White Paper

Traffic Management in Stratix GX Devices

Introduction

Data networks were designed to meet the rapidly increasing bandwidth requirements of Internet traffic, which increased line rates by 400 percent every 2 to 3 years. These initial networks were able to handle e-mail and web traffic, but were not able to make service providers a profit.

The communications industry is now looking to develop more intelligent networks that reduce operational expenditures and increase profits. The goal of current network planners is to make more efficient use of the existing bandwidth as well as offer additional services, like voice or video support. This will be enabled by efficient traffic management across the entire network.

This white paper discusses traffic-management concepts and the implementation of traffic-management functions within Altera® Stratix™ GX devices. It also provides a detailed discussion on memory and memory interfacing, which are critical to effective traffic-management implementations.

Traffic Management

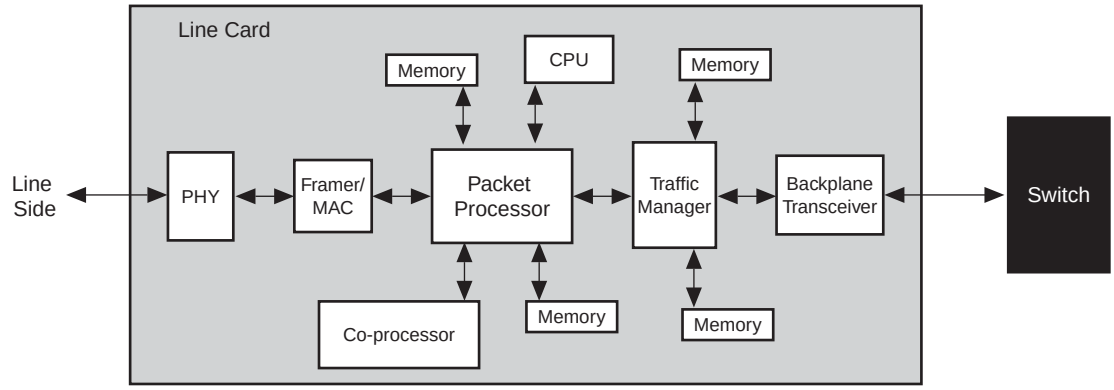
Traffic management enables bandwidth management through the enforcement of service level agreements (SLAs), which define the criteria that a network must meet for a specified customer or service. These criteria include: guaranteed bandwidth, packet loss, latency, and jitter. Guaranteed bandwidth, or throughput, can include minimum, average, and peak guarantees of bandwidth availability. Packet loss is the number or percentage of packets sent but not received, or received in error. Latency refers to the end-to-end delay of packets, and jitter is the delay variation for consecutive packets.

Line Card

You can use Stratix GX devices to implement traffic-management functions on a line card. Figure 1 illustrates a typical line card, featuring the traffic manager in the data path. (The traffic manager can also sit outside of the data path and communicate only with the packet processor, which is called lookaside mode.) The line card's packet processor classifies the ingress data traffic—or the data traveling from the line side toward the switch—and decides which port the data should exit through. The packet processor also modifies the data header by adding the appropriate class. This data header helps the traffic manager enforce the SLA requirements. For egress traffic—or the data traveling from the switch to the line side—the traffic manager will smooth out large spikes in traffic, allowing the networks to run more efficiently. Traffic managers typically reside on a line card next to a backplane transceiver because they implement the output queuing necessary for the switch.

Altera's Stratix GX FPGAs support the integration of backplane transcievers into a custom traffic manager. Stratix GX FPGAs feature 20 full duplex transceiver channels, each capable of operating at up to 3.125 gigabits per second (Gbps) per channel while consuming minimal power. Additonally, they feature up to 45 transmitter and 45 receiver channels based on the source-synchronous signaling scheme. The source-synchronous receivers also feature optional dynamic phase alignment (DPA) circuitry that simplifies skew problems arising from board layout complexities. Using these transceiver blocks and source-synchronous channels, Stratix GX FPGAs can interface with other high-speed devices on the board such as a packet processor, backplane, host processor, and memory.

