Switch Concept & Architecture

What types of architecture makes sense?
How to reduce the effect of blocking?
Example architectures
Switching Principles

- Switching is the transport of information/data from an incoming logical channel to an outgoing logical channel.

- Logical channels are characterized by
  - physical inlet / outlet identified by a physical port number
  - logical channel on the physical port identified by
    - ATM: VPI / VPI
    - L2: DA / SA
    - L3: IP Address
    - Multi-layer: Flow, TCP port, etc.

- Switch operations
  - cells / packets arrive at the input port
  - header lookup, switching decision and translation is performed
  - packets are routed through or queued in the switch to the appropriate output port.
Switching Principles

Switches have three main part.

- **Transport Network:**
  - Physical means for switching cells from one input port to one (or more) output port(s) in the switch
  - Performs functions of the User Plane

- **Transport Control**
  - Controls the transport network based on analysis of signalling information
    - decides which inlet connects to which outlet via routing, switching, etc. ...

- **Network Control**
  - Sets transport control parameters / tables.
    - Routing table, resource mgmt., call admission control (CAC), etc.
  - Performs functions in the Control Plane
    - Routing, network mgmt., signalling, signalling AAL (SAAL), etc.
Switching Components

Network Control

Transport Control

Decision Table
Forwarding Decision / Classification

Descriptor Memory
Buffer Mgt. (Switch)

Media Access Control
Packet Controller
Packet Memory
Scheduler

Statistic Counters

Extension Port
CPU Port

Transport Network
Classical Switching Systems

Space Switch

Time Switch

$\lambda$ Switch
Switching Fabric Classification

Switch Fabric

Time Division
- Shared Memory: $O(n)$
- Shared Medium: $O(n)$

Space Division
- Single Path
- Multi-path
  - Crossbar: $O(n^2)$
  - Fully Interconnected: $O(n^2)$
  - Banyan: $O(n \log n)$
    - Batcher Banyan: $O(n \log^2 n)$
    - Clos: $O(2n*m + m*r^2)$
    - Recirculating
The Bus is used to connect the inlet / outlet ports together.  
» Cells are transported via the Bus.  
» Outlet ports accepts cells/packets based on their address

Non-blocking bus: \[ \Sigma [\text{Port speed}] < \text{Bus throughput rate} \]  
» Blocking bus effect can be reduced via statistical muxing effect.

Advantages:
» Multicasting is easily supported  
» Bandwidth limited to about 2 Gbps.  
» Easy integration with existing LAN equipment based on backplane technology, such as hubs.
The **Shared Memory** is used to connect the inlet / outlet ports together.

- Cells are temporarily buffered
- Outlet ports pulls cells/packets based on their queue

**Non-blocking memory bandwidth required:**

\[ \approx \sum \text{[Port speed]} < \text{Memory throughput rate} \]

**Advantages:**

- Memory usage reduced via sharing
- Multicasting is easily supported
- Bandwidth limited by memory technology.
Crossbar Switching System

- Internally non-blocking:
  - each I/O pair has an unique path.
  - externally blocking at the output.

- Advantage:
  - high speed internal paths
  - multicast ready
  - simple

- Disadvantage:
  - difficult to scale
  - input queueing
  - switching elements, $O(n^2)$
Multiplane Network
Internally & externally non-blocking: "m" multiple paths between I/O pair.

Each output port consists of a concentrator and a shared buffer.

Advantage: high performance, multicast-ready

Disadvantage: costly to implement, switch elements $O(n^2)$, difficult to scale
Delta Network Switching System
Batcher-Banyan Switching System

Sort 2  Merge 4  Merge 8  Shuffle Exchange

Batcher Network

Banyan Network
Batcher-Banyan Switching System

**Category:**

- Space switching
- Internally non-blocking Multiple Interconnection Network (MIN)
- No internal buffering required
- Output contention still a possibility, can be solved by using
  - additional arbitration logic
  - input buffering
  - output buffering with fabric speed up.
Batcher-Banyan Switching System

- **Batcher Network = sorting network**
  - Sorts all cells according to destination
  - Cells to same destination are adjacent.

