Dave Bursky, **DIGITAL ICs/DSP EDITOR**

ETHERNET-PORT AGGREGATION CHIP SLASHES SYSTEM COST

Aggregating device reduces bandwidth requirements by combining two dozen 10/100/1000 Ethernet ports into a single 10-Gbit/s channel.

Today's Ethernet network connections typically operate at under 50% of the available bandwidth utilization due to many factors, but especially the bursty nature of the data being sent. However, equipment aggregating Ethernet links, like switches and routers, are usually designed for 100% link utilization. This results in the underutilization of bandwidth, increased equipment cost, and a reduced number of customers served per system.

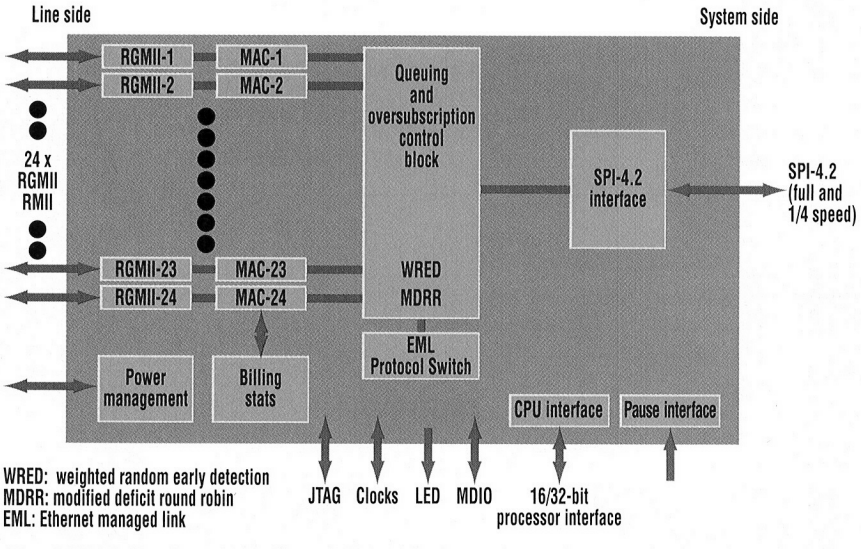
To maximize revenues per system, an oversubscription-based solution can be used. In this approach, unused bandwidth can be eliminated at the physical connection so the data stream can fully utilize shared hardware, such as the network processors, the backplane, and the switch fabric.

Ample Communications has developed such a solution. Its Harrier family of port aggregation chips provides intelligent oversubscription support. These chips can aggregate up to 24 Gbits of bandwidth from customers to a highly utilized 10-Gbit/s system datapath.

By using weighted-random-early-detection (WRED) and modified-deficit round-robin (MDRR) based algorithms, along with in-band PAUSE frame generation, the scheme can minimize frame drops and ensure that lower-priority and most offending customers are penalized in times of traffic congestion. The resulting system-level solution reduces hardware cost per customer by up to 40% and lets system vendors double the number of customers served per system, lowering the cost per customer.

Ample's oversubscription scheme uses intelligent buffering. It enables the aggregation of up to 24 1-Gbit/s ports into a single 10-Gbit/s channel, reducing the number of 10-Gbit/s interfaces by almost 2.5 times. The approach leverages the fact that data tends to be fairly "bursty," with large idle periods separating streams of bits.

By using that aspect and implementing large on-chip buffers and intelligent buffering algorithms, Ample crafted its highly integrated A2510 Harrier chip. It greatly reduces the number of components required to create an Ethernet solution with oversubscription and billing support. The chip also includes protection switching to better implement nonstop systems. The A2510 aggregates 24 Ethernet ports on the line side, with each port able to handle 10/100/1000-Mbit/s Ethernet traffic. On the host side, the chip implements a 10-Gbit/s SPI-4.2 interface *(see the figure)*. The system-peripheral-interface (SPI) port also will operate at quarter speed.



The A2510 Harrier chip offers 24 ports capable of 10/100/1000-Mbit/s traffic on one side and a 10-Gbit/s SPI-4.2 interface on the other. Its intelligent buffering scheme combines the 24 ports of traffic (up to 24 Gbits/s) into one 10-Gbit/s stream or unpacks a 10-Gbit/s stream into 24 separate 1-Gbit/s streams.

Each Ethernet port supports industry-standard RMII/RGMII version 1.3 interfaces that connect to readily available Ethernet physical-layer devices. For the 10/100-Mbit/s mode, the interfaces can be programmed to operate in the RMII mode to take advantage of the low-cost physical-layer chips. Each line-side port can be independently set to run at 10 or 100 Mbits/s, or 1 Gbit/s.

On the system side, the chip's standard optical-internetworking-forum (OIF) system packet interface level 4, phase 2 (SPI-4.2) port can be configured to run at full speed (311 to 400 MHz) or at quarter speed (77.7 to 100 MHz) for a total data throughput of up to 12.8 Gbits/s and 2.5 Gbits/s, respectively. The company will also offer two other versions of the Harrier chip. One is a 12-port line aggregation chip (the A2511). The other has 24 Ethernet ports, which are limited to 10- or 100-Mbit/s throughputs (the A2512).

