A Scalable Load Balanced Birkhoff-von Neumann Symmetric TDM Switch IC for High-Speed Networking Applications

Ching-Te Chiu*, Yu-Hao Hsu, Min-Sheng Kao, Hou-Cheng Tzeng, Ming-Chang Du, Ping-Ling Yang,
Ming-Hao Lu, Fanta Chen, Hung-Yu Lin, Jen-Ming Wu, Shuo-Hung Hsu, Yar-Sun Hsu
*Institute of Communications Engineering, National Tsing Hua University, Hsinchu, 300, Taiwan.
E-mail:ctchiu@cs.nthu.edu.tw

***Abstract*—For the first time, we implemented a scalable load balanced Birkhoff-von Neumann TDM switch IC with SERDES interface circuits for high speed networking applications. Any *N*×*N* Birkhoff-von Neumann TDM switch could be constructed recursively from the designed TDM switch IC to achieve switching capacity of hundred gigabits per second or higher. The TDM switch IC contained a digital 8×8 TDM switch core with 8B10B CODECs and analog SERDES I/O interfaces. In the I/O interfaces, eight 2.56/3.2Gbps dual-mode 16/20:1 SERDES with CML buffers were developed. The 16/20:1 instead of 8/10:1 serializer and deserializer were used to reduce the required operating frequency in the switch core by half. New half-rate architectures and all static CMOS gates were used in the 16/20:1 serializer and deserializer for the low power consumption. A wide-band CML I/O buffer with our patented PMOS active load scheme was developed. All the implementations were based on the 0.18 µm CMOS technology. Our test results showed a 20 Gbps switching capacity for the 8×8 TDM switch IC.**

I. INTRODUCTION

There is an urgent need to built high-speed switches that scale with the transmission speed of fiber optics. As the key limitation of an electronic switch is the memory accessing speed, input-buffered switches, capable of performing parallel read/write, have received a lot of attention recently. High-end routers, such as Cisco 12000 and Juniper T640, are based on conflict resolution of parallel buffers. However, conflict resolution requires additional computation and communication overheads, which prohibit it from building switches with much higher speed.

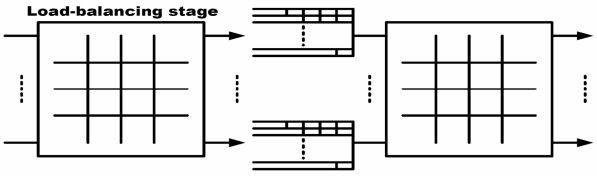


Figure 1. Load-balanced Birkhoff-von Neumann switch

The most important breakthrough switch architectures to overcome the problems of conflict resolution are the load balanced Birkhoff-von Neumann switch architectures [1,2,3]. They are proposed to resolve the memory access confliction and eliminate the extra computational and communication overhead by uniformly distributing input traffic. The load balanced Birkhoff-von Neumann switch consists of two crossbar switch fabrics (see Figure 1.) and parallel buffers between them. In a time slot, both the crossbar switch fabrics set up connection patterns that are periodically generated from a one-cycle permutation matrix. By so doing, the first stage performs load balancing for the incoming traffic so that the traffic coming into the second stage is uniform. As such, it suffices to use the same periodic connection patterns as in the first stage to perform switching at the second stage. In the load-balanced Birkhoff-von Neumann switch, there is no need to gather the traffic information. Also, as the connection patterns are periodically generated, no computation is needed at all. More importantly, the architecture is shown to achieve 100% throughput for any non-uniform traffic under a minor technical assumption [1,2].

The overall architecture of the switch is described in section II. The individual modules of the switch are presented in section III. Measurement results are given in section IV. A conclusion is summarized in section V.

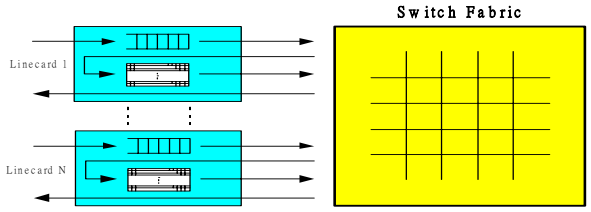


Figure 2. The Folded version of this switch

II. OVERALL ARCHITECTURE OF THE LOAD BALANCED TDM SWTICH

The folded version of load-balanced switch is used as shown in Figure 2 since the connection pattern is the same in the two stages N×N crossbar switch. The central buffers in the load-balanced Birkhoff-von Neumann switch are distributed among the linecards. As such, packets arriving at the linecard go through the switch fabric twice: one for load balancing and one for switching.