- **Banyan Network = self-routing network**
  - No internal blocking, if cells are sorted at inlet
  - No external blocking, if only one cell for each outlet.

- **Output port contention solved using 3 phase algorithm**
  - Arbitration phase
  - Acknowledgement phase
  - Sending phase
Batcher-Banyan Switching System

- **Phase I: Send and Resolve Request**
  - Send source-destination pair through sorting network
  - Sort destination in non-decreasing order
  - Purge adjacent requests with same destination.
Phase II: Acknowledge winning port

» Send ACK with destination to the port with winning contention

» Route ACK through Batcher-Banyan network
Batcher-Banyan Switching System

- Phase III: Send Packet
  - Acknowledged ports send packet through Batcher Banyan network
  - Buffers at port controller
Augmented Banyan
Recirculation Network
Blocking Effects in Switches

- **Blocking** is a fundamental problem of switching
  - occurs when two or more cells contend for the same resources
    - paths through the switching fabric (internal blocking)
    - output ports (external blocking / output contention)

- **Queueing** is the solution to the blocking effect
  - cells are temporarily stored in buffer memory until it can be safely transported to the intended output port.
  - basic queueing techniques
    - input, output, internal, shared buffering.
Input Buffering

- All arriving cells are queued at the input
  - cells are queued until the switch indicates that the cell has been successfully switched.
  - cells that fail to reach the output port stay in the buffer and can be tried again on the next round of arbitration.

- Advantage: Simple to implement.
- Disadvantage: Head of Line (HOL) blocking.
Output Buffering

- **Assumption:** All arriving cells must be switched to the output for buffering
  - When two or more cells are switched to the same output port in a cycle, extra cells are buffered until it can be transmitted.
  - Switching transfer speed must be performed at $N$ times the inlet speed. If transfer speed $< N \times$ inlet speed, then internal cell loss will occur.

- **Advantage:** Avoids HOL blocking, no arbitration required
- **Disadvantage:** Switch fabric may be difficult to implement
Internal Buffering

- Buffering of cells is done within the internal fabric of the switch.

- Advantage: Hides the concept of buffering

- Disadvantage: Difficult to implement efficiently.

- This technique is hardly used.
**Shared Buffering**

- **Assumption:** All arriving cells are switched to the shared output buffer.

- **When two or more cells are switched to the same output port in a cycle, extra cells are buffered until it can be transmitted.**

- **Advantage:** Memory requirement is greatly reduced, no HOL blocking

- **Disadvantage:** Limited memory bandwidth.
Other Queueing Strategies

- Multiple priorities may require more complicated queueing strategies.

- Partial buffer sharing
  - Buffer is "divided" for high and low priority cells.
  - High priority cells may only occupy high priority buffers.

- Push out buffer
  - Buffer is shared
  - High priority cells may overwrite low priority cells.
Queueing: Implementation Parameters

- Implementation complexity influenced by
  - Queue Size
    - Performance requirements
    - Queueing discipline
  - Memory Speed
    - Queueing discipline
    - Link speeds
    - Number of ports
    - Memory width
  - Memory Control
    - Queueing discipline
## Queueing: Comparison

<table>
<thead>
<tr>
<th></th>
<th>Input Queueing</th>
<th>Output Queueing</th>
<th>Central/Shared Queueing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single Port Memory</strong></td>
<td>W/2F</td>
<td>W/(N+1)F</td>
<td>W/2NF</td>
</tr>
<tr>
<td><strong>Example (ns)</strong></td>
<td>53.3</td>
<td>6.3</td>
<td>3.8</td>
</tr>
<tr>
<td><strong>Dual Port Memory</strong></td>
<td>W/F</td>
<td>W/NF</td>
<td>W/NF</td>
</tr>
<tr>
<td><strong>Example (ns)</strong></td>
<td>106.6</td>
<td>6.7</td>
<td>6.7</td>
</tr>
<tr>
<td><strong>Memory Speed</strong></td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td><strong>Control Logic</strong></td>
<td>FIFO</td>
<td>FIFO</td>
<td>Complex</td>
</tr>
<tr>
<td><strong>Memory size</strong></td>
<td>Very high</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td><strong>Multicast</strong></td>
<td>Difficult</td>
<td>Easy</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Cell size = 53 bytes  
W = 16 bits  
F = 150 Mbps  
N = 16
Ethernet Switch v1.x/v2.x Architecture

