A 32Gbps Low Propagation Delay 4×4 Switch IC for Feedback-Based System in 0.13μm CMOS Technology

Yu-Hao Hsu, Yang-Syu Lin, Ching-Te Chiu*, Jen-Ming Wu, Shuo-Hung Hsu,

Fan-Ta Chen, Min-Sheng Kao, and YarSun Hsu.
*Institute of Communications Engineering, National Tsing Hua University, Hsinchu, 300, Taiwan.
E-mail:ctchiu@cs.nthu.edu.tw

Abstract—**In this paper, a low propagation delay, low power, and area-efficient 4×4 load-balanced switch circuit for feedback-based system is presented. In this periodic and deterministic switch, only two DFFs are used to implement a pattern generator which is a $O(N^3)$ hardware complexity in traditional matching algorithm based $N×N$ switch. For packet reordering, a feedback path is established in series of symmetric patterns. As comparing with commercial switch systems, we implement a 4×4 switch IC directly in high speed domain without the use of SERDES interfaces to achieve low propagation delay and high scalability. In CML output buffer, PMOS active load and active back-end termination are introduced. A stacked current source and symmetric topology in CML-DFF are adopted. From our results, this work efficiently deducted 28ns propagation delay, 80% area and 80% power introduced by the SERDES interface. The throughput rate is up to 32Gbps (8Gbps/Ch).**

I. INTRODUCTION

As the optical communication improves, there is an urgent need to build a high speed switch in the core network. The load-balanced switch provided by C.S. Chang, etc. [1] is one of the most promising switch architecture since it has both 100% throughput and high scalability properties without computation overhead. One can implement the basic 2×2 or 4×4 switch block and then easily cascade these basic blocks to build a 16×16, 256×256, or even 1024×1024 switch system.

Recently, most researches focus on resolving the out-of-sequence issue in this two-stage switch architecture [2], [3]. The concept of feedback-based path in this two-stage system then is introduced. However, after scaling up, feedback-based system might degrade the system throughput since next packet has to stay for the feedback information from the last packet. The architecture of an $N×N$ ($N = 16$) load-balanced switch constructed by 4×4 switches is shown in Fig. 1(a). If the propagation delay is quite longer then the packet time, then throughput rate will reduced to the ratio of the packet time to the round-trip time (RTT) (as shown in left side of Fig. 1(b)).

Especially, to boost the bus bandwidth at each port, SERDES interfaces (serializer-and-deserializer) are commonly inserted in commercial switch systems to reduce pin counts and then reduce the routing complexity on printed circuit board. Taking the 16×16 load-balanced switch for example, four sets of SERDES in the switches and one in the linecard are included in one feedback path. Each pair of SERDES

interface contributes at least 200ns propagation delay [4], [5], and then results in over 1μs RTT in the system without the considering of routing delay.

Therefore, the feedback based system creates another scalability issue. One strategy is the pipeline method to fill the pipe and then gain the throughput (as shown in right side of Fig. 1(b)), but the penalty is to include the look-ahead block at each input port to predict the feedback information on the fly. To ease the complexity of look-ahead block, reducing the propagation delay introduced by SERDES interface is still needed. In this paper, we propose to implement the load balanced switch IC directly in high speed serial domain without the use of SERDES interface to reduce the RTT. PMOS active load and active back-end termination are introduced in the CML buffer. A stacked current source and symmetric topology in CML-DFF is adopted.