Now we describe how the connection patterns of these two crossbar switch fabrics are set up. In every time slot, both crossbar switches have the same connection pattern. For an $N\times N$ crossbar switch, the input port i will connect to output port j at time interval t, when

$$(i+j) \equiv (t+1) \mod N. \quad (1)$$

A switch fabric that implements the connection patterns in Eq. (1) is called a symmetric Time Division Multiplexing (TDM) switch.

An $N\times N$ symmetric TDM switch can be recursively constructed with an O ($N\times\log_2 N$) complexity. A two-stage construction of an $N\times N$ symmetric TDM switch (with $N = p\cdot q$) can be done in the following manner. The first stage consists of p $q\times q$ symmetric TDM switches and the second stage consists of q $p\times p$ symmetric TDM switches. The perfect shuffles are used at the input of the first stage and between the first stage and the second stage [3].

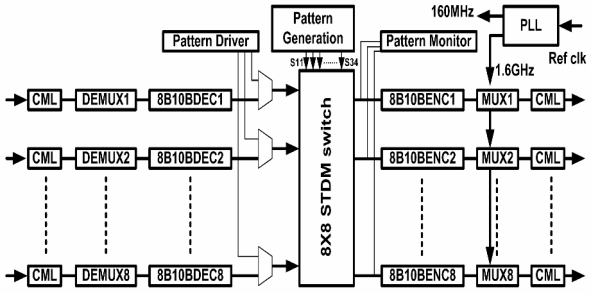


Figure 3. Symmetric TDM switch with SERDES interfaces

In this paper, we proposed and implemented an 8×8 scalable load-balanced Birkhoff-von Neumann Symmetric TDM switch IC with SERDES interfaces. Any $N\times N$ Birkhoff-von Neumann TDM switch could be constructed recursively from the designed TDM switch IC to achieve switching capacity of hundred gigabits per second or higher. The overall architecture includes an 8×8 load-balanced TDM switch core, eight 8/10B CODECs, eight SERDES ports, eight CML I/O interfaces and one PLL circuit as shown in Figure 3. Each receiving module contains a CML input buffer and a deserializer to convert the serial input into internal parallel data bus. Each transmitting module contains a serializer to convert the output data bus from the 8×8 TDM switch into a serial datum, which is sent out through the CML output buffer.

The 8×8 load-balanced TDM switch core is built from the 2×2 crossbar switches from Banyan network technology. The goal is to make the 8×8 crossbar switch can be decomposed into two independent 4×4 crossbar switch modules. The 8×8 crossbar switches can also be used as building modules for 64×64 or higher N×N switches.

A built-in-self-test (BIST) circuit is embedded in the TDM switch core. The BIST circuit provides input data pattern internally at full system clock rate. There is an on-chip monitor to verify the correctness of the switching outputs.

The 8B10B CODECs is used to generate DC-balanced data stream for the ease of clock data recovery at each input ports. The 8B10B CODECs can be bypassed for verifying regular patterns.

The SERDES (serializer/deserializer) interfaces are connected to the I/O ports. Data rate at each port is targeted at 3.2Gbps. We use half rate scheme to reduce the SERDES interface clock to 1.6GHz for 3.2Gbps data rate. We also use 16/20:1 MUX or 1:16/20 DEMUX in the SERDES to further reduce the switch core system clock to 160MHz.

The active load inductive peaking CML buffers are used in the input and output buffer in the SERDES interface. This active load inductive peaking technique not only achieve gigabit data transmission speed but also reduce 85% CML buffer area compared with passive inductive load circuit.

The power consumption is an important issue in the design due to large scale of the IC. We use all static CMOS design and half-rate scheme in the SERDES to reduce the power consumption. Low power design CML and PLL are also involved. In the digital switch core, we use 16/20:1(1:16/20) MUX and DEMUX to reduce the switch core frequency from 320MHz to 160MHz. From our simulation result, these techniques can save 200mW power consumption.

