Product Brief
June 1999

microelectronics group

Lucent Technologies
Bell Labs Innovations

LUC4AP10
High-Speed Switching ATM Port Controller (APC)

Introduction

The ATM port controller (APC) IC is part of the Lucent high-speed switching chip set that provides a highly integrated, innovative, and complete VLSI solution for implementing the ATM layer functionality/core of an ATM switch system. The chip set enables construction of high-performance, feature rich, and cost-effective asynchronous transfer mode (ATM) switches, scalable over a wide range of switching capacities.

Features

- Provides a comprehensive single-chip solution for implementing all ATM layer functions needed at an ATM switch port.
- Can be configured in a variety of switching modes for flexible operation:
  - — Performs as an ATM switch port card by supporting linear aggregation of up to 622 Mbits/s of ATM traffic at the physical layer interface (full duplex).
  - — Operates as a stand-alone, single-chip 32 × 32 shared memory switch or an N:1 concentrator (622 Mbits/s total switching).
  - — Operates in conjunction with another APC as a 2 × 2 (1.2 Gbits/s total switching capacity) dual APC-based switch (no separate external switch fabric needed).
- Performs ATM layer user network interface (UNI) and network node interface (NNI) management functions.
- Provides two independently operating full-duplex UTOPIA II compatible interfaces:
  - — Controls up to 31 full-duplex multiple physical layer (MPHY) ports on the physical layer side.
  - — Allows either UTOPIA interface to operate as 16-bit or 8-bit data.
  - — Allows any MPHY to be configured as a UNI or NNI.
- Supports up to 64K connections with scalable external memory:
  - — Manages virtual connection parameter table in external memory.
  - — Optionally performs the ATM Forum compliant dual leaky-bucket policing for up to 64K VCs.
  - — Facilitates call setup and tear down through VC parameter table update via high-performance microprocessor port.
  - — Optionally translates or passes the generic flow control (GFC) field of the egress ATM cell header for NNI or UNI applications.
  - — Performs virtual path identifier (VPI)/virtual channel identifier (VCI) translation for up to 64K connections on egress while allowing reusability of same VPI/VCI on different UNIs.
- Maintains a variety of optional per-connection, per-port, and per-device statistics counters in external memory and on-chip.
- Provides dual interfaces to high-speed switching switch fabrics to facilitate construction of redundant systems for fault tolerance.
- Supports spatial and logical multicasting for up to 32 destination ports on egress (31 MPHY ports and 1 microprocessor interface port).
- Provides a generic 32-bit microprocessor interface with interrupt.
- Supports high-speed read and write direct memory access (DMA) modes for cell extraction and insertion via the microprocessor interface.
- Provides input/output (bidirectional) queue management for an N × N switch fabric for over 100 Gbits/s capacity and up to 31 MPHY ports and one microprocessor port:
  - — Queues up to 512K ATM cells in external memory, organized in a per-VC, fully-shared queueing architecture.
  - — Supports five traffic classes via novel scheduling algorithms, including a WFQ type scheduler to provide per-VC QoS assurance.

Features (continued)

- Implements a flexible, efficient buffer/congestion management scheme based on a Bell Labs' patented adaptive dynamic thresholding (ADT) algorithm:
  - — Supports selective cell discard and early/partial packet discard (EPD/PPD).
  - — Provides capability for minimum buffer reservations on per-connection, per-class, and per-port basis.
- Performs ATM Forum-compliant available bit rate (ABR) explicit rate flow control using a Bell Labs' patented algorithm. Provides optional support for EFCI marking as well.
- Provides operations, administration, and maintenance (OA&M) Fault Management functions for loopback, continuity check, defect indication on all connections, and performance monitoring on up to 127 connections.
- Can be used in conjunction with an external switch fabric (LUC4AS01 (ASX) and LUC4AC01 (ACE) ICs, part of the high-speed switching chip set) to provide a scalable, nonblocking switch solution for over 100 Gbits/s capacity.
- Supports bandwidth scalability to OC-48 rates via use of external PIMUX devices on the PHY layer side.
- Provides an enhanced services interface (ESI) to support operation of an optional external adjunct device for comprehensive statistical data collection.
- Facilitates circuit board testing with on-chip *IEEE** boundary scan.
- Fabricated as a low-power, monolithic IC in submicron 0.25 µm, 3.3 V CMOS technology, with 5 V tolerant and TTL-level compatible I/O.
- Available in a 600-pin LBGA package.

