TEXAS INSTRUMENTS

ThunderLAN™ Adaptive Performance Optimization™

Technical Brief

1995

Networking Products Group

Texas Instruments

Printed in U.S.A., July 1995

SPWT089

ThunderLAN™ Adaptive Performance Optimization™

Technical Brief

SPWT089
July 1995

IMPORTANT NOTICE

Texas Instruments (TI) reserves the right to make changes to its products or to discontinue any semiconductor product or service without notice, and advises its customers to obtain the latest version of relevant information to verify, before placing orders, that the information being relied on is current.

TI warrants performance of its semiconductor products and related software to the specifications applicable at the time of sale in accordance with TI's standard warranty. Testing and other quality control techniques are utilized to the extent TI deems necessary to support this warranty. Specific testing of all parameters of each device is not necessarily performed, except those mandated by government requirements.

Certain applications using semiconductor products may involve potential risks of death, personal injury, or severe property or environmental damage ("Critical Applications").

TI SEMICONDUCTOR PRODUCTS ARE NOT DESIGNED, INTENDED, AUTHORIZED, OR WARRANTED TO BE SUITABLE FOR USE IN LIFE-SUPPORT APPLICATIONS, DEVICES OR SYSTEMS OR OTHER CRITICAL APPLICATIONS.

Inclusion of TI products in such applications is understood to be fully at the risk of the customer. Use of TI products in such applications requires the written approval of an appropriate TI officer. Questions concerning potential risk applications should be directed to TI through a local SC sales office.

In order to minimize risks associated with the customer's applications, adequate design and operating safeguards should be provided by the customer to minimize inherent or procedural hazards.

TI assumes no liability for applications assistance, customer product design, software performance, or infringement of patents or services described herein. Nor does TI warrant or represent that any license, either express or implied, is granted under any patent right, copyright, mask work right, or other intellectual property right of TI covering or relating to any combination, machine, or process in which such semiconductor products or services might be or are used.

TRADEMARKS

Adaptive Performance Optimization and ThunderLAN are trademarks of Texas Instruments Incorporated.

Ethernet is a registered trademark of Xerox Corporation.

Contents

List of Tables

List of Illustrations

Introduction

The Texas Instruments (TI) ThunderLAN™ TNETE100 is a single chip Ethernet™ controller for the Peripheral Component Interconnect (PCI) local bus that includes an integrated 10Base-T and 10Base-5 AUI physical layer interface (PHY). When equipped with the appropriate external modular PHY, ThunderLAN can handle 100-Mbps Ethernet protocols to accommodate growth in the user's networking bandwidth demand. TI suggests thorough testing and measurement of system parameters using the TNETE100 device to provide high quality operation in a network meeting IEEE 802.3 (10Base-5 and 10Base-T) approved standards and 802.3u (100Base-T) and 802.12 (100VG) draft standards.

Table 1. Customer Assistance Contacts

TYPE OF INFORMATION	ASSOCIATED CONTACT
Asking about product operation or reporting suspected problems	TNETE Technical Support Line on the Internet: **TLANHOT@micro.ti.com** or send a FAX to: **(713) 274–4027**
Requesting more information about Texas Instruments Networking Products	Texas Instruments Incorporated Market Communications Manager, MS 736 P.O. Box 1443 Houston, Texas 77251–1443
Ordering Texas Instruments documentation	TI Literature Response Center: **(800) 477–8924**
Reporting mistakes in this document or any other TI documentation	Texas Instruments Incorporated Technical Publications Manager, MS 702 P.O. Box 1443 Houston, Texas 77251–1443 Electronic mail: comments@books.sc.ti.com

Adaptive Performance Optimization

Adaptive Performance Optimization™ (APO) refers to specific buffering and pacing techniques geared toward improving adapter performance by adjusting the resources and the transmit procedures to achieve an optimal transmission rate and minimal CPU utilization for varying system and network conditions. APO, as implemented in TI's ThunderLAN architecture, uses automatically adjustable pretransmit buffering to compensate for system bus latency, transmit pacing to reduce server backlog while reducing server / client collision rates, and interrupt pacing / thresholding to reduce adapter-to-host interrupt activity and minimize host CPU context switching.