Switch IC V1.0
(9 x 9 Fabric)

Port IC V1.0
(100Mbps)

Port IC V2.1
(100Mbps)

Port IC V2.1
(100Mbps)

Port IC V1.1
(10Mbps)

Memory

Memory

CPU

Memory

Memory

Memory

Memory
Ethernet Switch v3.0 Architecture

ESIC V3.0

Serial EPROM

16 bits

64/32 bits

32K X8

10/100 MAC/MII

MAC10

256K X16

1 - 8 Mbytes EDO RAM
• Non-Blocking for 3 Chips
• Maximum Connections for 8 Chips
Galileo System Architecture
Galileo Buffer Management

- **Read Pointer**
- **Write Pointer**
- **M/U Byte Count Blk Addr**
- **Frame #n**
- **Rx Empty List**
- **Tx Descriptors: 1024 X 9**
- **Rx Buffer (for all ports and PCI)**
- **GT-48001A**
- **DRAM**
Galileo Memory Structure
I-Cube Switch Architecture
I-Cube LS101
I-Cube LS101 Switch Matrix
MAC Layer Controller

[Diagram of MAC Layer Controller]

- Physical Interface
  - TXD[3:0]
  - TX_EN
  - TX_CLK
  - COL
  - CRS
  - RX_DV
  - RX_EN
  - RX_CLK
  - RXD[3:0]

- Transmit Data Serializer
- Transmit Protocol Control
- Transmit MUX
- Transmit Data Synchronizer
- Preamble/Sync Jam Pattern Generator
- Transmit Packet Type Detector

- Receive Protocol Control
- Receive Data De-Serializer
- Receive Data Synchronizer
- Receive Packet Type Detector

- Flow Control
- Data Frame Decoder

- Transmit Data from Buffer Manager
- Transmit Data to Buffer Manager and ATC
Memory & Buffer Mgt.
Queue Dataflow

[Diagram showing the dataflow.]
PMC Sierra System Architecture
PMC Sierra Chip Architecture
PMC Sierra Chip Architecture
PMC Sierra L3 System Architecture

Stacking SERDES Transceiver

PM3370 (Layer 3 OAR)

PM3370

EXACT RING

PM3370

PM3370

PM3370

PM3370

Stacking SERDES Transceiver

PM3390

PM3370

PM3370

PM3370

PM3380 (Master)

Gigabit PHY

1000M Ethernet

8 x 10/100 Ethernet link

10/100 Quad PHY

10/100 Quad PHY

10/100 Quad PHY

10/100 Quad PHY

External HOST CPU

loop back

8 x 10/100 Ethernet link

8 x 10/100 Ethernet link
PMC Sierra PM3390

CLK135A
CLK135B
CLK135C
CLK135D
CLK135E
CLK135F
CLK125A
CLK125B
CLK125C
CLK125D
CLK125E
CLK125F
CLK25A
CLK25B
CLK25C
CLK25D
CLK25E
CLK25F
SYSCLKA
SYSCLKB
SYSCLKC
SYSCLKD
SYSCLKE
SYSCLKF

PM3380 GB uplink-1
PM3380 GB uplink-2
HDMP1636
HDMP1636
HDMP1636
PM3390
EX-1
EX-2
EX-3
EX-4
EX-5
EX-6
EX-7
EX-8

HCMP636 expansion-1
HCMP636 expansion-2
HCMP636 expansion-3

CLK135[?]
@135MHz
SYSCLK[?]
@83MHz
ENSYSCLK
@CLK135/2
CLK25[?]
@25MHz
CLK125[?]
@125MHz
EXACT BUS
@1Gbps
(135MHz)