Applications

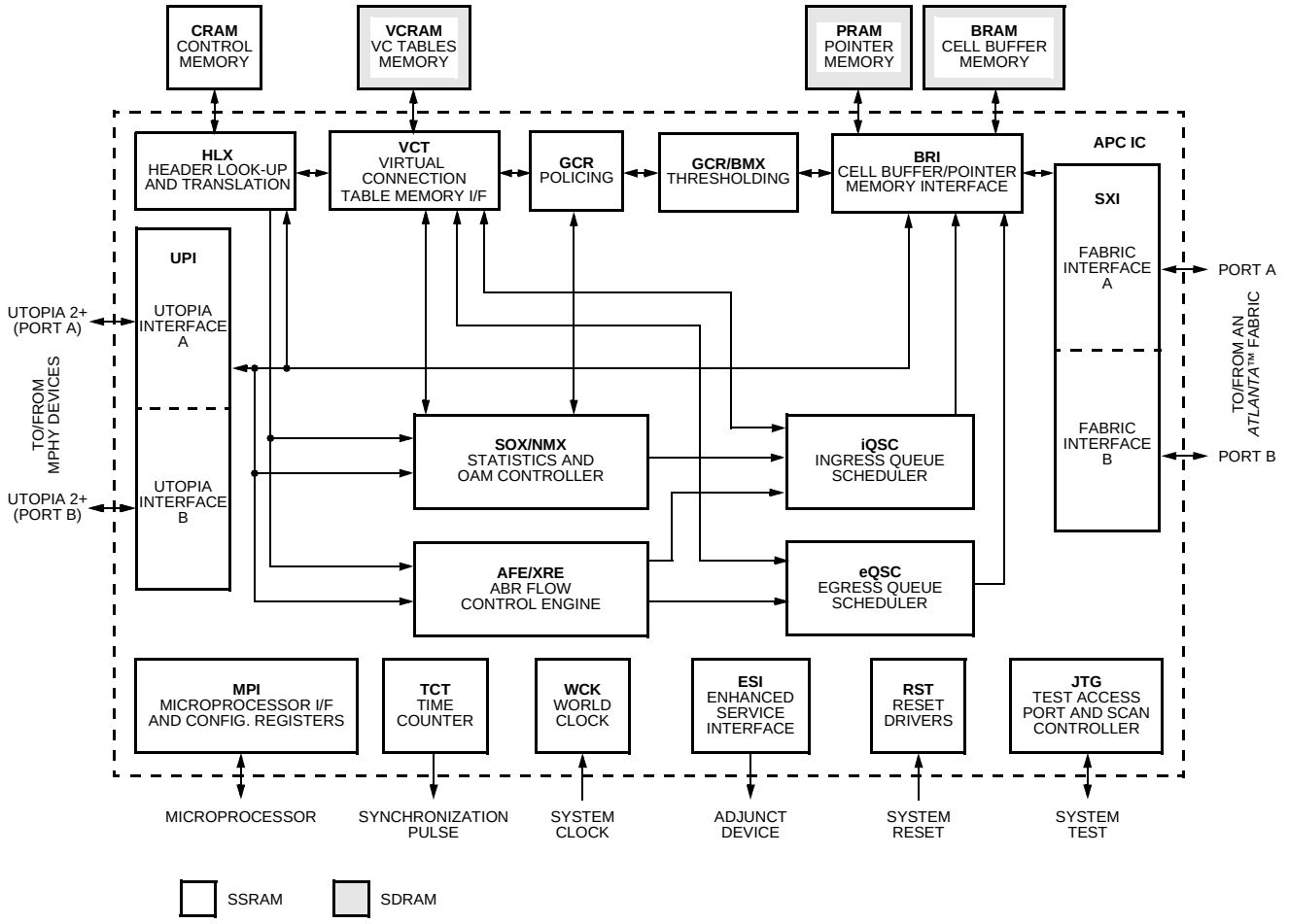
- ATM switches
- x-DSL systems
- DLC systems
- ADM
- Access multiplexers
- Routers
- PBX
- Wireless infrastructure equipment
- VP rings

* *IEEE* is a registered trademark of The Institute of Electrical and Electronics Engineers, Inc.