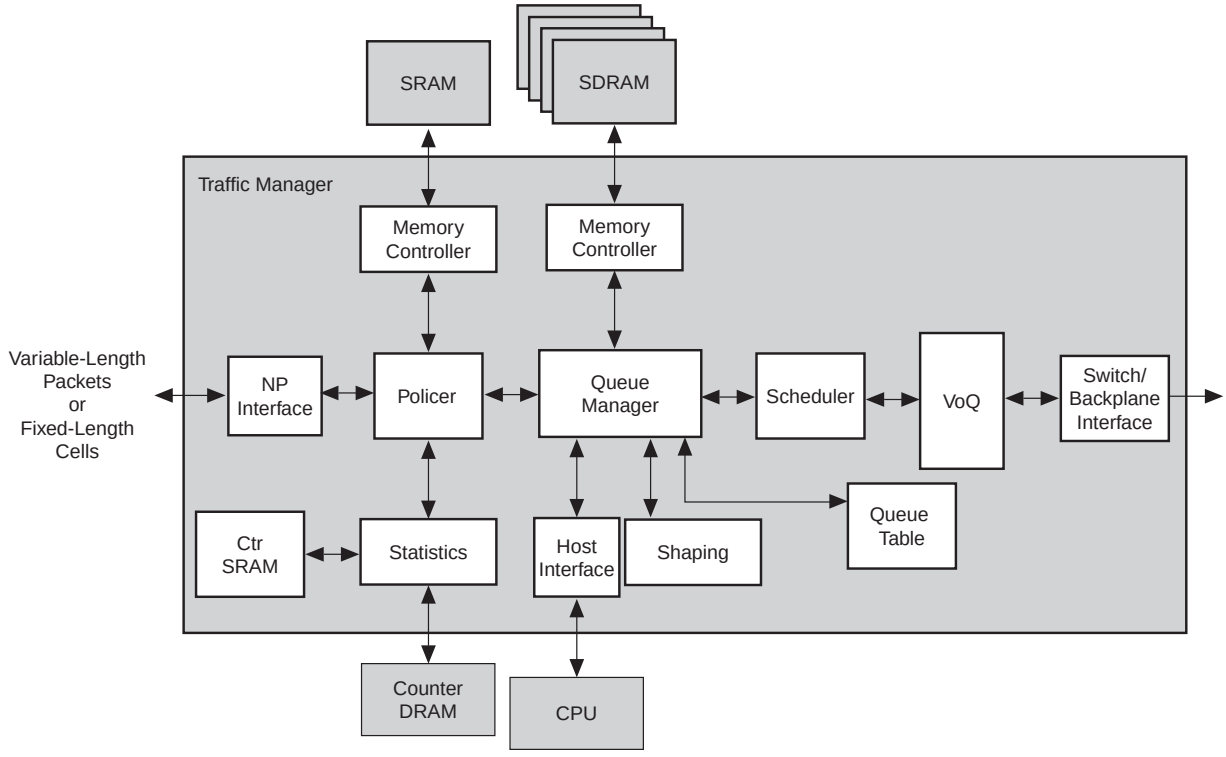
Figure 1. Line Card Diagram



Traffic Managers

Figure 2 shows a block diagram of a generic traffic manager. The data arriving into the traffic manager could be complete variable-length packets or it could be fixed-length cells. Many traffic managers or packet processors will segment packets into fixed-length cells because switch fabrics can be optimized for switching those fixed-sized cells. The modified header of incoming data traffic allows traffic managers to prioritize and decide which packets should be dropped and retransmitted, when packets should be sent to the switch fabric, and how traffic should be shaped when sending onto the network.

Figure 2. Traffic Manager Diagram (Note 1)



Note to Figure 2:
(1) All traffic managers do not implement all of the blocks in the diagram. Additionally, some of the functions in the diagram may be implemented within the packet processor.

Traffic Manager Line Interfaces

Historically, the data path interfaces for the traffic manager were proprietary and had many different available options. The interface to the packet processor now is standardizing on the source-synchronous interface SPI-4.2 because of pin efficiency and the availability of compliant products. The switch fabric interface is still in flux, with some current solutions supporting Common Switch Interface Layer 1 (CSIX-L1), and many others that are still proprietary. The Network Packet Switching Interface (NPSI) was recently finalized to improve the pin efficiency of CSIX-L1 while keeping some of its protocol advantages.

Using the embedded transceiver blocks as well as the source-synchronous transmitter, receiver, and DPA features, Stratix GX FPGAs support a variety of emerging high-speed protocols such as 10 Gigabit Ethernet (XAUI), SPI 4.2, NPSI, CSIX-L1, Utopia Level 4, HyperTransport™, and others. The embedded high-speed transceiver circuitry is complemented by very high-speed I/O buffers that support I/O standards such as LVDS, PCML, and LVPECL that are needed for high point-to-point data transfer rates.

Policer

The policer makes decisions based on which packets to drop. A dropped packet will force a re-transmission by the TCP layer from the source node. Dropping packets intelligently is an important method to avoid and prevent congestion within a network. When buffers overflow, the need to drop packets is obvious. Methods to prevent this overflow in advance by dropping packets intelligently can be used to increase the efficiency of the network. Algorithms used for dropping packets include weighted random early detection (WRED), which pseudo-randomly drops packets based on a priority weighted by their SLA. Instead of dropping packets, policers may choose to mark packets that are not obeying the SLA, and drop them first when congestion occurs.