This paper is organized as follows. In section II, the circuit design techniques to boost switching speed are presented. In section III, the measurement results are provided. Then, in section IV, we give a short conclusion.
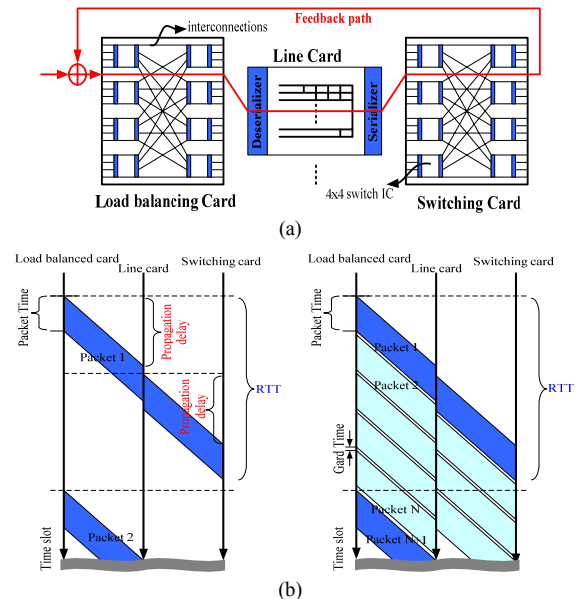

Figure 1.   (a) A 16×16 switch fabric constructed by 4×4 switch ICs;
(b) round-trip time of the feed-back path.

II. CIRCUITS DESIGN TECHNIQUE

The overall architecture of a 4×4 load balanced switch is shown in Fig. 2. The pattern generator, the CML-DFF in 2×2 switches, and the CML output interface are three key blocks to guarantee high-speed data transmitted from one input to another output directly in high-speed domain without the use of SERDES interfaces. In this section, we describe more detail about the circuit design techniques of these three blocks.
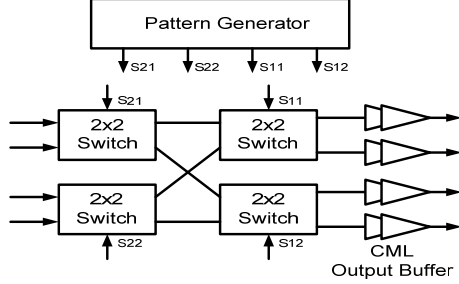

Figure 2. Block diagram of 4×4 load-balanced switch fabric.

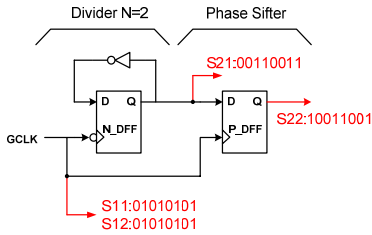### A. Pattern generator block design

Figure 3. The concpet of pattern generator block

Since connection patterns are periodic and deterministic, there is no need to find a $O(N^3)$ matching at every time slot. In this sub-section, we demonstrate that only two DFFs (as shown in Fig. 3) are used to implement the pattern generator for this periodic and deterministic switch as comparing with the matching algorithm based switch.

The switch pattern generation block produces connection patterns for the all 2×2 switches. The connection pattern of each 2×2 switch depends on its position in the $N×N$ symmetric TDM switch module and the current time slot. The column stage index l of each 2×2 switch is defined from right to left as $1, 2, ..., log_2N$ and the row stage index m is defined from top to bottom as $1, 2, ..., N/2$. The connection pattern of the mth switch of the lth stage at time t can be determined by Eq. (1).

$$\Psi(l,m,t) = \left\lfloor \frac{\left(t - \Phi(l,m)\right) \bmod 2^l}{2^{l-1}} \right\rfloor \quad (1)$$

where $\Phi(l,m) = \left((m-1) \bmod 2^{l-1}\right) + 1 \quad (2)$

We set the bar connection pattern if Eq. (1) equals to zero, and set the cross connection pattern otherwise.