Description

Block Diagram

Figure 1 presents a block diagram of the APC chip. Each high-level block function is summarized in this section.



5-8009(F)r.2

Figure 1. High-Level Block Diagram of the APC IC

Description (continued)

UTOPIA Interface (UPI)

The UPI controls the transfer of ATM cells between the APC and multiple physical layer devices (MPHYs) connected to it, via two full-duplex UTOPIA Level II compatible interfaces operating independently at up to 52 MHz.

The UPI supports linear aggregation of ATM traffic up to 622 Mbits/s. Each UTOPIA interface can be enabled independently and is always a bus master. Receive parity checking can be enabled or disabled. Transmit parity generation can be configured as odd or even. Each transmit interface can be configured to internally loopback cells to its receive interface independently for diagnostic purposes.

Additionally, the UPI can support an extended cell format and an external port inverse multiplexer (PIMUX) device to interface with link rates greater than 622 Mbits/s such as OC-48 trunks.

Each interface can be independently configured to operate in 16-bit or 8-bit mode (for UTOPIA Level 1 compatibility), with a standard or extended cell format and its own contiguous UTOPIA poll address range to support up to 31 MPHY devices*.

For special cells (OA&M and RM), the UPI also checks (on receive) and inserts (on transmit) the payload CRC-10 field.

Header Look-Up and Control Memory Interface (HLX)

The HLX performs a three-level ingress look-up and one-level egress look-up in external memory to fetch connection information for up to 64K ingress and egress connections. The HLX supports MPHY port configuration (UNI or NNI), VP or VC switching, ATM header validation, ingress cell capture and cell recognition (e.g., RM and OA&M cells).

Additionally, the HLX accesses control memory tables maintained in external synchronous static memory, SSRAM (CRAM), to support queuing, scheduling (QSC), and OA&M (NMXSOX) operations.

* Only 31 MPHY devices can be supported over the two UTOPIA interfaces combined.

Virtual Connection Table Memory Interface (VCT)

The VCT handles all the necessary operations needed to read, write, and refresh the ingress and egress virtual connection parameter tables, maintained in external SDRAM (VCRAM) for up to 64K connections.

Policing Generic Cell Rate (GCR)

The GCR implements two instances of the generic cell rate algorithm (GCRA) for each connection to police cells for conformance to their negotiated traffic contracts. It uses the virtual scheduling algorithm outlined in ITU-T I.371 to determine conformance. Seven possible policing configurations are supported for each GCRA instance including all the conformance definitions specified in ATM Forum Traffic Management Specification Version 4.0.

Policing can be enabled or disabled on a per-connection basis, or globally enabled or disabled. The policing action, monitor, tag, or drop can be programmed separately for each individual leaky bucket on a per-connection basis. The range of policed rates goes from 424 bits/s (one cell per second) to the maximum supported rate of 622 Mbits/s.

Buffer Management Controller (BMX)

The GCRBMX block performs buffer management functions, by checking a variety of per-connection, per-port, and per-traffic class thresholds. They are used to drop a cell before it is queued.

A CLP0+1 discard threshold, an optional CLP1 discard threshold, and an optional early packet discard (EPD) threshold are provided for each connection on ingress and egress. These thresholds can be configured as static or dynamic.

The dynamic threshold is based on an effective buffer allocation for the connection. This allocation is not used directly as the discard threshold. The overall reserve and specific congestion conditions add to or subtract from the buffer allocation to continuously adjust the threshold. These adjustments are based on the available reserve of the common buffer pool and the congestion threshold established for the traffic class to which the connection belongs. This balances the losses of the given circuit against the overall state of the system. A guaranteed minimum buffer space is supported during periods of heavy buffer use.

