASE-T network, it should not cause disruption of the entire existing 10BASE-T installation..

So, key concerns drove the need for Auto-Negotiation:

- Manufactures wanted to make "speed agile" devices, which would operate at 10 Mb/s now, and would automatically upgrade to 100 Mb/s operation as the network infrastructure was upgraded.
- The 100 Mb/s signaling solutions were incompatible with each other, and 10BASE-T, and threatened to undermine the plug-and-play reputation of Ethernet.
- Additional solutions which may be developed in the future, would have to take account of all other solutions that had been previously deployed, and operate with a signaling scheme which did not disrupt the installed base.

Overview of Specification

Auto-Negotiation is located in Clause 28 of the 100BASE-T Standard suite (a "Clause" is IEEE terminology for a section or chapter).

The Auto-Negotiation clause defines several key mandatory and optional features that a device needs to posses to be considered compliant with the specification.

The mandatory features are basically:

- The device must transmit a special pulse train to identify itself after power up (this pulse train is referred to as a "Fast Link Pulse Burst").
- The device must recognize a similar pulse train from the "link partner" device at the opposite end of the link.
- If the local device recognizes the link partner has compatible operating modes, the optimal mode is selected, and both devices move to their operational (link pass) state.

Optional features include:

- A device can send additional information using the special pulse train if both devices provide this feature (referred to as "Next Pages").
- A device may indicate to its link partner using the special pulse train that it has a fault, and may further identify the type of fault.

AUTO-NEGOTIATION

Building on the 10BASE-T Legacy

The Auto-Negotiation (AutoNeg) algorithm builds on a key characteristic that all 10BASE-T devices provide, namely the 'link test" or "link integrity" pulse.

When there is no packet data being transmitted by a 10BASE-T device, it is responsible to periodically transmit a single, short duration (approximately 100ns) pulse as a "heartbeat" indication. If the receiver of a 10BASE-T device does not see either packet data, or a link test pulse within a defined time window (16ms ±8ms) it will enter a "link fail" condition, which disables data transmission and reception, and can be locally detected by the device as a medium failure. So if the receive pair is disconnected, the 10BASE-T device will enter the link fail state, and further transmission is disabled. During link fail, the transmission and reception of link test pulses continues. To re-establish the link, at least 2 consecutive link test pulses or a single receive packet must be received.

10BASE-T incorporates the link test pulse function in order to ensure end-to-end connectivity between the two devices at either end of the twisted pair link (typically an end-station and a repeater port). 10BASE-T uses separate transmit and receive conductor pairs in the UTP cable, unlike the original coax based Ethernet and Cheapernet systems where transmit and receive were connected to the same single inner conductor of the coax. A mechanism is necessary to detect if a single pair is disconnected while the other pair remains intact. Connectivity of both pairs is fundamental to the correct operation of the CSMA/CD MAC protocol, and to ensure the Ethernet MAC observes no difference between the underlying media types (whether coax or UTP).

The link test pulse capability permitted two additional side benefits. Firstly, the detected "link status" was displayed via an external LED on each 10BASE-T device, a feature universally welcomed by installers, network administrators and users alike, as a quick and simple indication of basic connectivity. Secondly, since the link test pulse is unipolar (a positive excursion), unlike the normal differential data transmission (both positive and negative excursions), many implementations used this as a simple and fast indication of the polarity of the received signal, and provided automatic correction in the 10BASE-T device to rectify simple wiring errors (although robust auto-polarity detect/correct algorithms use additional criteria other than the polarity of the link test pulse alone).

Given conservative estimates of some 35 million 10BASE-T nodes (and corresponding repeater ports) in the installed base (the number is difficult to estimate since the current deployment rate is in the order of 25 million adapters and repeater ports annually), the incentive to create an interoperable solution with 10BASE-T was undeniable.

Auto-Negotiation Signaling Scheme

Auto-Negotiation makes use of the timing relationship that 10BASE-T specifies for link test pulses, by replacing the single link test pulse with a burst of pulses.