Transmit Latency Limits Transmit Performance

One way the network interface controller adapts to the system in which it is installed is by monitoring the system bus latency and adjusting the amount of pretransmit buffering to obtain the highest transmit performance level for a particular system. This process of adapting to compensate for the system bus latency makes pretransmit buffering a key component of the APO feature. In order to understand how pretransmit buffer sizing increases transmission throughput, it is important to understand transmit latency and how it affects transmission performance.

Transmit latency, encompassing system bus latency, is defined as the time delay between the point when the host computer initiates frame transmission and the point when the first bits of data appear on the

network. Although network access time is a contributing factor, generally most of the delay is due to the setup and buffering requirements of the LAN adapter.

With perfect, unhindered access to host memory, a LAN adapter requires no transmit data buffering. However, a LAN adapter has to compete with other peripherals and the host CPU for access to memory, so it needs to buffer data to cope with the bursty nature of these accesses. Generally, this is done in a first in/first out (FIFO) buffer. The critical factor in adapter transmit latency is the amount of data buffered before initiating transmission. If insufficient data is buffered, then a transmission can fail because of an inability to fetch more data before sending the buffered data. Insufficient buffered data is referred to as an underrun. If too much data is buffered, a large transmit latency results. Since transmit performance is degraded when the buffer is either too big or too small, it is important to adjust the buffer size to achieve the optimum transmit throughput.

Pretransmit Buffer Sizing

Pretransmit buffer sizing refers to a method of automatically adjusting the amount of pretransmit buffering, based on system bus latency, to obtain the optimum performance level for a particular system. Instead of using a pretransmit buffer of fixed size, the size of this buffer is allowed to change in order to compensate for system bus latency and provides an overall reduction in transmit latency. Over time, the adapter learns the latency of the PCI bus and adjusts the size of its FIFO buffer to account for this latency.

Pretransmit buffer sizing is implemented in hardware using a counter that maintains the number of transmit data underruns and uses this counter to determine the level of pretransmit buffering to use. This counter is called the commit level counter as it controls the amount of transmit data required before the transmitter commits to begin network transmission. Initially the commit level is zero, corresponding to the minimum pretransmit buffering. If the host system is capable of operating with this amount of pretransmit buffering, no transmit underruns occur and the zero commit level remains in use. Should a transmit underrun occur, the commit level is incremented, increasing the amount of pretransmit buffering used. If this level provides insufficient buffering, another underrun occurs, and the level is stepped up, again.

An exponential mapping between the commit level and the number of pretransmit buffering data bytes provides a rapid adaptation to the host system's capabilities. Doubling the number of data bytes buffered before data transmission for every commit level means that the adapter quickly adapts to the system capabilities, while only failing to send a few frames. The adapter can be trained to the system's capabilities by sending a number of frames before beginning normal operation. Table 2 shows the relation of commit level to buffer size.

Table 2. Commit Level Code Corresponding To Transmit Buffer Size

COMMIT LEVEL	BUFFER SIZE
0	64 bytes
1	128 bytes
2	256 bytes
3	512 bytes
4	1024 bytes
5 – 7	Whole frame

Figure 1 shows an example of pretransmit buffer sizing. When reset occurs, ThunderLAN assumes a commit level of zero (64 bytes). As the PCI controller begins filling the FIFO buffer with data during the

adapter's first transmission, the data level in the FIFO begins to rise. When the data level reaches the commit level, the PHY begins transmitting the data onto the network. If the latency of the PCI bus causes the PCI controller to fill the FIFO slower than the transmitter empties it, the data level begins to fall. If the transmitter exhausts all of the data in the FIFO before the PCI controller has delivered the entire frame to the FIFO, the transmission is aborted (due to the underrun condition). An underrun occurs when there is not enough data in the transmit buffer to support a continuous transmit rate on the network. In the event of a transmit underrun, ThunderLAN increments the commit level and retransmits the frame. The process is repeated until the pretransmit buffer commit level adequately satisfies a continuous transmit rate without underruns.