Description (continued)

Buffer Management Controller (BMX) (continued)

The buffer is logically partitioned so that each traffic class can support a different buffer reservation independent of other classes. Additional mechanisms are provided to control the sharing of buffers between ports with minimum guarantees for each port.

An optional partial packet discard (PPD) mode is supported for each connection. A static selective explicit forward congestion indicator (SEFCI) marking threshold is also provided. EFCI marking can be selectively enabled for each connection.

Cell Buffer Memory Interface (BRI)

The BRI handles all the necessary operations needed to read, write, and refresh the cell data in ingress and egress cell buffer pools maintained in external SDRAM (BRAM). The BRI supports up to 512K cell buffers which can be statically partitioned between ingress and egress without restriction. The BRI performs cell enqueue and dequeue operations and serves also as intermediate cell storage to support NMXSOX and XREAFE operations.

Additionally, the BRI manages the communication between the APC and the ingress and egress cell pointer memory maintained in external SDRAM (PRAM).

Ingress Queue Scheduler (iQSC)

The iQSC implements a sophisticated per virtual connection queue architecture and hierarchical scheduling algorithms for ingress.

The iQSC performs the virtual connection enqueue and dequeue operations, which are supported by the HLX, VCT, and BRI blocks.

The iQSC provides novel scheduling algorithms to service up to 64K VC queues and five traffic classes (CBR, rt-VBR, nrt-VBR, ABR, and UBR) using a three-level hierarchy. The first (VC) level of the hierarchy contains per-VC schedulers for each of the five traffic classes. Each VC scheduler can support all 64K VC queues as needed. The second (class) level of the hierarchy contains two class schedulers: one to provide guaranteed bandwidth and one to distribute excess bandwidth to each work-conserving VC scheduler. The third (port merge) level of the hierarchy merges the CBR VC scheduler and the two class schedulers together.

The CBR scheduler uses a non-work-conserving shaped virtual clock (ShVC) algorithm to shape CBR virtual connections to any one of up to 32 user-programmable rates.

The rt-VBR scheduler uses a work-conserving starting potential fair queuing (SPFQ) algorithm to schedule virtual connections using any one of up to 16 user programmable rates/weights.

nrt-VBR, ABR, and UBR schedulers each use a work-conserving variable length frame based weighted round-robin (WRR) algorithm to schedule virtual connections using one of up to 256K different weights.

The guaranteed traffic class scheduler (GTS) uses the ShVC algorithm to provide guaranteed bandwidth to each of four classes (the rt-VBR, nrt-VBR, ABR, and UBR VC schedulers). A user-programmable rate is provided for each class.

The excess bandwidth class scheduler (EBS) uses a work-conserving self-clocked fair queuing (SCFQ) algorithm to fairly allocate excess bandwidth to each of four classes (the rt-VBR, nrt-VBR, ABR and UBR VC schedulers). A user-programmable weight is provided for each class.

The port merge scheduler merges the CBR VC scheduler and the GTS class scheduler together with strict priority over the EBS class scheduler. CBR and GTS service is arbitrated at this level using timestamps.

Switch Fabric Interface (SXI)

The SXI implements an interface to *ATLANTA* switch fabrics, as well as a loopback mode for a single APC stand-alone operation. The SXI interface consists of an 8-bit parallel data bus plus additional control signals and clock, operating at up to 100 MHz. The SXI terminates the simple *ATLANTA* fabric interface protocol. The transmit interface can be configured to generate odd or even parity. The receive interface checks for odd parity and protocol violations. It also extracts the fabric backpressure information from the cell header.

This interface is duplicated to allow construction of redundant switch fabric architectures for fault tolerance, and 1.2 Gbits/s fabric-less switches using two APCs. Each transmit interface can be independently enabled and disabled.

Description (continued)

Egress Queue Scheduler (eQSC)

The eQSC implements a sophisticated per virtual connection queue architecture and hierarchical scheduling algorithm for egress.

The eQSC performs the virtual connection enqueue, dequeue, and subport spatial and logical multicast operations, which are supported by the HLX, VCT, and BRI blocks.