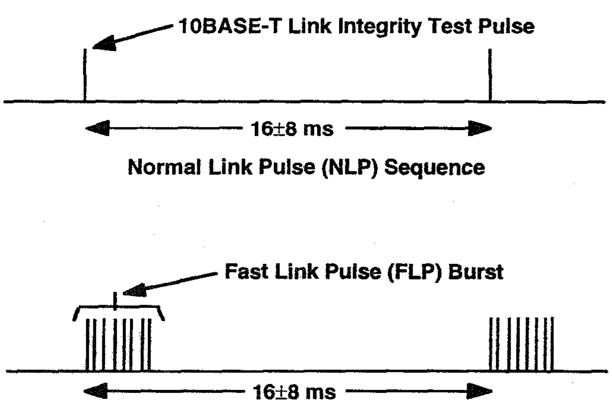


Figure 2 - NLP to FLP Comparison

The link test pulse of 10BASE-T is defined as a Normal Link Pulse (NLP) and the burst of pulses that Auto-Negotiation devices issues is defined as the Fast Link Pulse (FLP) Burst. Figure 2 above shows the relationship of 10BASE-T NLPs to the Auto-Negotiation FLP Burst.

Each FLP Burst consists of a series of clock and data pulses, as shown below in Figure 3. The entire FLP Burst is completed within 2ms. This ensures that existing 10BASE-T devices observe the FLP Burst as a single link test pulse (due to the timing specifications of 10BASE-T). As can be seen from Figure 3, odd numbered pulse positions in the FLP Burst are designated as clock pulses, and are separated by 125µs (nominally), while even numbered pulse positions are designated as data pulses, and occur (if present) 62.5µs from the previous clock pulse position.

The data pulses within the FLP Burst are used to form a 16 bit Link Code Word (LCW). The clock pulses are used only for timing and recovery of the data pulses.

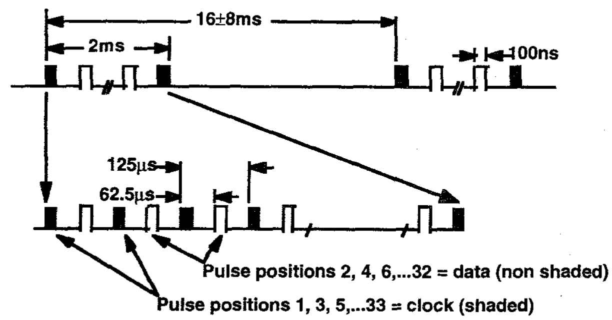


Figure 3 - FLP Burst Timing

Figure 4 shows the construction of an FLP Burst in more detail. The entire FLP Burst consists of a minimum of 17 pulses (assuming all data pulses are absent), and a maximum of 33 pulses (assuming all data pulses are present). While clock pulses are always present, data pulses may or may not be present. In the case that a data pulse is present, this indicates a value of 1 in the LCW, whereas absence of a data pulse indicates a value of 0 in the LCW.

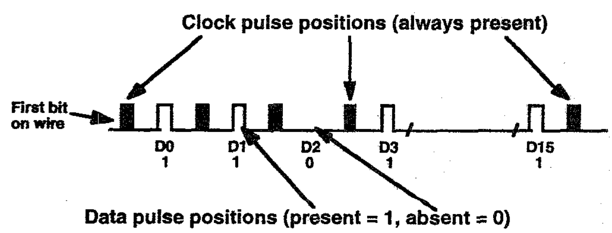


Figure 4 - FLP Burst Content

Received signals are evaluated to ensure that an FLP Burst is correctly identified and the embedded Link Code Word is decoded: Timers are used to:

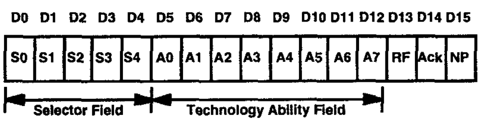
(i) Detect the presence of an FLP Burst (versus an NLP).
(ii) Detect the presence or absence of a data pulse (between adjacent clock pulses).