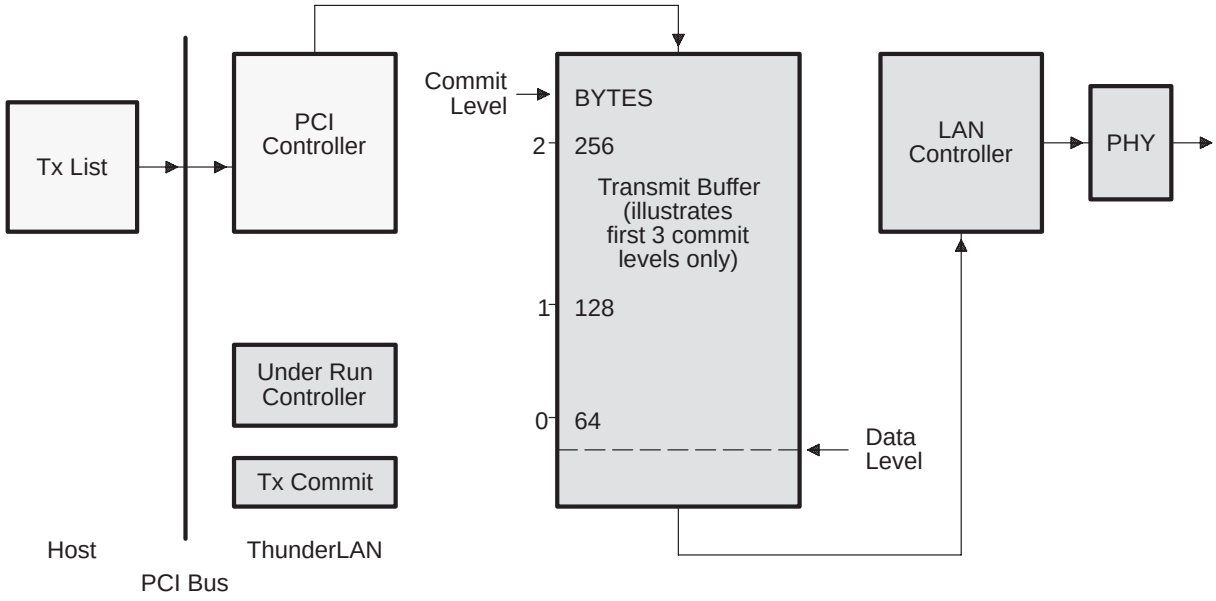

Figure 1. Pretransmit Buffer Sizing

Transmit Pacing of CSMA/CD Frames

Transmit pacing addresses a condition in which a server is unable to service all pending backlog requests because it is preoccupied with the retransmission of collided frames. In addition to temporarily throttling down the server to allow service of backlogged client requests, the pacing delay algorithm ensures a higher probability for successful frame transmission, which helps to resolve high server/client collision rates.

Transmit pacing of transmitted CSMA/CD frames improves the way the network uses its file servers by reducing the amount of contention between server and client workstations. Figure 2 shows how the pacing algorithm automatically delays new adapter frame transmission in contention situations. If a transmitted frame either collides with another frame, or has to defer to another transmission, a pacing delay is inserted between new frame transmissions. This pacing delay continues to be inserted until 31 sequential frames are transmitted without collision or deference.

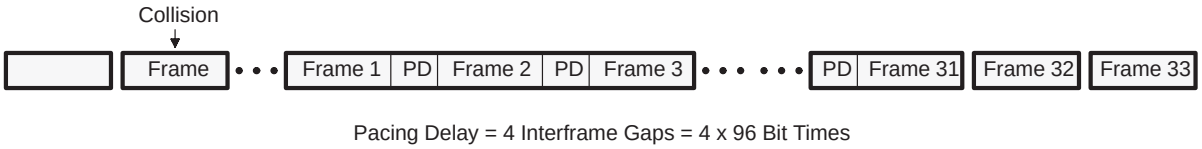

Figure 2. Transmit Pacing

Interrupt Pacing

The overall performance of a LAN adapter is determined by the performance of the hardware and the efficiency of the software network drivers. An adapter may be capable of high frame throughputs, but unless the host network drivers are capable of transferring frames to and from the adapter at these rates, that performance is not realized in a working system. A key factor for optimum performance is the rate of network interrupts experienced by the host system. Emerging high-speed LANs such as 100-Mbit/s CSMA/CD have frame rates of over 100 Kfps. Modern highly-cached CPUs have poor interrupt latencies due to the time required to switch context; these high interrupt rates degrade their performance.

APO addresses the condition of high interrupt rates using an interrupt pacing timer and an interrupt threshold counter. The pacing timer ensures a minimum time period between interrupts, while the interrupt threshold counter restricts interrupt generation.

