Broadcom BCM5600
StrataSwitch

A Highly Integrated Ethernet Switch
On A Chip

Andrew Essen and James Mannos
Broadcom Corporation
Outline

- Introduction
- Networking Basics
- Description of BCM5600
- Design Process
- Vital Statistics
- Conclusion
Background

• A single modern processor can now saturate a LAN which used to be easily shared by multiple users.

• Multimedia applications, especially those with real-time requirements, place a huge strain on bandwidth
New Technology to the Rescue

- **New Network Architectures** - replace shared media and hubs with dedicated media and switches

- **Provision for Increased Bandwidth** - replace 10Mb/s with 100Mb/s (FE) and 1000Mb/s (Gig)

- **New Protocols** - L4 to L7 filtering, Class of Service (802.1p), Virtual LANS (802.1Q)

- **Advancing Semiconductor Technology** - DSM (< .25μ), System-on-chip architectures, and large amounts of integrated memory.
# Project Goals

<table>
<thead>
<tr>
<th>Features</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switch On Chip</td>
<td>- Lowest Cost/Port</td>
</tr>
<tr>
<td></td>
<td>- Faster System Development</td>
</tr>
<tr>
<td></td>
<td>- Modules can be easily “bolted” on</td>
</tr>
<tr>
<td>24 10/100 &amp; 2 Gig Ports Non-blocking</td>
<td>- High Capacity</td>
</tr>
<tr>
<td></td>
<td>- High Performance</td>
</tr>
<tr>
<td>Non-blocking L2/L3 Switch</td>
<td>Line speed switching and routing</td>
</tr>
<tr>
<td>Line speed L4-L7 filtering</td>
<td>- Fast Filter Processor &amp; Flexible Rules Engine</td>
</tr>
<tr>
<td></td>
<td>- Content Aware traffic classification based on any combination of designated fields.</td>
</tr>
<tr>
<td>Advanced Features</td>
<td>- 4-levels of COS (Class of Service).</td>
</tr>
<tr>
<td></td>
<td>- Support of Virtual LAN Trunking, Flow Control, Mirroring, Tagging &amp; Spanning Tree</td>
</tr>
<tr>
<td></td>
<td>- Elimination of Head of Line Blocking</td>
</tr>
<tr>
<td></td>
<td>- Broadcast/Multicast</td>
</tr>
<tr>
<td>On board SRAM caches</td>
<td>- Allows use of low-cost, external SDRAM</td>
</tr>
<tr>
<td>PCI interface</td>
<td>- Flexible connectivity and expandability.</td>
</tr>
</tbody>
</table>
What Makes a SOC so Tough?

- Network traffic is chaotic and asynchronous
- Can have large differential flows of data
- Behavior often appears non-deterministic because of packet drop
- Differing packet sizes & formats along with a huge amount of state - very difficult to verify
- 60M transistors, 1 MB SRAM, 4M+ gates.
Traditional Network

(Before Switched LAN’s)

Each L2 broadcast domain is connected by a router
Layer 2 Switching

- Appears to end stations as though everybody is still on a single shared media
- Lookup destination MAC address and forward packet to correct port
- Numerous packets can be “on the wire” simultaneously
  - Up to 52 with the BCM5600
    - 26 full duplex ports
  - Versus 1 with a traditional shared Ethernet
Layer 3 Switching

- End stations must be kept aware of routing configuration

- L3 Switching is performed when the L2 destination is the router itself
  - Lookup IP destination in routing table
  - Update L2 addresses
  - Decrement TTL
  - Forward packet to correct port
Layer 4-7 Processing

• Pattern matches result in actions
  – Kind of like a simple awk script
  – Actions include drop, change priority, change destination, send to CPU

• Applications include:
  – Detect traffic type based on TCP port and reprioritize
  – Support VoIP and multimedia applications such as real-time video
  – Simple firewall by examining IP addresses
**StrataSwitch 24+2 Switch**

<table>
<thead>
<tr>
<th>RJ45 x 24</th>
<th>Fiber Xor or RJ45+Magnetics</th>
<th>Fiber Xor or RJ45+Magnetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEDs</td>
<td>LEDS</td>
<td>LEDS</td>
</tr>
<tr>
<td>Octal PHY</td>
<td>Octal PHY</td>
<td>Octal PHY</td>
</tr>
</tbody>
</table>