(iii) Detect the FLP Burst repetition rate (to reject a spurious noise event as a potential FLP Burst).

Additional robustness is provided by the fact that 3 identical Link Code Words must be received before the contents are considered reliable, and used for the negotiation process.

Base Link Code Word

Figure 5 below defines the bit positions of the initial Link Code Word that is exchanged between Auto-Negotiation devices. This is defined as the Base Link Code Word, and is transmitted after power-on, reset, or re-negotiation is requested by one of several mechanisms.



Selector Field
S[4:0] = 00001 = 802.3
S[4:0] = 00010 = 802.9

Technology Ability Field	
A0	10BASE-T
A1	10BASE-T FD
A2	100BASE-TX
A3	100BASE-TX FD
A4	100BASE-T4
A5	Reserved
A6	Reserved
A7	Reserved

Other Fields	
RF	Remote Fault
Ack	Acknowledge
NP	Next Page

Figure 5 - Base Link Code Word Definition

Selector Field - This 5 bit value indicates the device type. Even though Auto-Negotiation was specifically developed for 802.3 devices, extensibility to other technologies was incorporated. Other technologies need only apply for a unique Selector Field value. When another technology uses the Auto-Negotiation function, the contents of certain fields in the LCW are permitted to be re-defined to contain data appropriate to that technology.

Technology Ability Field - Identifies the subset of technologies that are supported (i.e., 10BASE-T, 100BASE-TX/T4, Full Duplex, etc.). All technologies are simultaneously advertised in the Base Link Code Word. Currently, 3 additional bits are available for expansion, with additional expansion provided by the Next Page function (explained later). Each device maintains a prioritization table, used to ensure the highest common denominator ability is chosen. The priority is assigned by technology type and administered by 802.3, and is not based on the bit ordering in the Link Code Word. The priority table is currently defined as (from highest to lowest):

1. 100BASE-TX (Full Duplex) (Highest)
2. 100BASE-T4
3. 100BASE-TX
4. 10BASE-T (Full Duplex)
5. 10BASE-T (Lowest)

New technologies can simply be inserted at any point in the priority table. Existing devices which have no knowledge of these new technologies will ignore the bits in the Link Code Word, and so will not have their priority hierarchy altered.

Remote Fault Bit - Indicates the presence of a fault detected by the remote link partner. In addition, the Next Page Bit may be set to identify the precise type of remote fault. One of the simplest uses of Remote Fault is to set this bit if the device enters the "link fail" state, which may the result of a broken receiver (or receive cable). Using this mechanism for instance, a hub can be informed by a DTE that no receive path appears to exist from the hub to the DTE, although the DTE is able to report this using the intact transmit path to the hub.

Acknowledge Bit - Acknowledges the successful receipt of 3 identical LCWs from the link partner. An LCW with ACK=1 is transmitted a minimum of 6-8 times to the link partner. This ensures the partner will detect the LCW with ACK=1 also 3 consecutive times. The ACK bit is one field that is fixed across all Selector Field values.

Next Page Bit - Indicates a device wishes to send additional Link Code Word(s) following the current word being exchanged. In order for additional pages to be sent, both ends must be "Next Page Able" (both must set this bit in the Base LCW). Next Page information will follow the same protocol and timing as the Base LCW. Encodings for Next Pages have been defined to allow expansion of the Technology Ability Field, to more precisely define the nature of a Remote Fault, and to allow vendor specific extensions.