The interrupt pacing timer disables adapter interrupts for the time period of the timer after an interrupt has been acknowledged. This ensures a minimum time period between interrupts, making certain that some host CPU time is spent outside the interrupt service routine. It also causes interrupt-bunching so more than one event can be serviced per interrupt, improving driver efficiency by reducing processor context switches.

The interrupt threshold counter can restrict interrupt generation until a number of frames have been transmitted. The servicing of interrupts for transmitted frames is a background operation, freeing up the transmit data buffers. Receive frames, unlike transmit frames, cause immediate host actions, and though a limited amount of extra latency (imposed by the pacing timer) is acceptable, each must be serviced individually. It is not acceptable to wait for three frames to be received before acting on the first because a second frame may not occur.

Figure 3 illustrates how a pacing timer can bunch interrupts for short transmit and receive frames, while interrupt thresholding restricts interrupt generation for larger transmit frames. In the first scenario for short transmit/receive frames, interrupts are not posted until the pacing timer expires. For large transmit frames, interrupt thresholding imposes an added constraint. Scenario 2, shown in Figure 3, illustrates how interrupt thresholding restricts interrupt generation for large transmit frames after the pacing timer expires.

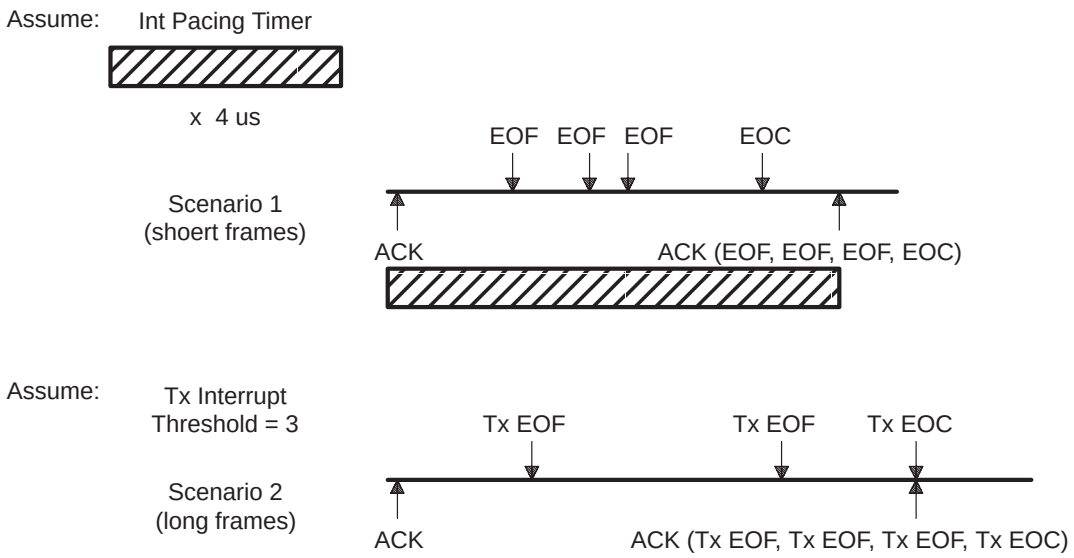

Figure 3. Interrupt Pacing and Thresholding

Two methods of reducing the number of interrupts are required to deal with variation in packet size and the different characteristics of frame reception and transmission. An interrupt pacing timer alone is effective at bunching the interrupts from short frames, but the transmission time of long packets may exceed the period of the timer. By having both pacing mechanisms, the interrupt threshold counter(s) can be used to bunch transmit frames, while the interrupt pacing timer sets the latency for, and bunches, receive frames.

Conclusion

APO uses pretransmit commit buffering together with transmit and interrupt pacing techniques to allow the adapter to adjust resources and transmit procedures to optimize adapter performance for varying system and network conditions. As system bus latency fluctuates, the pretransmit buffer size is adjusted to prevent underruns. When frame collisions occur during server transmission, transmit pacing is implemented automatically to give the server the bandwidth needed to service backlogged requests and to reduce contention between the server and client workstations. Finally, during conditions of high interrupt activity, interrupt pacing as well as the suppression of interrupts below a threshold reduces the number of adapter-to-host interrupts and the number of host operations required to service a frame. All of the APO features in the ThunderLAN drivers and architecture work together to achieve optimal performance for varying system and network environments.