There are three methods to implement the pattern generation block:
1) *Direct mapping from math-equations*
2) *Using shift registers to memories all the states*
3) *Using divider with a phase shifter*

Method 1 directly implements math equations (1) and (2) that deal with power-of-two modulus divisions and many time consuming arithmetic operations, such as the addition and subtraction. In method 2, equations are expanded in advance and then all states have to be memorized in considerable registers. Actually, connection patterns expressed by Eq. (1) are periodic. After observing the behavior of all states expanded by Eq. (1), the third method is proposed and only two DFFs are necessary for constructing a 4×4 switch circuit.
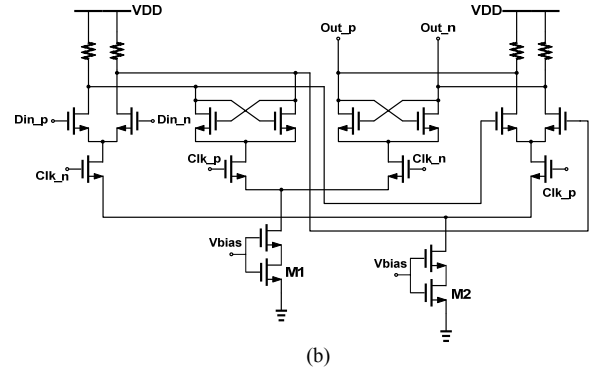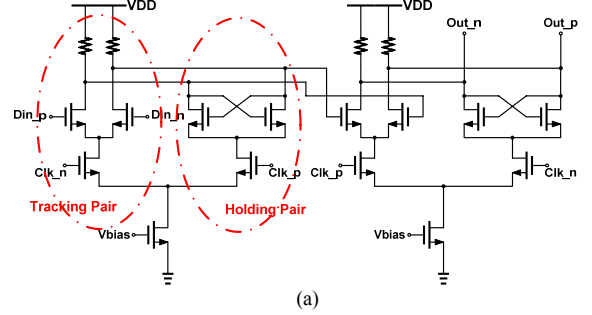
B. CML-DFF design

(a)

(b)

Figure 4. (a) Traditional DFF design; (b) modified DFF design.

In the switch system, the most important circuit block is the CML-DFF, which is composed of two D-latches, since it is responsible for transmitting serial data from one input port to another output at high speed.

Fig. 4(a) demonstrates a traditional CML-DFF circuit. Each CML latch consists of an input tracking pair, which is utilized to track the input data signal while the clock transistor pair switches the current to the left branch, and a cross-coupled regenerative pair (also called the holding pair), which is utilized to hold the data while the current is switched to the right branch. A few drawbacks exist in this circuit. Especially, two inherently different branches, tracking and holding, share the same current source, which in turn tie up the bias condition of these two circuits. At high-speed data-rates, the parasitic capacitances of the transistor degrade the required minimum small-signal gain for proper tracking operation. Therefore, the tail current source must be sufficiently high to achieve a wider range of linearity and a larger trans-conductance. On the other hand, the holding pair does not need a large bias current at ultrahigh-frequencies [6].

To solve these problems, a traditional CML-DFF is modified so that the tracking sides in the two latches share a single current source and the holding sides share another current source, as shows in Fig. 4(b). With this modification, the DFF also becomes more symmetric and thus results in a lower level of switching noise at 10Gbps data-rate [7]. In addition, each of the tail current sources in the DFF is replaced by a stacked current source, which consists of two cascaded

NMOS transistors [8]. The upper transistor is a low threshold voltage device and the bottom one is a regular threshold voltage device, as shows in Fig. 4(b). This configuration results in a flat current source characteristic since output resistance increase from r_o to r_o^2. Here is our derivation:

$$R_o = r_o + r_o(1 + g_m r_o) = 2 r_o + g_m r_o^2 \doteqdot g_m r_o^2 \qquad (3)$$

C. Back-end termination design

The CML output interface is shown in Fig. 5. This output interface consists of two-stage CML buffers. In the first stage, we use our patent of PMOS active load inductive peaking technique [9] to improve the high-frequency performance. In the last stage, we propose the active back-end termination for impedance match of the 50Ω load.

The traditional CML output interface is with resistor load. To improve the high-frequency bandwidth, one can choose the on-chip inductors to replace resistors. However, on-chip inductors occupy largest chip area and introduce significant parasitic capacitance. In our design, we use PMOS active load inductive peaking technique [9] in the first stage of CML output buffer (see Fig. 9[a]) to enhance the bandwidth. It includes active inductors formed by PMOS transistors (M9-M10) that act as active resistors connected to NMOS transistors load (M7-M8). They act as the on-chip inductors to employ inductive-peaking. Compared to on-chip inductors, active inductors require much lower chip area and consume less power but have the same frequency response. We also incorporate negative Miller capacitance (M3-M4) to meet high-speed requirement.