**Additional Silicon Required**
- 3-5218
- 2-5400
- Integrated Microcontroller
- 4-64 Mbytes SDRAM

**Optional I2C Components**

**System Bus**

- PCI Bridge
- SDRAM
- Flash
- Serial Port
- CPU

**PCB Components**

- EPICs
- GPICs
- PMMU
- CMIC
- ORION SA

**Network Interfaces**

- RJ45
- Fiber Optic

**Clocks**

- 128 bits/125 MHz
- I2C Bus 100/400 KHz (very low bandwidth path)

**Memory Buses**

- 32-bit 33/66 MHz
- 64-bit 133 MHz

**Bus Interfaces**

- PCI
- System Bus

**Microarchitecture**

- ORION SA
- CMIC
- EPICs
- GPICs
- PMMU

**Silicon Components**

- ORION
- SDRAM
- Flash
- Serial Port
- CPU
All packets travel as cells along a very wide central bus
Ingress Details

MAC → Data Buffers → L2 Logic → L2 Tables → L3 Logic → L3 Tables

- 8k entries
- L2/L3 determines destination port(s)
- 2k entries
- 128 entries
- L4-L7 processing
- Priority assignment
- Cell slicing
- Drop decision

Internal Bus
Memory Allocation & Packet Storage Control

Packet Queues

SRAM Packet Buffer

SDRAM Interface

Packet Read Control & Memory Dealloc

Internal Bus

27 concurrent packet writes, backpressure status, write buffer mgt

4 - 64 Mbytes of SDRAM (2 Gbytes/s peak bandwidth)

27 concurrent packet schedules, DRAM to SRAM prefetches, SRAM to Egress Reads
Egress Details

Data Buffers

Packet reassembly & transmit control

MAC

MMU requests

Internal Bus
PCI Details

• 32 bit/66 MHz operation

• Also behaves as port 27
  – Contains mini-ingress & egress
  – Allows easy processing of unusual packets
    • E.g., routing non-IP packets
Simulation

• “Cycle accurate” C++ model

• Validate architecture
  – First implementation of initial concepts
  – Try out various algorithms
  – Find bottlenecks

• Optimize queue and buffer sizes
  – With regards to finite real estate
Simulation Results

24+2 Streaming to 24+2

Wire speed operation with various packet sizes
More Results

2 Gig Spraying to 20 Fast

Wire speed

24+2 Spraying to 24+2

Fast ports are overloaded
Simulated Packet Flow

Timeline of individual packets in simulation

TOASTView 1.11

Legend: invalid ex xma eng los noc unc txc txcu txcu losma smt class sch tx drop
Verification

- Verification is complicated by...
  - Multithreaded, asynchronous nature
  - Long operations
    - 160k cycles for 1500B packet at 10 mbits
  - Long time to steady state
    - Filling 64 Mbytes takes a while at 100 Hz
  - Dropped or redirected packets may be correct

- Multiple solutions
  - Extensive unit testing
  - Full chip testing with automated checking
  - Emulation
Checker

Ingress

Ingress

Ingress

MMU

Egress

Egress

Egress

Verilog/C

Checker

RTL/gates

PLI

C++

Real-time pass/fail
• 50 KHz is a lot better than 50 Hz!
  – Allows software development
  – Millions of packets per day
  – Test environment is similar to that for silicon debug
Emulation Setup

SmartBits unit, Host PC, and speedbridges
The Emulator

One of three cabinets
System Bringup

SmartBits and board
Reference Board
Design Timeline

6/98 - 7/99

6/98 Architecture Spec

10/98 RTL Integration

1/99 RTL Feature Complete

8/98 Simulator Complete

5/99 Quickturn Stable

2/99 Begin Quickturn

6/99 Tape-out

7/99 Switching Packets!
Vital Statistics

- 24+2 ports
- 60 million transistors
- 1MB of embedded SRAM
- 133 MHz
- 0.25u 5 metal CMOS
- 2.5V core, 3.3V I/O
- 600 ball TBGA package

(Die Photo)
Summary

• First integration of 24FE + 2GE, line-speed, L2-L7 switch on a single chip

• Enables convergence of voice, video, and data to the desktop.

• First pass silicon in less than 12 months, read PCI ID in minutes, switching packets the next day!

• Good flow from solid spec through emulation was crucial to the success.