Scheduler

The scheduler determines which of the non-empty queues should be used for transmitting the next packet. Algorithms used include weighted round robin (WRR) or weighted fair queuing (WFQ). WRR services non-empty queues in a round robin fashion, but accepts more data from higher-weighted queues. This prevents highly congested, high priority queues from completely blocking out lower priority queues, while also enforcing the necessary prioritization. WFQ is similar to WRR, but more complex in that it also accounts for variable-sized packets.

Queue Manager

The queue manager maintains the traffic queues, which are pointers to the packet data stored in external memory. A separate queue can be kept for each traffic class or traffic flow. An example of these pointer memories can be found in the section titled “Memory” on page 4.

Shaper

The shaper is similar to policing in that it uses token or leaky buckets. While policing drops packets, shaping delays packets, but it provides temporary buffering to ensure the outgoing traffic fits an appropriate profile.

Statistics

Also called metering, statistics are kept to provide information on whether packets do not meet appropriate SLAs. Metering is also used to enable dynamic billing based on usage.

High-speed counters are needed to perform traffic-management metering at high speeds. The results of these counters are stored in memory. A hierarchical memory structure can be used to support the large number of counters necessary for keeping statistics. An example of this type of memory can be found in the section titled “Memory” on page 4.

You can implement high-speed counters in Stratix GX logic elements (LEs). Such counters are capable of running at speeds more than 300 MHz.

Memory

While communications link speeds have grown by 4 times every 2 to 2.5 years, memory performance has only increased by 10% per year. This growth discrepancy has led to a critical memory bottleneck in communications systems. Current traffic managers require large amounts of embedded memory as well as support for high-speed external memories.

The amount of external memory required is application-dependent, but there are a few general guidelines. Because the data will be written into memory and eventually read out of memory, the memory throughput must be at least two times the line rate. If header information is added to the data as it is processed, the throughput requirements will increase to up to four times the line rate. The total size of the memory in many cases is bounded by the round trip time (RTT) of the TCP. This is the average round trip time between active hosts, and it can range from 200 to 300 ms. In this case, a 10 Gbps interface would require 2 to 3 Gbps of memory.

Table 1 shows the different external memories that are used in traffic-management applications.

Table 1. External Memory Guidelines

Feature	SDRAM	SRAM	CAM
Latency	High	Low	Very low
Density	High	Low	Low
Cost	Low	High	Very high
Power	Low	Medium	Very High
Applications	Packet Buffer	Pointers Flow tables Rate tables	Search Classification

SRAM is more expensive than SDRAM and too costly to use for packet buffering. This limits SRAM usage to smaller, latency-critical memories like pointer lookup. Any external SRAM that can be integrated onto the traffic management chip will improve the overall latency of the solution. Content-addressable memory (CAM) has the highest latency, but it comes at the expense of cost and power. CAM is used in packet processing functions such as classification.

SDRAM is inexpensive and has high bandwidth but also higher latency compared to SRAM. Therefore, SDRAM is used for functions with very high density needs. SDRAM is used to buffer the data as it is being processed. SDRAM also requires many pins. The example below shows the number of pins required to interface to a 64-bit SDRAM. Many types of high-end networking equipment will require several of these devices, leading to very high pin requirements for traffic-management devices (see Table 2.)

Table 2. Pin Requirements for Traffic Management Devices

Pin Name	Function	Pins
A[0..12]	Address bits	13
BA[0..1]	Bank address	2
DQ[0..63]	Data in/out	64
DQS[0..7]	Data strobe	8
CK[0..2]	Clock	3
!CK[0..2]	!Clock	3
CKE[0..1]	Clock enable	2
CS[0..1]	Chip select	2
RAS	Row address	1
CAS	Column address	1
WE	Write enable	1
DM[0..7]	Data – in mask	8
Total Pins		108

Stratix GX FPGAs are available in advanced pin packages that provide board area savings as well as high pin-counts. For example, Stratix GX devices are offered in the 1,020-pin FineLine BGA® package with up to 589 user I/O pins. These high pin counts provide easy access to I/O pins while interfacing with multiple memory chips and other devices on the board.