Next Page Function

The Next Page function allows the exchange of LCWs in addition to the Base LCW. However, in order for any Next Pages to be exchanged, both devices must indicate they are "Next Page Able" in the Base LCW. If both devices indicate they are Next Page Able, both must send at least one Next Page. Next Page exchange occurs until both devices have no additional Next Pages. If one device has no more information, it continues to send a "Null Message Page" until the partner device has completed its Next Page transmissions. The Next Page protocol consists of a two message sequence. A "Message Page" is transmitted first (see Figure 6 below), which indicates both the number and type of "Unformatted Pages" (if any) that will follow (in the Message Code Field). Unformatted Pages (see Figure 7 below) are subsequently transmitted as necessary, until the transfer completes. Unformatted Pages effectively provide the 11 bit Unformatted Code Field of the Next Page to be used for data, while the top 5 bits are protocol overhead.

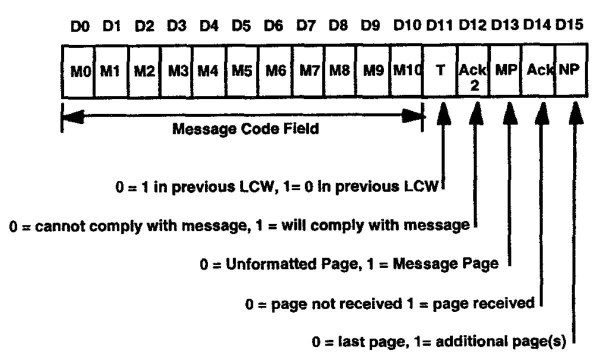


Figure 6 - Encoding of Next Page Message Page

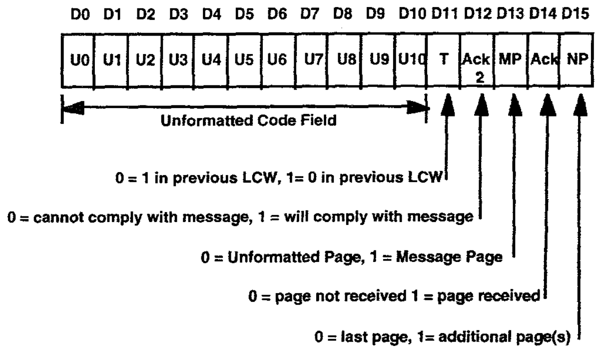


Figure 7 - Encoding of Next Page Unformatted Page

Management Interface
Auto-Negotiation allows for management via the MII Management Interface (or an equivalent if the MII is not provided). The mandatory Control (Reg. 0) and Status (Reg. 1) features provide most of the enable/disable and ability reporting for Auto-Negotiation. Three additional Auto-Negotiation specific registers are defined as mandatory if Auto-Negotiation is implemented:
Advertisement Register - Used to store the devices LCW for transmission.
Link Partner Ability Register - Stores the advertised ability of the link partner device (the received LCW).
Expansion Register - Miscellaneous AutoNeg specific information (i.e., Link Partner AutoNeg Able, Page Received, etc.).
Next Page Transmit Register - Required only if the Next Page function is implemented. Used to store the Next Page of information to be transmitted.

Functional Overview
Figure 8 below shows a simple block diagram of an 802.3 device with multiple technologies supported, including 10BASE-T. The signals passed between the MDI (the RJ-45) and the AutoNeg function are the transmit and receive data signals from the UTP medium, which always use the same pins of the RJ-45 for all 802.3 networks.

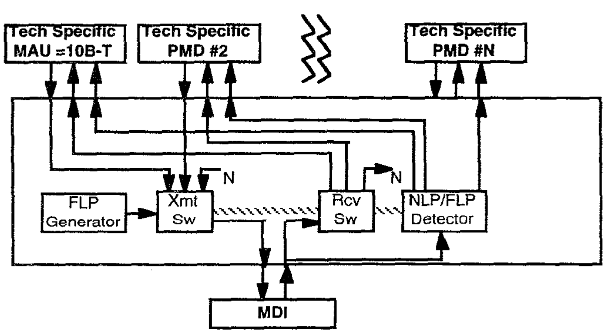


Figure 8 - AutoNeg Functional Block Diagram

Each of the technologies reside above the AutoNeg layer, which polices what is transmitted from them to the medium, and also directs any receive signals from the medium. Each technology must encompasses its' own native detection scheme for "link good" (e.g., 10BASE-T uses normal link pulses), to provide a mechanism to verify the underlying medium once control is handed off from AutoNeg, and a single technology is enabled.