It provides novel scheduling algorithms to service up to 64K VC queues, five traffic classes (CBR, rt-VBR, nrt-VBR, ABR, and UBR), and 32 subports using a four-level hierarchy.

Subports 0 to 15 are designated as type 1 and have dedicated three-level hierarchical schedulers similar to iQSC. The first (VC) level of the hierarchy contains per-VC schedulers for each of the five traffic classes. Each VC scheduler can support all 64K VC queues as needed. The second (class) level of the hierarchy contains two class schedulers: one to provide guaranteed bandwidth and one to distribute excess bandwidth to each work-conserving VC scheduler. The third (port merge) level of the hierarchy merges the CBR VC scheduler and the two class schedulers together.

The type 1 CBR scheduler uses a non-work-conserving asynchronously shaped virtual clock (AShVC) algorithm to shape CBR virtual connections to any one of up to 32 user-programmable rates.

The type 1 rt-VBR scheduler uses a work-conserving starting potential fair queuing (SPFQ) algorithm to schedule virtual connections using any one of up to 16 user-programmable rates/weights.

The type 1 nrt-VBR, ABR, and UBR schedulers each use a work-conserving variable length frame based weighted round-robin (WRR) algorithm to schedule virtual connections using one of up to 256K different weights.

The guaranteed traffic class scheduler (GTS) uses the ShVC algorithm to provide guaranteed bandwidth to each of four classes (the rt-VBR, nrt-VBR, ABR, and UBR VC schedulers). A user-programmable rate is provided for each class.

The excess bandwidth class scheduler (EBS) uses a work-conserving self-clocked fair queuing (SCFQ) algorithm to fairly allocate excess bandwidth to each of four classes (the rt-VBR, nrt-VBR, ABR, and UBR VC schedulers). A user-programmable weight is provided for each class.

The port merge scheduler merges the CBR VC scheduler and the GTS class scheduler together with strict priority over the EBS class scheduler. CBR and GTS service is arbitrated at this level using timestamps.

Subports 16 to 31 are designated as type 2 and have dedicated two-level hierarchical schedulers. The first (VC) level of the hierarchy contains per-VC schedulers for each of the five traffic classes. Each VC scheduler can support all 64K VC queues as needed. The second (class) level of the hierarchy merges the VC schedulers together.

The type 2 CBR, rt-VBR, nrt-VBR, ABR, and UBR schedulers each use a work-conserving variable length frame based weighted round-robin (WRR) algorithm to schedule virtual connections using one of up to 256K different weights.

The type 2 class schedulers can be configured (for all 16 subports as a group) to use either an AShVC algorithm to shape each of the VC schedulers or a SCFQ algorithm to fairly allocate bandwidth to each of the classes. A user-programmable rate or weight is provided for each class. When the type 2 schedulers are configured for AShVC, they each provide five shaped pipes of virtual connections.

The top level (subport scheduler) of the hierarchy uses a non-work-conserving shaped virtual clock (ShVC) algorithm to shape each of the subport schedulers to its user-programmable rate. Two additional configuration options are provided to allow excess bandwidth to be granted to subport 31 (the microprocessor capture port) and a user-identified SAR port for work-conserving operation on those ports.

Following the subport scheduler is a virtual MPHY mapping function. The default configuration provides a one-to-one mapping between subport scheduler and MPHY address. This function can be configured by the user to map multiple subport schedulers to the same UTOPIA port. This allows multiple virtual MPHYs to be created within one physical MPHY device.

Description (continued)

Statistics/OA&M Controller (NMXSOX)

The NMXSOX implements ingress and egress OA&M functions as specified in the ITU and Bellcore recommendations.

Its main functions are OA&M cell type recognition and validation, and implementation of fault management defect indication, loopback, continuity check on all connections, and performance monitoring for up to 127 OA&M processes.

The NMXSOX also maintains per-connection as well as per-port and global cell event information (arriving/departing, enqueued/dequeued cells, and related timestamps) which can be communicated to an external adjunct device for connection statistics collection and performance monitoring data collection.