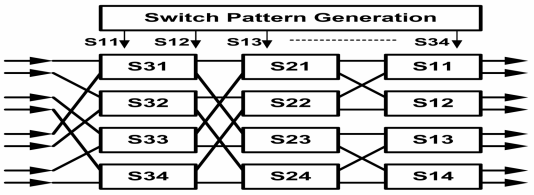


Figure 4. 8X8 Symmetric TDM switch core

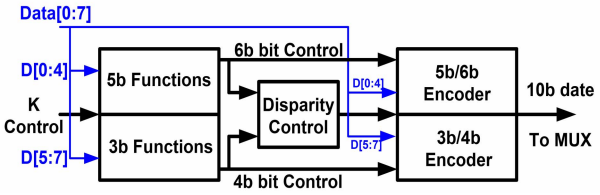


Figure 5. Architecture of an 8B10B Encoder

III. The Individual Module Architecture of the Symmetric TDM IC

A. Symmetric TDM Switch Core with 8B10B CODEC

An 8×8 symmetric TDM switch module can be constructed by using three stages of 2×2 crossbar switches. Each stage contains four 2×2 crossbar switches. These twelve 2×2 crossbar switches are labeled as $S_{l,m}$ (l=1,2,3; m=1,2,3,4) as shown in Figure 4. The pattern generation block produces the connection pattern for each 2x2 crossbar switch according to Eq. (1).

Two 8B10B CODECs [4] are used in parallel to perform the1:16/20 or 16/20:1 data conversion. Fig. 5 shows the

architecture of an 8B10B encoder. It contains three stages. First stage is the 5B function and 3B function that are used to generate classification signals for disparity checking. The second stage contains the disparity classification and complementary control signal generator. The final stage is the 5B to 6B and 3B to 4B data conversion.

B. Ultra-low-power 16/20:1 Dual Mode Serializer and Deserializer

When integrating the serializer/deserializer with the TDM switch core, the 8/10:1 serializer (deserializer) requires 320MHz clock frequency to achieve 2.56/3.2Gbps data rate. We propose a 16/20:1 serializer (deserializer) cooperating with two 8B10B encoders (decoders) that can reduce the operating frequency requirement in digital core by half. Therefore lots of core power consumption can be reduced in the system. Moreover, the 16/20:1 serializer (deserializer) implemented by all static CMOS gates effectively reduces the power from 50mW to 5mW as compared with the one implemented by the Source Coupled Logic (SCL).

The 16/20:1 serializer converts the sixteen (or twenty) bit input data into a serial datum. For the 20:1 mode, since it is not the power of 2, the general tree type multiplexer cannot be used. Therefore, we adopt the shift register approach to store the parallel inputs and send it out serially (see Fig. 6). We could treat the last 2:1 MUX and two DFFs in the dashed box in Fig. 6 as a double-edge-triggered DFF, which means that only a half rate clock is needed. For example, a 3.2Gbps data stream requires a 1.6GHz clock. The similar approach is done in the 16/20:1 deserializer.

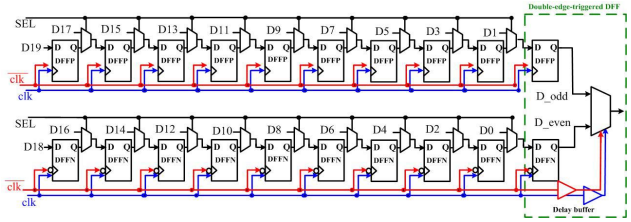


Figure 6. Block diagram of a 16/20:1 serializer

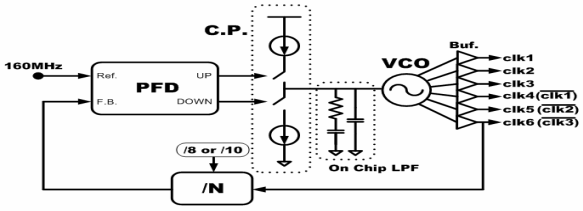


Figure 7. PLL Block Diagram

C. Phase Locked Loop

Multi-phase phase locked loop (PLL), shown in Fig. 7, contains a ring-type voltage controlled oscillator (VCO), a phase frequency detector (PFD), a charge pump (C.P.), a divided-by-8/10 two-mode divider, and a 2^{nd} order loop filter. The dual phase clocks, clk1 and clk4, from VCO modules are used as differential clocks for the 2-to-1 MUX and DFFs in serializer modules.