PM3370 Ports 1-8
PM3370 Ports 9-16

BCM 5208
BCM 5208
BCM 5208
BCM 5208
BCM 5208
BCM 5208

CCL/N300; Paul Huang 1999/4/13
PM3370 Octal Fast ES Port Controller

- 8 channels 10/100 MII interface
- ENTRE Packet interface
- Link transmit controller
- Egress queue manager
- EXACT Bus Interface
- Link receive controller
- Ingress DMA
- Ingress queue manager
- Global registers
- Embedded SmartPath RISC CPU
- Address resolution logic
- JTAG interface
- Microprocessor Interface
- Memory Interface A
- Memory Interface B
- JTAG
- External CPU
- SSRAM A
- SDRAM B
- EXACT Ring Devices
- 8 x 10/100 PHY
- OR
- ENTRE external device
Address Table: Hash Bucket

Hash bucket pointer array (max. 32768)

Hash Bucket

Hash Bucket

Hash Bucket (max. 65536)

NextBktPtr

NextBktPtr

NextBktPtr

MACadr[15:0]

NextBktPtr[15:0]

MACadr[47:16]

timestamp[15:0]

MACstasBlk[15:0]

Forwarding context word 0

(1 to 5 forwarding context words)

Forwarding context word N-1

2 x [15:0]
Output Queue Structures

Queue Control Frames (on-chip, one-per-queue)
PM3380 GE Switch Port Controller

- **Gigabit Ethernet MII Interface**
- **Link transmit controller**
- **Link receive controller**
- **Egress queue manager**
- **Ingress queue manager**
- **EXACT Bus Interface**
- **EXACT Bus Ring Devices**

- **Global registers**
- **Embedded SmartPath RISC CPU**
- **Address resolution logic**
- **Ingress DMA**
- **data to PPB**
- **queue management**
- **SSRAM A**
- **SSRAM B**
- **JTAG interface**
- **Microprocessor Interface**
- **Memory Interface A**
- **Memory Interface B**
- **JTAG**
- **External CPU**
- **Free list access**
- **translated**
Output Associated Input Queueing

- Packets segmented into 240-byte blocks of Partial Packet Buffer (PPB).

- Input packets are classified by destination and priority (256-port x 4-class).

- IQM issues Queue Allocated (QA) to destination if queue is not empty.

1. Packets segmented into 240-byte blocks of Partial Packet Buffer (PPB).
2. Input packets are classified by destination and priority (256-port x 4-class).
3. IQM issues Queue Allocated (QA) if queue is not empty.

(1-a): DA,SA extracted to ARL
(1-b): packet segmented into PPB
(2): DA/SA address lookup
(3): routing result to IQM
(4): Classify packets into output queues
(5): Issue QA to EQM if queue not empty
Multicast is achieved by duplicating Queue Elements (QE).

PPB is released after transmission if REF_COUNT = 1.
“PULL-ed” by Output

- When output FIFO becomes available for pending requests, EQM issues Queue Fetch (QF).
- When received QF, IQM responds with Data Block (DB).
- Packets are “pulled” from input by output.
- IQM frees PPBs to IDMA after packets are delivered (REF_COUNT=1) or discarded.
- Data Flow:
  QA - QF - (DB - QF)^n - DB_{last}

(5): Issue QF when output available
(6): Queue Fetch received
(7): find QE of requested packet
(8): send packet segments as DB
(9): reassemble DBs and issue next QF
(10): Free QE and PPBs
Vertex L3 Switch Architecture

- **SC220** XpressFlow Engine
  - Address Mapping Table
  - Buffer RAM

- **EA208E** 8-Port Ethernet Access Controller
  - 8 Ethernet ports

- **EA208E** 8-Port Ethernet Access Controller
  - 8 Ethernet ports

- **EA234** 4-Port Ethernet Access Controller
  - Four 100M Fast Ethernet ports

- Flash ROM
- Switch Manager CPU
- DRAM
- RS232 Local Control Console
- Management Bus
- XpressFlow Bus

