A Single Chip Terabit Switch

Speaker: Fred Heaton

Authors: Bill Dally, Wayne Dettloff, John Eyles, Trey Greer, John Poulton, Teva Stone, Steve Tell
Architecture Overview

- Re-usable, Low Power Tx/Rx Module
- Rx Clock/Data recovery
- Adaptive termination
- Lane-by-lane selectable data rates
  - 3 selectable Reference clock inputs
  - 5 selectable multiplier settings
- Asynchronous switch operation
- User selectable Transmitter Pre-emphasis
- Scalable architecture
Applications

- 3/5 Stage Clos Network configurations
  - 9800 port SNB 3 stage Clos network (30+ Tbps)
- Protection switching
- Video routing
- DWDM OEO switching applications
Receiver

140 x 140
3.125Gb/s
Switch Core

Receiver assembly N
receiver assembly 2
receiver assembly 1
receiver assembly 0

Rx
RxDat
RxToggle

CDR
Clock Gen

To Switch

Host Interface, Controller, and Global logic

8 bit uP Interface
MDIO/MDC

RX0 ±

RFCC ±
RFCB ±
RFCA ±
Switch

From Rx

- 140 x 140
- 3.125Gb/s
- Switch Core

Crosspoint switch fabric

140 x 140 x 9

To Tx

- Full reconfiguration and multicast support
- Asynchronous operation
- Differential data path design
- Byte wide datapath + 1 “clock”
Transmitter

- Differential, CML signaling
- Transmitter Pre-emphasis
- On-chip adaptive termination

Switch control

Host Interface Controller, and Global logic

8 bit uP Interface MDIO/MDC

© 2000 Velio Communications, Inc.
Technology

- 0.18 micron, 7 Layer Metal
- Flip-chip die attach
- Ceramic BGA package
  - 1mm ball pitch
  - 36 x 36 full array (1296 pins)
  - 168 Vdd pins (1.8 volt nominal)
  - 13 Ivdd pins (2.5/3.3 volt nominal)
  - 473 Ground pins
Electrical/Performance Specs

- Rx Input Sensitivity - 35 mV
- Jitter Tolerance - 0.65 UI
- Jitter Generation - 0.25 UI
- Plesiochronous tracking of 200ppm difference between reference and embedded data clock
- Switch Reconfiguration time ~30uS
- Clock Lock time ~9uS
- Selectable output levels (675 mV max)
- Power Dissipation
  18 Watts @ 2.5 Gbps (1.8 V operation)
  22 Watts @ 3.125 Gbps (1.8V operation), 130mW/Tx-Rx pair
Early Late Logic

\[ \text{dclkN} \]
\[ \text{eclkP} \]
\[ \text{dclkP} \]
\[ \text{eclkN} \]

Logic threshold

- Data sample

\( ! = \text{late} \)
\( = \text{early} \)
Switch Design

- Asynchronous, low-swing differential design
- Pre-emphasis technique used to drive heavily loaded row/column lines
- Weak driver and “resistors” used to maintain differential voltage at DC
- 3 Watts worst case power dissipation at 3.125 Gbps, 2 volt operation
Nominal Spice simulation, 2.5 Gbps operation
Transmitter

2^{10} - 1 PRBS Generator

Driver

Programmable Termination

720° Phase Interpolator

Clock Multiplier

EQL

8/10

dclkP

dclkP

eclkP, eclkN

dclkP, dclkN

TXDAT

tx CDR

from switch

TXTOGGLE

Quadradure clocks

DACLK+ , DACLK-

PRBS Generator

TX+, TX-
Transmitter  Current Steering Element

- 4 identical sets of current steering elements
- Separate staggered clocks provided to each for slew rate control
- Programmable bias generator (5 settings)
- Equalization tap (5 settings)
Interconnect Performance @ Gigabit rates

- Skin effect attenuation - $\sqrt{\text{freq}}$ relationship
- Dielectric loss - to first order, linear attenuation with frequency
- Serial transmitters use pre-emphasis to compensate for signal distortion effects of FR4
Example Waveform (No Pre-emphasis)

Transmitted pulse

Received pulse
Example Waveform (Pre-emphasis)

- Pre-Emphasis 1-tap equation:
  \[ V_{pr}(n) = a \times V_i(n) - b \times V_i(n-1) \]
  Narrows pulse which opens the transmit eye at the receiver
Pre-Emphasis at 2.5Gb/s, PRBS data:

At Transmitter: No Pre-Emphasis

At Transmitter: Pre-Emphasis Enabled

Signal at Receiver

Signal at Receiver
On-chip Processor

- 8 bit Microprocessor
  - 4K x 14 bit instruction ROM
  - 4K x 14 bit instruction RAM
  - 512 x 8 bit data RAM
  - External EEPROM interface
- On reset processor performs
  - On-chip register initialization
  - Default switch configuration
  - Receiver offset trim
  - Termination resistor calibration
- Ongoing polling loop
  - Register updates, termination resistor updates
$2^{10} - 1$ PRBS, 3.577 Gbps, 4” FR4

- Duty cycle distortion present due to internal clock load imbalance
- Eye reduction in both the time domain and amplitude
Terabit Switch Building Blocks

- Crosspoint switch demonstrates two key building blocks
  - High-speed I/Os
  - High-performance switch core
- Different switches can be realized by adding application-specific logic between these
  - Grooming switch
    - Framers and time-slot interchangers
  - Cell switch
    - Scheduler and queueing
Other Products

- **VC2002 SONET/SDH Grooming Switch**
  - 72 Integrated 2.5Gb/s I/O port pairs
  - SONET Input and Output Processing
  - ST-192(c) Support
  - STS-1 level switching
    - (3456 x 3456 STS-1 Switch)
  - multi-stage scalable architecture

- **VC1001/2/3 Octal SERDES**
  - full-duplex, 1Gbps - 3.125 Gbps, 1.5 - 2.1 Watts
  - Quad version also available