Flow Control in the Cortina Systems® IXF1010 and IXF1110 10-port Gigabit Ethernet Media Access Controllers

Application Note

25 June 2007

Document Number 250236

Revision 2.0

Contents

Tables

Revision History

Revision 2.0 **Revision Date: 25 June 2007**
First release of this document from Cortina Systems, Inc.

Revision 001 **Revision Date: April 2002**
Initial release.

1.0 Introduction

This Application Note details the flow control implementation in the Cortina Systems® IXF1010 and IXF1110 10-port Gigabit Ethernet Media Access Controllers (IXF1x10 MAC). The methodology used to throttle throughput to minimize data loss is known as flow control. The focus of this document is on methods of flow control and related system implications. This document discusses the following areas of flow control:

- IEEE 802.3x Flow Control Mechanism
- Flow Control Features in the IXF1x10 MAC
- FIFO Threshold Determination
- System Implications

1.1 IEEE 802.3x Flow Control

The Ethernet MAC sublayer is required to perform two main functions: Data Encapsulation and Media Access Management. To perform these functions, First-in, First-out (FIFO) queues are used to store frame data. In an ideal world, there would be no data loss due to limitations in network performance or FIFO sizing. In the real world, data is lost without a mechanism to limit frame transmission. Data is lost when packet data is received faster than it is transmitted, resulting in filling of the FIFOs. MAC devices either overwrite data in the FIFOs (loss of oldest data) or stop writing to the FIFOs (loss of newest data). To avoid this data loss, MAC devices must slow down the receive-data stream until the transmit stream has caught up. Flow control is the methodology used to throttle the receive data stream to keep from completely filling the FIFOs. This has created an industry challenge to balance the amount of throughput loss due to flow control versus the amount of data loss without flow control.

The IEEE 802.3x flow control mechanism is accomplished within the MAC control sublayer. The FIFO begins to fill as packets are received. Once the FIFO has reached a pre-programmed threshold, the MAC Control sublayer signals an internal state machine to transmit a “PAUSE” frame. This signal informs the link partner to halt transmission for a specified length of time, referred to as “X_{off}”. The MAC continues to transmit “PAUSE” frames with the programmed idle time as long as the threshold has been exceeded. If the FIFO level falls below the threshold prior to the expiration of this time, another “PAUSE” frame is sent with a zero time specified to re-enable transmission, referred to as “X_{on}”.

1.2 Flow Control Features

The IXF1x10 MAC support the IEEE802.3x flow control mechanism. These devices have fully programmable FIFO thresholds, along with a programmable pause time, to be transmitted as data in the “PAUSE” frame.

In addition to this method of flow control, these devices have a system flow control mechanism to allow for higher layer devices to signal a need to transmit a “PAUSE” frame.

This intelligent system pause control interface allows higher layer devices to:

- select all ports to send out a X_{off} “PAUSE” frame with the programmed pause time
- select all ports to send out a X_{on} “PAUSE” frame with zero time specified
- select individual ports to send out “PAUSE” frames with the programmed pause time

Implementing flow control via this interface forces the link partner to pause transmission for the specified pause time.

1.3 FIFO Threshold Determination

The IXF1x10 MAC has per-port programmable FIFO high and low thresholds. The high threshold is the threshold above which flow control is implemented, and the low threshold is the threshold below which the flow control is terminated. Proper FIFO threshold selection determines the effectiveness of the implemented flow control. To ensure no data is lost, the FIFO threshold should be set low enough to allow for storage of the maximum amount of data that could be received prior to the flow control taking effect. To ensure maximum throughput, the FIFO threshold should be set as follows:

- High enough to not empty the FIFO prior to the flow control being released
- High enough to limit the percentage of time that flow control needs to be activated

There are many system constraints that must be considered to ensure that the optimal FIFO threshold selection is made.

1.4 Flow Control System Implications

Many system constraints must be considered when implementing flow control. Flow control effectiveness is greatly compromised if any of the following constraints are not considered:

- Packet Size
- Duplex Mode
- Link Speed
- Media Link Segment Length
- Media Link Segment Type
- MAC-PHY Latency

When selecting the FIFO threshold, both overflow and underflow conditions must be taken into account for optimal throughput at a given loss rate.