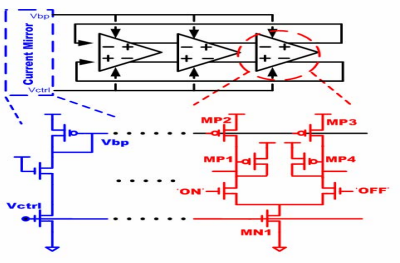
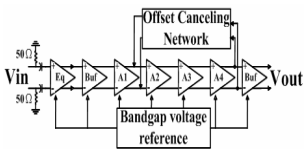


Figure 8. Three Stage Ring VCO

Since the 16/20:1 MUX is not a tree type, the PLL block doesn't need to generate quadrature phase clocks [5], but only to get half rate dual clocks. For this reason, a 3-stage ring VCO is enough to generate a wide-band clock. Fig. 8 shows the circuit schematic of the VCO, including a current mirror and a 3-stage ring oscillator generating up to 6 output phases [6]. It provides fixed swing and common mode operation, and makes this VCO suitable for a wide range of operating frequencies and supply voltage. To obtain 50% duty cycle full-swing output, a differential-to-single-ended converter circuit is designed in every VCO multi-phase output [7].

The PFD is the well-known PFD based on four RS latches [8] and a delay chain, which eliminates the dead zone. Only two transistors are used in charge pump to eliminate charges stored on parasitic capacitors [6]. Furthermore, the PLL design parameters, such as bandwidth and damping factor, change with the division ratio N in the feedback path. To compensate loop parameters for changes in N, the charge pump also turns on another current branch while the frequency divider is in N=10 mode.

The frequency divider circuit provides /8 and /10 two modes. We use the 2^{nd} order on chip loop filter to suppress the reference spurs. The capacitors are made up of MOS capacitors to reduce area consumption.



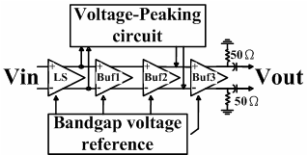


Figure 9. (i) CML input interface (ii) CML output buffer

D. CML I/O Interface

The architecture of the CML I/O interface is shown in Fig. 9. The CML input interface consists of an equalizer, an inductive-peaking active feedback CML limiting amplifier and a DC offset canceling circuit. The typical input sensitivity is 4mV and the limiting amplifier output swing is around 250mV. The CML output interface consists of a level-shift circuit, a voltage-peaking circuit and three-stage CML buffers, used as a backplane driver. The last stage of CML output buffer can provide approximately 8mA driving

current in order to drive 50 ohm load and let an output swing range up to 250mV.

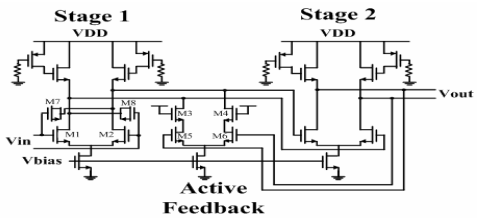


Figure 10. Block diagram of a basic CML Buffer

The architecture of a basic differential current-mode logic buffer circuit is shown in Figure 10. It includes an active inductor formed by PMOS transistors that act as active resistors connected to NMOS transistors load. They act as the on-chip inductors to employ inductive-peaking. Compared with on-chip inductors, active inductors require much lower chip area and consume less power but have the same frequency response. This CML buffer circuit also incorporates active feedback and negative Miller capacitance to meet high-speed requirement.

IV. Measurement

This symmetric TDM switch IC was fabricated using 0.18µm CMOS technology. The overall chip area (including PLL) is 3.65×3.57mm^2. Fig. 11 shows the die micrograph. Fig. 12 shows the measurement result of one channel 16/20:1 multiplexer with CML output buffer at 3.2Gbps. The eye diagram of the CML output buffer at 3.2Gbps is shown in Fig. 13. The differential output voltage is 250mVp-p with 20.2ps jitter at 3.2Gbps. The receiver sensitivity is 25mVp-p. The power consumption is summarized at Table I. The implementation results show that the 8×8 TDM switch with SERDES interface can achieve a 20 Gbps switching rate.