With the increasing operation speed of the communication network, the signal reflection is getting worse due to the impedance mismatch and it impacts the performance of the transmission. To resolve this problem, some circuit designs use passive back-end termination but it costs 50% modulation current. Some other circuit designs use AC-coupled back-termination but it is very difficult to design a high quality capacitor in chip process and it occupies large chip area too. As shown in Fig. 5(a), we propose the active back-end termination technique in the last stage of CML output buffer to match the 50Ω load environment. This scheme provides high current driving efficiency than passive back-end termination. As comparing with AC-coupled active back-end termination, it occupies less chip area due to no need for on-chip capacitor. Fig.5(b) shows the overall output interface output impedance for 50Ω load system varied with operation speed of circuitry.
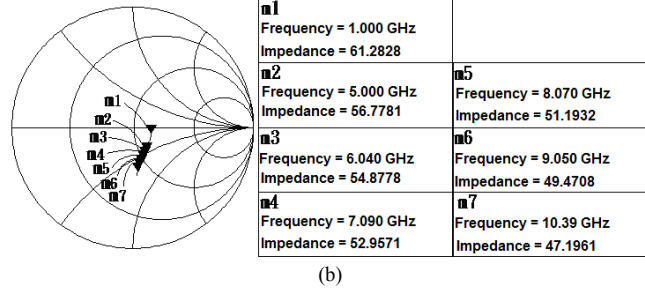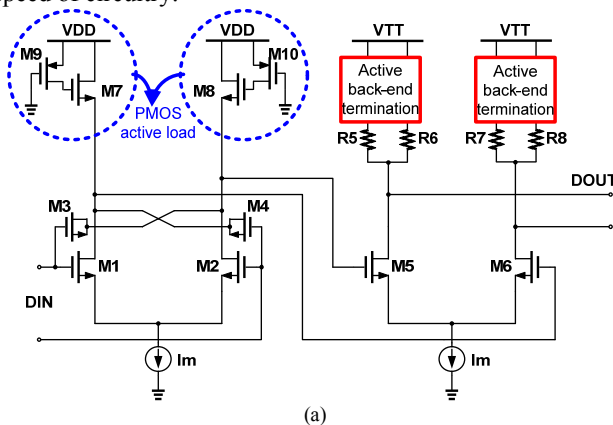

(a)

(b)

Figure 5.   The CML output interface:
(a) active load inductive peaking and active back-end termination;
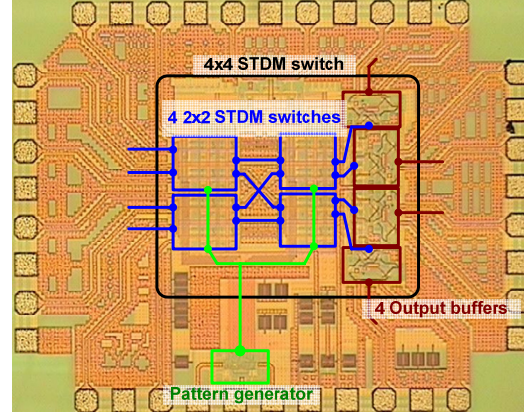(b) impedance char of back-end termination.

Figure 6.   The die poto of the 4×4 load-balanced switch.
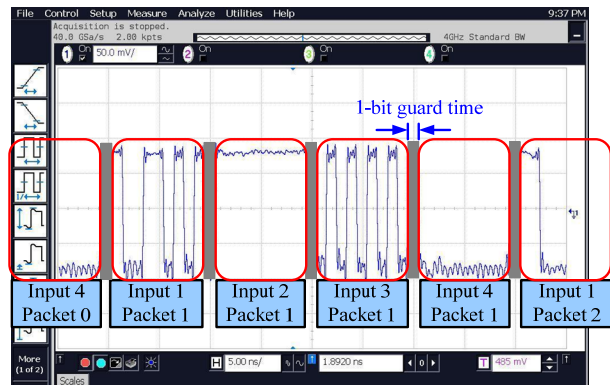

Figure 7.   Evaluation board.