CCL/N300; Paul Huang 1999/4/13
Vertex XpressFlow Engine

SC220 XpressFlow Engine

CAM (optional)
- Address Mapping Table
- SRAM
- Control Buffer Memory
- Control Buffer Memory I/F
- CAM I/F
- HISC I/O Registers
- XpressFlow Bus I/F
- Auto Buffer Mgt.
- Mgt. Bus I/F
- HISC Core
- Management Bus

XpressFlow Bus
SMAS Chip Architecture

Signal Pin Count: 535
SMAS Switch

OC-12

1st Stage

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

3rd Stage

OC-12

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch

2.5 Gb/s ATM Switch
### SMAS Timing Diagram

<table>
<thead>
<tr>
<th></th>
<th>FP_B</th>
<th>FP_Q</th>
<th>TP</th>
<th>HP</th>
<th>Reg</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>HP_IBM ← FP_B(HP_IBM)</td>
<td>HP_IFP ← FP_Q(HP_IFP)</td>
<td>R1 ← TP_AQkp</td>
<td>EF_AQkp ← 0</td>
<td>R2 ← HP_IBM</td>
</tr>
<tr>
<td>S2</td>
<td>FP_Q(R3) ←&lt;R1,R2&gt;</td>
<td>TP_AQkp ← R1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>&lt;R4,R_MPp'&gt; ← FP_B(HP_MPpQp)</td>
<td>FP_Q (HP_IFP) ← FP_Q(HP_IFP)</td>
<td>R1 ← TP_AQkp</td>
<td>EF_AQkp ← 0</td>
<td>R1 ← HP_IFP</td>
</tr>
<tr>
<td>S4</td>
<td>FP_Q (R3) ←&lt;R1,R2&gt;</td>
<td>TP_AQkp ← R1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>&lt;R2, R1&gt; ← FP_Q(R5)</td>
<td>R3 ← TP_AQkp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>FP_B(TP_IBM) ← R1</td>
<td>FP_Q(TP_IFP) ← R5</td>
<td>HP_AQjp ← R2</td>
<td>TP_IBM ← R1</td>
<td>TP_IFP ← R5</td>
</tr>
<tr>
<td>S7</td>
<td>Same as above.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S8</td>
<td>Same as above.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td>HP_IBM ← FP_B(HP_IBM)</td>
<td>R2 ← HP_IBM</td>
<td>R7 ← MP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S10</td>
<td>FP_B(TP_MPpQp) ← &lt;R2,R7&gt;</td>
<td>TP_MPpQp ← R2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:**
1. Total registers: R1, R2, R3, R4, R5, R6, R7, R_MP
2. Set R_MP: "0000 (F)" when initializing the system
3. *: executed when R_MPp = "0000(F)", meaning just initialized or the last multicast cell was served completely
4. **: executed when completing service of the current MAQ cell (i.e., R_MPp = only one bit is '1')
5. ***: read out all EF_AQj (4 bits), read/write simultaneously (dual prot mem.)
Fujitsu Chipset

SRE Self-Routing Switch Element
NTC Network Termination Controller
ALC Adaptation Layer Controller
ATC Address Translation Controller
- **ATM Switch Element SRE (Self-Routing switch Element)**
  - 4x4 155 Mbps cell switch building block
  - Selectable high and low priority queues
  - Output queued for nonblocking operation
  - Multicast support
SRE Switch Matrix

Switch Configuration Parameters

Input Ports

Output Port

*Delta Configuration allowed
Network Termination Controller (NTC)

- On-chip DMA controller for high-speed transfer of statistics to system memory.
- Interface to Address controller to perform real time cell header translation.
- Cell rate decoupling
Address Translation Controller (ATC)