TABLE I. Summary of The STDM Switch

Process	0.18µm CMOS
Supply Voltage	1.8V
Overall Data rate	20Gbps
Total Area	3650x3570µm^2
Switch & 8 Ch CODEC Power	222mW
MUX& Driver Power /Ch	31mW
DEMUX & Driver Power /Ch	28mW
PLL Power	24mW

V. Summary and Conclusions

For the first time, the scalable load balanced TDM switch IC is implemented. The module is simple and can easily be scaled up to an *N*×*N* TDM switch. A 64×64 TDM switch can be recursively constructed from the 8x8 TDM modules to reach 160Gbps switching capacity using a 0.18µm CMOS technology. In the SERDES interface, we developed low power 8-channel CML transceivers using half rate and dual-mode 16/20:1 multiplexing schemes.

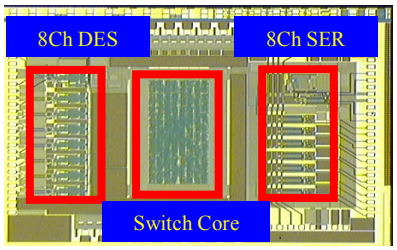


Figure 11. Die microphotograph

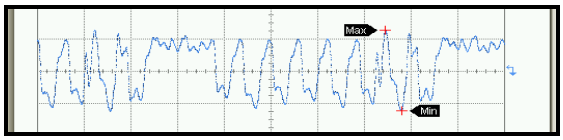


Figure 12. The measurement of one channel 16/20:1 multiplexer with CML output Buffer@ 3.2Gbps

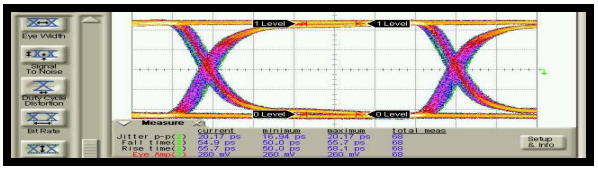


Figure 13. The eye diagram of the CML output buffer@3.2Gbps 2^7 – 1 PRBS input

Acknowledgment

The authors would like to acknowledge the support of this work by National Science Council (Taiwan) (NSC-95-2752-E-007-002-PAE), and the chip fabrication support of National Chip-Implementation Center (Taiwan) and Taiwan Semiconductor Manufacturing Company (TSMC). We also like to thank President W.T. Chen, Prof. C.S. Chang and Prof. D.S. Lee for support of this work.

References

[1] C. S. Chang, D. S. Lee and Y. S. Jou, "Load balanced Birkhoff-von Neumann switches, part I: one-stage buffering," Computer Communications, Vol. 25, pp. 611-622, 2002.

[2] C. S. Chang, D. S. Lee and C. M. Lien, "Load balanced Birkhoff-von Neumann switches, part II: multi-stage buffering," Computer Communications, Vol. 25, pp. 623-634, 2002.

[3] C. S. Chang, D. S. Lee, and Y. J. Shih , "Mailbox switch: a scalable two-stage switch architecture for conflict resolution of ordered packets, Proceedings of IEEE INFOCOM 2004.

[4] Widmer, A.X&Franaszek, P.A., A DC-Balanced, Partitioned-Block, 8/10B Transmission Code. IBM J. Res. Develop., Vol. 27, No. 5. September 1983.

[5] H. W. Lu, C. C. Su, "A 5Gbps CMOS LVDS Transmitter with Multi-Phase Tree-Type Multiplexer", 2004 IEEE Asia-Pacific Conference on Advanced System Integrated Circuits, 2004

[6] P. Larsson, "A 2-1600-MHz CMOS clock recovery PLL with low-Vdd capability", Solid-State Circuits, IEEE Journal of Volume 34, Issue 12, Dec. 1999.

[7] J.G. Manteatis, "Low-Jitter Process-Independent DLL and PLL Based on Self-Biased Techniques", Solid-State Circuits, IEEE Journal of Volume 31, Issue 11, Dec. 1996.

[8] B. Razavi, "Design of Analog VMOS Integrated Circuits", Mc.Graw-Hill . 2001.