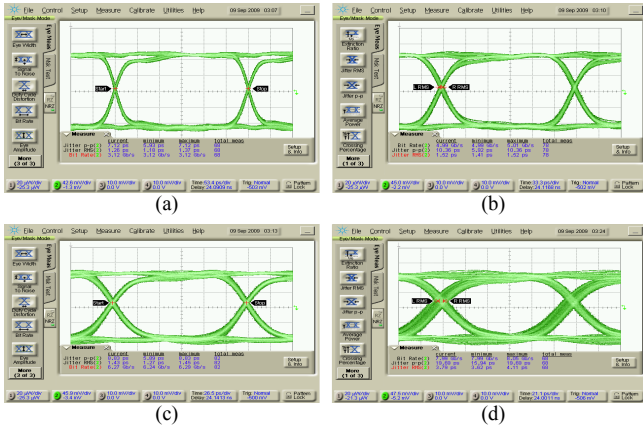Figure 8.   One of the output waveform of the 4×4 load-balanced switch IC.

583

Figure 9. Measured eye diagrams at different specifications
(a) Jitter$_{p-p}$ = 7ps @ 3.125Gbps;   (b) Jitter$_{p-p}$ = 10ps @5Gbps;
(c) Jitter$_{p-p}$ = 9ps @6.25Gbps;   (d) Jitter$_{p-p}$ = 20ps @8Gbps.
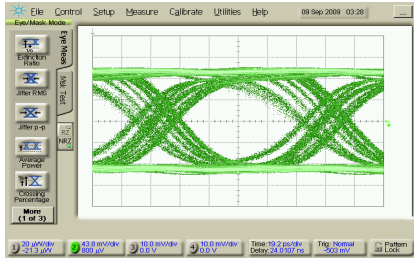

Figure 10. Measured eye diagram at 9Gbps@2^7-1PRBS input.

III. MEASUREMENT RESULTS

The 4×4 load-balanced switch IC has been implemented in 0.13μm CMOS technology. The total area including PADs is 1380×1080μm^2, which is almost 20% of previous works as shown in Table I. Fig. 6 shows the chip micro photo of the 4×4 switch. The printed circuit board configuration is shown in Fig. 7. Four layers PCB is fabricated with Nelco 4000-13(10Gbps transmission rate guaranteed). One of the output waveform is presented in Fig.8. For the ease of the demonstration, we input series of packets at input 1 with PRBS content, series of packets at input 2 with logic '1' content, series of packets at input 3 with '0101…' content, and series of packets at input 4 with logic '0' content. Packets are evenly switched to each output port with 1-bit guard time. We test it with different data rate from 3.125Gbps, 5Gbps, 6.25Gbps, 8Gbps to 9Gbps. Eye diagrams are shown in Fig. 9 and 10 respectively. Table I shows the comparison with previous works. In table I, type-I switch system in [11] and Type-II switch system in [10] are implemented with different SERDES interface structures.

IV. CONCLUSIONS

A low propagation delay, low power, and area-efficient 4×4 load-balanced switch circuit for feedback-based system is presented. In this periodic and deterministic switch, only two DFFs were used to implement the pattern generator which is a $O(N^3)$ complex combinational block in traditional matching algorithm based switch. For packet reordering, a feedback path is established in series of symmetric patterns. In CML output buffer, PMOS active load and active back-end termination are introduced. We adopt a stacked current source

and the symmetric topology in CML-DFF. Traditionally, the SERDES interface was adopted in commercial switch system to deduce interconnections for scalability. However, the long propagation delay introduced by SERDES interfaces caused another high complex look-ahead block in a feedback-based system and this result in another scalability issue. In this paper, we implement a 4×4 switch IC directly in high speed domain without the use of the SERDES interface. This work not only reduces the interconnections but also deducts at least 28ns propagation delay, 80% area and 80% power introduced by the SERDES interface. The throughput rate is up to 32Gbps (8Gbps/Ch).