After power up. reset, or upon a re-negotiation request, the Xmt Sw (Transmit Switch) connects the FLP Generator to the MDI transmit path and FLP Bursts are generated to advertise the device's abilities to the link partner. All other technology transmit functions are isolated from the MDI during negotiation. After negotiation, the Xmt Sw connects a single technology transmitter to the MDI transmit, and the technology is then responsible to use its own test for link good.

The NLP/FLP Detector detects NLPs (which indicates the other device is 10BASE-T only), or FLP Bursts in which case the Base LCW is decoded to determine the optimal mode shared by both devices. Once a single technology has been chosen by the AutoNeg function, it is informed and control is passed to it, to perform its' own link integrity function.

During the negotiation phase, the Rcv Sw (Receive Switch) blocks MDI receive traffic to the technology receivers, passing it to the NLP/FLP Detector only (the Parallel Detection function allows a modification to this, which is explained below). After negotiation completes, the Rcv Sw connects the MDI receive to a single technology receiver.

Parallel Detection Function
Auto-Negotiation assumes that devices generate NLP and/or FLP Sequences. However, since the AutoNeg work started later than the definitions of 100BASE-TX and T4, early implementations did not provide FLP compatible signaling. The "Parallel Detection" scheme was developed to permit interoperation with 100BASE-T technologies under development prior to initiation of AutoNeg. Only 100BASE-TX/T4 technologies are permitted to use Parallel Detection, all future 802.3 technologies must use the FLP Burst transaction.

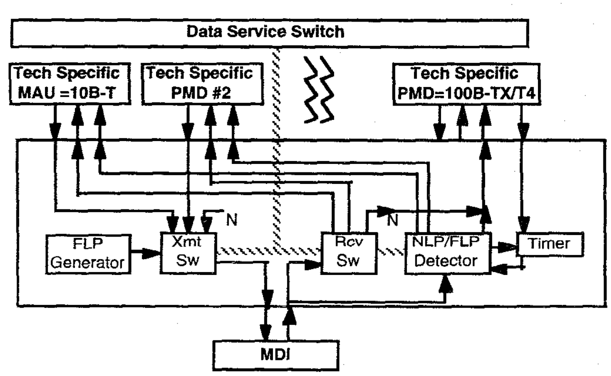


Figure 9 - AutoNeg With Parallel Detection Function

Figure 9 shows how Parallel Detection is incorporated into the AutoNeg function. A Timer is added, and the RCV Sw and NLP/FLP Detector are modified. Now, as signals are received from the medium, they are passed in parallel to the NLP/FLP Detector, and to any 100BASE-TX or T4 PMD which exists in the implementation. If the received signal causes the TX/T4 PMD to enter the link good condition (it is a native TX or T4 signal, rather than an FLP), the link good event is detected by the AutoNeg function and the Timer started. When the timer expires, if only one technology indicates link good, control is passed to it. If more than one link good condition exists after the time expires, then the signal at least appeared acceptable to more than one technology. AutoNeg will flag this as an error, and the device remain in link fail, isolated form the network.

INTEROPERABILITY EXAMPLES

Interoperability With 10BASE-T Installed Base

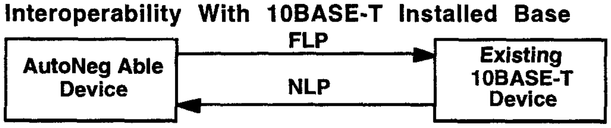


Figure 10 - AutoNeg to Existing 10BASE-T

- The AutoNeg able device powers up in link fail and transmits consecutive FLPs (after "break link" time delay).
- The 10BASE-T device transmits NLPs or traffic (10BASE-T devices are allowed to power up in link good).
- The 10BASE-T device will go into link fail due to inactivity for the break link time.
- If the AutoNeg device has a 10BASE-T mode, AutoNeg passes control to it, the 10BASE-T MAU transmits NLPs and the 10BASE-T mode is configured