- Real time translation of ATM header information up to 155 Mbps
- Supply 3 byte switch-internal routing tag.
- 1024 entry CAM
- Full 28 bit comparison for each entry, with optional bit-masking
- Supports multiple matches for multicast operation.
- Multiple ATCs can be cascaded to support larger addressing range
- Supports CLP and congestion indication/removal for each entry.
Adaptation Layer Controller

Host CPU  \rightarrow  SAR Memory  \rightarrow  ALC  \rightarrow  NTC  \rightarrow  XVR

Host CPU  \rightarrow  System Memory

ALC  \rightarrow  NTC  \rightarrow  NTC  \rightarrow  NTC

ATM Switch

ATM TE
Scorpio Chipset

STS-3

2xSTS-3

3xDS3

6x15.6 Mbps

* each JC needs a SRAM
* each OC needs a DRAM
Scorpio Chipset

- **ATM Junction Controller (JC)**
  » Provides x-point switching and buffer mgt. for 4 junctions of a switch fabric.

- **ATM Output Controller (OC)**
  » Provides cell buffer mgt. and scheduling for up to 16 junctions.

- **ATM Dual Port Interface Controller (Dual PIC)**
  » Interfaces a 640 Mbps fabric channel to 2 UTOPIA interfaces.

- **ATM Port Interface Controller with SAR (PIC w/ SAR)**
  » Interfaces a single 640 to a single UTOPIA interface and includes SAR functionality to allow fabric access by a host processor.

- **ATM Multiple Physical Support Chip (MultiPHY)**
  » Multiplexes up to 6 UTOPIA interfaces, with an aggregate bandwidth of less than or equal to 155 Mbps to a single UTOPIA interface.
Scorpio Chipset – Cont'd

- **Switch Features**
  - NxN switch fabric (640 Mbps/channel; N < 8; < 5.12 Gbps)
  - Non-blocking architecture
  - Distributed output cell buffering (combination of central and output buffering)
    - Buffer sharing to allow for statistical variance in port buffer usage
    - Average buffer sizes up to 64k cells per output channel
    - Guaranteed minimum buffer space per physical port (between 2 to 64 cells).
    - Configurable per-port limits on buffer usage to maintain fairness.
    - Configurable per-port thresholds at which cells with CLP=1 are discarded.
  - Full support for multicast
Scorpio Chipset - Cont'd

- **Port Interface Module**
  - Support for up to 14 ports (max. aggregate capacity of 640 Mbps)
  - ATM Layer mgt. - VP/VC label swapping
  - SAR function integration
  - Back pressure indication for flow control between physical ports and the SF.
  - Multiplexes input port data onto a single Fabric channel
  - Demultiplexes Fabric channel data to physical ports.
Scorpio Chipset – Cont’d

- **Junction Controller**
  - Total capacity of 1.28 Gbps
  - Ability to manage 124 individual queues (112 unicast; 12 multicast)
  - Supports 2 x 14 physical port per output bus.
  - Dynamic allocation of 1k to 32k of ext. cell memory among all managed queues

- **Output Controller**
  - Coordinates queue management for up to 4 JC.
  - Provides multicast functionality
  - Provides cell transmission scheduling for all queues based on a combination of weighted round robin and strict priority algorithms.
Port Interface Controller
- Interfaces a 640 Mbps fabric channel to two (for dual) UTOPIA devices
- Provides VPI/VCI mapping
- Provides cell drop accounting
- SAR functionality (for PIC w/ SAR)

MultiPHY
- Multiplexes/demultiplexes up to six UTOPIA devices
- Provides back pressure mechanism to the switch fabric by monitoring FIFOs within the physical layer devices
Transswitch - CellBus Switch
Transwitch Chipset

**CUBIT Device: Cellbus Switch TXC-05801**

- Inlet-side address translation and routing header insertion (ext. SRAM)
- Cellbus access request, grant reception and bus transmission
- Cellbus cell reception and address recognition
- Outlet cell queueing: various modes
- Master bus arbiter included in each CUBIT
- Interface port to translation table SRAM

**CUBIT**

- 37 line common bus @ 38 Mhz clock rate = 1 Gbps bandwidth.
- max 256 multicast sessions.
- 4 different service queues at output.
Brooktree Chipset