Enhanced Services Interface (ESI)

NMXESI implements a 16-bit parallel interface between the APC and an optional external adjunct device to support value-added enhanced services. The APC reports a rich set of events through this interface to support the implementation of virtually any customer desired per-connection, per-port, or per-device statistic in the external device.

ABR Flow Control Engine (XREAFE)

The ABR flow control engine (AFE) operates on ingress and egress ABR traffic to control switch and network congestion. The user has the option to globally select among different flow control algorithms implemented by the AFE. These are the selective EFCI marking algorithm and a Bell Labs' patented algorithm.

The ABR flow control algorithm can selectively operate on a minimum cell rate (MCR) plus equal share or proportional to MCR basis.

Only one ABR flow control algorithm can be active at a given time. The AFE also supports backward RM cell consolidation for up to 1023 fabric port multicast connections and 1023 egress subport multicast connections.

To support its operations, the AFE performs RM cell validation and calculates queue lengths and traffic load on subports (egress) or fabric ports (ingress).

Microprocessor Interface (MPI)

The MPI provides a generic 32-bit asynchronous microprocessor interface with maskable interrupts. This allows an external processor to access the APC for configuration, maintenance, statistics, internal and external memory reads and writes, as well as interrupt services.

The MPI also provides high-performance DMA support for fast cell insertion and extraction via FIFOs at the microprocessor interface.

Global Time Counter (TCT)

The TCT generates all the signals necessary for internal event synchronization. It also generates the external synchronization pulse GTSYNC.

Global Clock (WCK)

WCK generates all the clocks needed in the APC, as well as the output fabric clocks.

Reset Driver (RST)

RST synchronizes an externally applied, active-low, asynchronous reset pulse, and generates all the signals needed to reset the internal blocks of the APC. Reset is applied synchronously on each of the internal clock domains of the APC.

JTAG Test Access Port (JTG)

The APC incorporates logic to support a standard 5-pin test access port (TAP), compatible with the *IEEE* 1149.1 standard (JTAG), used for boundary scan. TAP contains instruction registers, data registers, and control logic. It is controlled externally by a JTAG bus master. The JTG gives the APC board-level test capability.

For additional information, contact your Microelectronics Group Account Manager or the following:

INTERNET: **http://www.lucent.com/micro**
E-MAIL: **docmaster@micro.lucent.com**
N. AMERICA: Microelectronics Group, Lucent Technologies Inc., 555 Union Boulevard, Room 30L-15P-BA, Allentown, PA 18103
1-800-372-2447, FAX 610-712-4106 (In CANADA: **1-800-553-2448**, FAX 610-712-4106)
ASIA PACIFIC: Microelectronics Group, Lucent Technologies Singapore Pte. Ltd., 77 Science Park Drive, #03-18 Cintech III, Singapore 118256
Tel. (65) 778 8833, FAX (65) 777 7495
CHINA: Microelectronics Group, Lucent Technologies (China) Co., Ltd., A-F2, 23/F, Zao Fong Universe Building, 1800 Zhong Shan Xi Road, Shanghai 200233 P. R. China **Tel. (86) 21 6440 0468**, **ext. 316**, FAX (86) 21 6440 0652
JAPAN: Microelectronics Group, Lucent Technologies Japan Ltd., 7-18, Higashi-Gotanda 2-chome, Shinagawa-ku, Tokyo 141, Japan
Tel. (81) 3 5421 1600, FAX (81) 3 5421 1700
EUROPE: Data Requests: MICROELECTRONICS GROUP DATALINE: **Tel. (44) 1189 324 299**, FAX (44) 1189 328 148
Technical Inquiries: GERMANY: **(49) 89 95086 0** (Munich), UNITED KINGDOM: **(44) 1344 865 900** (Ascot),
FRANCE: **(33) 1 40 83 68 00** (Paris), SWEDEN: **(46) 8 594 607 00** (Stockholm), FINLAND: **(358) 9 4354 2800** (Helsinki),
ITALY: **(39) 02 6608131** (Milan), SPAIN: **(34) 1 807 1441** (Madrid)

June 1999
PN99-088ATM

microelectronics group

Lucent Technologies
Bell Labs Innovations