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, April 2002.

[2] C. S. Chang, D. S. Lee, Y. J. S, and C. L. Yu, "Mailbox switch: a scalable two-stage switch architecture for conflict resolution of ordered packets," *IEEE Transactions on Communications,* vol. 56, pp. 136-149, January 2008.

[3] C. L. Yu, C. S. Chang, and D. S. Lee, "CR switch: a load-balanced switch with contention and reservation," *IEEE/ACM Transactions on Networking,* accepted for future publication, 2009.

[4] Texas Instruments 10 Gigabit (XAUI) Ethernet Transceivers Datasheet: http://focus.ti.com/lit/ds/symlink/tlk3138.pdf.

[5] XILINX RocketIO™ Transceiver User Guide: www.xilinx.com/support/documentation/user_guides/ug024.pdf

[6] P. Heydari and R. Mohanavelu, "Design of ultrahigh-speed low-voltage CMOS CML buffers and latches," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 12, pp. 1081–1093, Oct. 2004.

[7] T. Otsuji, M. Yoneyama, K. Murata, and E. Sano, "A super-dynamic flip–flop circuit for broadband applications up to 24 Gbit/s utilizing production-level 0.2-um GaAs MESFETS," *IEEE Journal of Solid-state Circuits*, vol. 32, pp. 1357–1362, Sep. 1997.

[8] H. D. Wohlmuth and D. Kehrer, "A low power 13-Gb/s 27^{-1} pseudo random bit sequence generator IC in 120 nm bulk CMOS," in *proceedings IEEE Symposium on Integrated Circuits and Systems Design*, Pernambuco, Brazil, Sep. 7-11, 2004, pp. 233–236.

[9] U.S. patent: M. Kao, C. Jen, J. Wu, C. Chiu, and S. Hsu, "Transmission circuit for use in input/output interface," No. US20070069769A1.

[10] C. T. Chiu, Y. H. Hsu, M. S. Kao, H. C. Tzeng, M. C. Du, P. L. Yang, M. H. Lu, F. T. Chen, H. Y. Lin, J. M. Wu, S. H. Hsu, and YarSun Hsu, "A Scalable Load Balanced Birkhoff-von Neumann Symmetric TDM Switch IC for High-Speed Networking Applications," in *Proceedings IEEE International Symposium on Circuits and Systems (ISCAS'07)*, New Orleans, Louisiana, USA, May 27-30, 2007, pp. 2754-2757.

[11] Y. H. Hsu, M. H. Lu, P. L. Yang, F. T. Chen, Y. H. Li, M. S. Kao, C. H. Lin, C. T. Chiu, J. M. Wu, S. H. Hsu, and YarSun Hsu, "A 28Gbps 4×4 switch with low jitter SerDes using area-saving RF model in 0.13μm CMOS technology," in *Proceedings IEEE International Symposium on Circuits and Systems (ISCAS'08)*, Seattle, Washington, USA, May 18-21, 2008, pp. 3086-3089.

TABLE I.   COMPARISON TABLE

	This work	Type-I [11]	Type-II [10]
Technology	0.13μm	0.13μm	0.18μm
Supply Voltage	1.2V	1.2V	1.8V
Max. Speed/Ch	9Gbps	8.8Gbps	3.2Gbps
Overall Throughput	32Gbps	28Gbps	25.6Gbps
Jitter	20ps	18ps	21ps
Chip Size (including PAD)	1380×1080 μm^2	3000×2480 μm^2	3650×3570 μm^2
Overall Power	134mW	850mW	730mW
Propagation delay	0.8ns	29.5ns	50ns