Input Dual-Port RAM (512x36)

Read Pointer

Flags

Write Pointer

Bidirection SW Control

Write Pointer

Flags

Read Pointer

Output Dual-Port RAM (512x36)
BrookTree Chipset - Cont'd

- **Bt8215 Bidirectional Cell Buffer**
  - Simplifies interface between the processors and the peripherals by integrating memory and control logic.
  - Replaces eight separate FIFO memories and associated control logic for the byte-to-word format conversion; 8 bit to 32 bit data alignment.
  - Cascade with off-the-shelf FIFOs for greater depth.
  - Supports fixed-length cell switching.
  - 33 Mhz operation for 36 bit port; 20 Mhz operation of 9 bit port.

- **Additional Features**
  - 2 Bt8215s may be cascaded to form a 64-bit interface.
  - Full, empty, almost-empty, almost-full, and half-full flags provide for buffer control.
  - Bidirectional 36-bit port with integral parity check.
  - Separate unidirectional 9 bit ports with integral parity check.
  - 512 x 36 bit buffer memory in each direction.
  - Synchronous or asynchronous interfaces on all ports.
AT&T System Architecture (1)
AT&T System Architecture (2)
Switching capacity ~ 6x6 622 Mbps (3.7 Gbps total bandwidth)

Integrated internal buffer (shared) ~ 512 cells (No external buffering necessary)

Scalable
- 18x18 622 Mbps (11 Gbps total bandwidth) on a single fabric card
- 36x36 622 Mbps (24 Gbps total bandwidth) on a dual fabric card
- Support up to 1080x1080 UNI / NNI ports
- Redundant fabric possible

Functionality
- 4 WRR (16 programmable weights) delay priorities/port, each with 2 loss priorities and configurable dynamic thresholding
- Smart backpressure specific to congested fabric port and priority level
- HOL blocking avoided via flow control
- Multicasting (only one cell copy)
Scalable buffer size up to 32k cells/fabric port

4 separately configurable WRR (16 programmable weights) delay priorities/port, each with 2 loss priorities and configurable dynamic thresholding

  » ensures higher priority sub-queues served more frequently
  » prevents starvation of lower priority levels (i.e. guaranteed some BW sharing)
  » If empty or blocked, serves highest, non-empty, non-backpressured delay priority

HOL blocking avoided via flow control

"Smart" backpressure from fabric to ingress and (w/ override) from egress to fabric

Rate scheduling for all subports on output (1.5 to 622 Mbps)

Provides multicasting to all subports
ATLANTA Chipset - ALM

- **ATM Layer functions**
  - Up to 30 physical UNI/NNI supported via MultiPhy Utopia II interface
  - Up to 32k VCs (in/out-bound) per fabric port
  - VPI/VCI translations, optional HEC check & header correction.

- **Traffic Management**
  - per-VC configurable dual-leaky bucket UPC
    (scalable granularities: 64 kbps to 622 Mbps, with ~ 0.1% steps)
  - Cell monitoring: CLPO, CLP1, CLPO+1
  - Parameter monitoring: PCR, CVT, SCR, BT.
  - Extract/Insert OAM cells with optional routing to/from local µP interface.

- **per-VC statistics collection (6 counters)**
ATLANTA - Memory requirements

- **VC Tables:** ~ 2200 VCs per 1 Mbit SRAM
  - VC connection parameters, Traffic statistics, Policing parameter
  - w/o policing ~ 4000 VCs per 1 Mbit SRAM
  - w/o policing & statistics ~ 8000 VCs per 1 Mbit SRAM
- **Cell Buffer:** ~ 2k cells per 1 Mbit SRAM
  - Minimum 4k cells needed w/ 2 units of 32k x 32 SRAM
  - Does not include associated pointer space
- **Example:** 4 STS--3c ports on a line card
  - ~ 8k VCs (no policing): 2 Mbits
  - 4k cell buffer: 2 Mbits
  - Pointer space: 1 Mbits