Protection From Incompatible Equipment

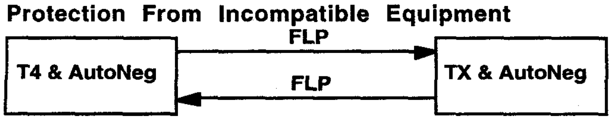


Figure 11 - 100BASE-T4 AutoNeg to 100BASE-TX AutoNeg

- Both devices power up in link fail and transmit FLPs.
- Upon receiving 3 consecutive and consistent FLP Bursts, the capabilities of the far end station are recognized.
- Based on the capabilities communicated, no common denominator exists (assumes no common 10BASE-T data service).
- Both devices remain in link fail - any network connected to either end will not be disrupted.
- The link partner's ability information is saved in local register(s) and could be used for management purposes (report incompatible device type).

Interoperability With Non AutoNeg 100BASE-T

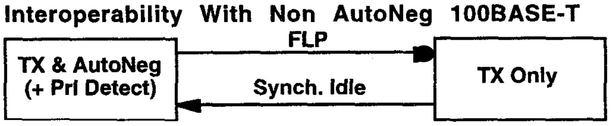


Figure 12 - 100BASE-TX AutoNeg to 100BASE-TX Only

- AutoNeg able device powers up in link fail state and transmits FLPs.
- 100BASE-TX/Non-AutoNeg device powers up and sends "Idle Symbols" (continuous activity during idle).
- Providing 100BASE-TX network technologies exist at both ends, 'Parallel Detection" will allow NLP/FLP detector to be bypassed.
- Devices will operate in 100BASE-TX mode.
- AutoNeg could also be disabled and the mode forced if by management or manually (least optimal choice) if required.

SUMMARY

Interoperability

Auto-Negotiation solves the growing potential for interoperability conflicts between the broadening array of 802.3 compatible products. It is interoperable with the huge installed base of 10BASE-T devices. It enables new "agile" technology devices to plug-and-play with existing equipment now, and to automatically upgrade as the network infrastructure is enhanced in the future. The algorithm provides a path for early adopter, non-agile 100 Mb/s implementations using the Parallel Detection function. Adequate code space is provided for future expansion, while preserving future interoperability.

Performance

Auto-Negotiation provides built-in redundancy for signaling robustness in the UTP noise environment. The algorithm completes in a bounded time, negotiating to a Highest Common Denominator mode where one exists, or indicating failure through management when there is no common operating mode. Auto-Negotiation does not test all aspects of the UTP cable performance, leaving this to the appropriate technology interface where these are critical. The complexity of Auto-Negotiation is modest in current VLSI process technology.

Benefits

The primary benefit that Auto-Negotiation provides is that it protects operating 802.3 networks from "alien" equipment. There are clearly many other features of merit, such as "remote fault" detection, future expandability, and provision for vendor enhancements. However, the real benefit will be if Auto-Negotiation can help make 100BASE-T as pervasive in the future as 10BASE-T already is.

---

[1] Draft Supplement to 1993 version of ANSI/IEEE Std 802.3, Document # P802.3u/D5, MAC Parameters, Physical Layer, Medium Attachment Units and Repeater for 100 Mb/s Operation (Version 5.3, dated March 23, 1995), IEEE Standards Department.

[2] ISO/IEC 8802-3 :1993 (E) ANSI/IEEE Std 802.3-1993 Edition, Section 14.

[3] ISO/IEC 8802-3 :1993 (E) ANSI/IEEE Std 802.3-1993 Edition, Section 8.

[4] ISO/IEC 8802-3 :1993 (E) ANSI/IEEE Std 802.3-1993 Edition, Section 10.