1.4.1 FIFO Depth Determination Factors to Avoid Overflow

To avoid an overflow condition, the FIFO must be able to store the amount of data that can be received prior to the flow control taking effect. This amount of data is quantified as the combination of the following:

- Media Delay 1: The amount of time for data transmitted from the link partner to reach the MAC receiver after the FIFO threshold is exceeded. This delay corresponds to data that has been transmitted but not received (data currently traveling along the media).

  All media has an inherent time delay and the length of the delay is dependent upon the type and length of the media. This delay can easily translate into large numbers of bytes at faster speeds.
- MAC Latency to Respond to Over Threshold: Preparation time to send out a "PAUSE" frame.
- Wait to Transmit: This is dependent upon duplex and supported packet size:
  - In half-duplex (not supported), the time associated with receiving the remainder of the current packet.
  - In full-duplex, the time to wait for the current transmission to end.

  The supported packet sizes impact the time in either case.
- Inter Packet Gap (IPG): The required time to wait between transmissions.
- Pause Packet Transmission: The amount of time to transmit the pause packet that needs to be accounted for.
- Media Delay 2: The amount of time for the pause packet to reach the link partner.
- MAC-PHY Latency: The amount of time to transfer the pause packet through the PHY to the MAC.
- MAC Reaction Time: The amount of time for the receiving MAC to react to the "PAUSE" frame. The maximum is specified by IEEE.
- Packet Delay: The receiving MAC could have just started transmission of another packet, which has to be completed before flow control takes effect.

1.4.2 FIFO Depth Determination Factors to Avoid Underflow

To avoid an underflow condition, the receive FIFO must have enough stored data to continue transferring for the time to terminate the flow control. The amount of stored data must exceed the combination of the following:

- MAC Latency to Respond to Under Threshold: Preparation time to send out the new "PAUSE" frame with zero time.
- Wait to Transmit: This is dependent upon duplex and supported packet size:
  - In half-duplex (not supported), this is instantaneous.
  - In full-duplex, the amount of time to wait for the current transmission to end.

  In either case, the supported packet sizes impact the time.
- IPG: The required time to wait between transmissions.
- Pause Packet Transmission: The amount of time to transmit the pause packet that needs to be accounted for.
- Media Delay: The amount of time for the pause packet to reach the link partner.

- MAC-PHY Latency: The amount of time to transfer the pause packet through the PHY to the MAC.
- MAC Reaction Time: The amount of time for the receiving MAC to react to the “PAUSE” frame. The maximum is specified by IEEE.
- Media Delay: Once the link partner has decided to re-start transmission, the delay through the media must be accounted for before the data reaches the other end.

Table 1 on page 7 documents the FIFO depth calculations for a system transmitting and receiving maximum length standard Ethernet frames at a 1000 Mbps data rate across 2000 meters of fiber.

Table 1 FIFO Depth (in Bytes) Calculations Example

FIFO Depth Calculations To Avoid Overflow	2000 Meter Fiber 1530 Byte Frame
Media Delay – already transmitted	1250
MAC Latency – respond to threshold	10
Wait To Transmit – current transmission	1530
IPG – IEEE Standard	12
Pause Packet – IEEE Standard	72
Media Delay – delay for pause to reach link partner	1250
MAC-PHY Latency – time through PHY to MAC	4
MAC Reaction Time – IEEE Standard	128
Packet Delay – just started transmission	1530
	5786 Bytes
FIFO Depth Calculations To Avoid Underflow	
MAC Latency – respond to threshold	2
Wait to Transmit – current transmission	1530
IPG – IEEE Standard	12
Pause Packet – IEEE Standard	72
Media Delay – delay for pause to reach link partner	1250
MAC-PHY Latency – time through PHY to MAC	4
MAC Reaction Time – IEEE Standard	128
Media Delay – time for new data to reach IXF1x10 MAC	1530
	4248 Bytes
Total – Overflow and Underflow	10034 Bytes

1.5 For More Information

For your specific application, please contact your local FAE for calculations to avoid underflow and overflow conditions for different packet and media combinations.

For additional product and ordering information:

www.cortina-systems.com

~ End of Document ~