Each chip includes an extensive set of statistics counters on both the transmit and receive sections for billing applications. These counters collect statistics on different types of frames sent, received, and dropped, as well as CRC errors, jabbers, byte counts, and aborts. Designers also packed enough on-chip memory (over half a megabyte) to support transmit, receive queues, and FIFO buffers so each port can handle the Ethernet jumbo frames of up to 9.6 kbytes without any external memory. On-chip media-access controllers (MACs) can also generate and process in-band PAUSE frames for flow control.

An on-chip Ethernet management link (EML) pass-through permits flexible management and redundancy protection of the Ethernet ports. When deploying this feature, Ethernet frames are optionally forwarded over the SPI-4.2 interface with the first 8 bytes designated as preamble and the SFD bytes unchanged. Systems can then perform operation, administration, and management functions based on information embedded in these bytes. Also, the A2510 Harrier chip's internal logic helps implement 1+1 or 1:N protection switching between independent line ports.

The chip's ability to oversubscribe the Ethernet ports starts with a pool of queues associated with the 24 Ethernet ports. The queues are allocated dynamically in page increments and can grow to support multiple jumbo frames. Once a port is serviced, the corresponding queues are de-allocated and go back into the available pool. To avoid congestion, a flexible layer 2 WRED-based scheme is used to limit incoming data rate. In this approach, frames are dropped with some probability if a set threshold is exceeded. Anticipating congestion and dropping frames early prevents congestion due to bursty traffic.

WRED provides up to four programmable thresholds associated with each of the two queues (high priority and low priority). Corresponding to the four thresholds are four drop-programmability levels used to create four threshold-probability pairs. For priority traffic, thresholds can be set on selected ports to guarantee no frame drops.

Thus, the A2510 can aggregate 24 multirate ports on the line side and, similarly, support up to 24 ports on the system side over the SPI-4.2 port. If a network processor can handle 24 SPI ports, then mapping is 1:1 and no special tagging is required. Yet a number of current 10-Gbit network processors only handle up to 16 SPI ports.

To compensate, the chip allows m:1 RGM11:SPI mapping, where m can be either 1, 2, or 3. In this scheme, a tag is added to each frame. The tag identifies the line-side port during transmit and receive operations. Complete frames are then transmitted across the SPI-4.2 interface for m>1.

As mentioned earlier, the chips support protection switching. This is done in conjunction with a higher-layer management function. One or more ports on the chip can be dedicated as the "protect" ports, and others as the "working" ports. In case of port failure, the higher-layer management software can command the Harrier to switch from the failing port to a protect port. Then it will remap the RGMII to SPI ports as part of the swap. After the switch, data flows in and out of the protect port. When the failed port is repaired, the chip can be commanded to switch back to the original port.

Other chip features include power management, an external PAUSE control interface, and four LED indicator drivers. The power-management function enables software to control the ports and interface to shut down and reactivate each port. As a result, the system can reduce power when ports are idle. The external PAUSE interface lets the system generate PAUSE frames "out-of-band" for system-level flow control and diagnostic purposes. For basic control operations, the aggregation chips all include a standard 16/32-bit processor interface and a master MDIO interface.

The aggregation chips support an end-to-end flow-control mechanism using the PAUSE frame as specified in IEEE 802.3x. On the line side, each port has its own independent flow-control processing. On the transmission side, the received PAUSE frame causes the MAC to stop transmission for the required number of PAUSE quanta. When the quanta expire, the controller resumes transmission.

The chips also allow for fiber latency by providing fiber-latency buffering via on-chip storage. Distances of up to 5 km (one way) can be buffered for every port on the chip. Each port's range can be extended or reduced via program control as long as the total buffering doesn't exceed the available memory.

All three versions of the Harrier chip will be housed in 784-contact 31- by 31-mm FCBGA packages. The chip's core can operate from a 1.8-V supply. The I/O pins for the control and CPU interface run from a 2.5-V supply but are 3.3-V tolerant. Along with the three versions, Ample offers development support, an evaluation board, and associated software drivers. ED

PRICE & AVAILABILITY

The Harrier family of Ethernet port aggregation devices consists of three chips: the A2510, 2511, and 2512. The A2510 packs 24 1-Gbit/s-capable Ethernet ports and can aggregate them via oversubscription to a single 10-Gbit/s SPI-4.2 port. It costs $195 in 10,000-unit quantities. Packing a dozen 1-Gbit/s-capable ports, the A2511 sells for $150, and the A2512 offers 24 10/100-Mbit ports and runs $150, both also in 10,000-unit quantities. Samples will be available later this quarter.

AMPLE COMMUNICATIONS INC.
Ken Madison, (510) 657-1500, ext. 138
www.amplecomm.com