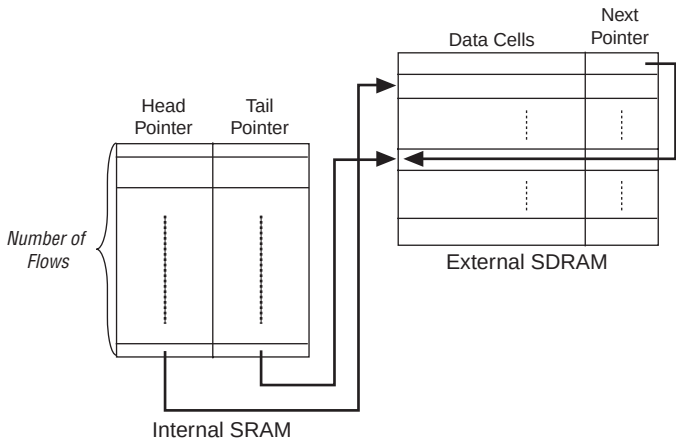
Stratix GX FPGAs provide interfaces to high-speed external memory devices such as SDR SDRAM, DDR SDRAM, FCRAM, RLDRAM, quad data rate (QDR) SRAM, and zero-bus turnaround (ZBT) SRAM. Altera also provides memory controllers (such as DDR memory controllers) in the form of intellectual property (IP) that can be implemented in LEs.

Internal Memory

The use of internal SRAM reduces pins, power, board space, cost, and latency. Stratix GX FPGAs provide abundant embedded TriMatrix™ memory that is capable of handling the memory requirements of traffic management. For example, Stratix GX devices offer up to 3.4 megabits (Mbits) of memory that provides an internal memory bandwidth of 4.4 terabits per second. The TriMatrix memory consists of three types of memory blocks: M512, M4K, and M-RAM blocks. M512 blocks support 512 bits of memory, M4K blocks support 4 Kbits, and M-RAM blocks offer up to 512 Kbits of memory per block. Each of these blocks can be configured in different operation modes.

For traffic managers, queue memory pointers are used to keep track of the various flows being serviced. Figure 3 shows one implementation of these memory structures. The internal SRAM stores both the head and tail pointers for each flow, while the external SDRAM contains the actual data plus pointers to the next piece of data within the flow. This creates an overall linked list for each flow. When a new piece of data is added to a flow, the tail pointer is updated for that flow. When a piece of data is scheduled to be sent to the switch fabric, the head pointer for that flow is updated.

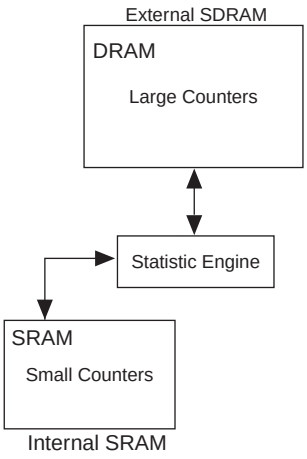
Figure 3. Memory Structure Example



The Stratix GX M-RAM blocks are optimal for storing pointers to external memory locations. For example, with a 128-Mb SDRAM containing fixed 64-byte cells, Stratix GX M-RAM blocks can store both head and tail pointers for 64K flows.

Another example of the use of internal memory is for creating a hierarchical memory structure needed to support statistics counters. The need for hierarchical memories again arises from the external DRAM memory bottleneck. The throughput of current DRAM technologies cannot meet the requirements of updating numerous counters per cell at line rate due to the inherent latency of DRAM. This requires temporary counters to be stored in SRAM, which then are used to occasionally update the external DRAM (see Figure 4).

Figure 4. Internal Memory Structure Example



In this example, the DRAM latency is only incurred periodically, which is determined by the size of the SRAM counters. The statistics engine will update the appropriate small counter values in SRAM as packets are received. Periodically, it will then read the large counter values from external DRAM and add the small counter values. Then it will reset the small counter value to zero in the internal SRAM. The M4K blocks within Stratix GX FPGAs can be configured to temporarily store the count values for each counter. For example, for 64K flows, the M4K blocks can be used to store up to three 8-bit counters for each flow. This will theoretically reduce the number of times the external DRAM needs to be accessed by up to 2^8.

Conclusion

FPGAs are a natural fit for implementing traffic managers because of the limitation of this risk and the ability for equipment vendors to differentiate their traffic management solution. Additionally, a re-configurable solution can be used to add and support new services in the future.

Featuring integrated transceivers that enable direct connection to the backplane, Stratix GX FPGAs meet the requirements for efficient traffic management solutions. Additionally, the TriMatrix memory blocks allow flexible memory management at today's highest traffic rates, with support for future memory standards. The embedded memory structure of Stratix GX devices enables storage of pointer tables into the large M-RAM blocks and statistic caches in M4K blocks. Stratix GX FPGAs offer a complete fabric to implement high-speed traffic-management solutions.

101 Innovation Drive
San Jose, CA 95134
(408) 544-7000
http://www.altera.com