Design of Ultra High-Speed CMOS CML buffers and Latches

Payam Heydari, Ravi Mohavavelu
Department of Electrical and Computer Engineering
University of California
Irvine, CA 92697-2625
E-mail: payam@ece.uci.edu, RMOHAVA1@irf.com

Abstract - **A comprehensive study of ultra high-speed current-mode logic (CML) buffers and regenerative CML latches will be illustrated. A new design procedure to systematically design a chain of tapered CML buffers is proposed. Next, a new 20GHz regenerative latch circuit will be introduced. Experimental results show a higher performance for the new latch architecture compared to a conventional CML latch circuit at ultra high-frequencies. It is also shown, both through the experiments and by using efficient analytical models, why CML buffers are better than CMOS inverters in high-speed low-voltage applications.**

1. INTRODUCTION

High-speed Buffers and latches are the circuit cores of many high-speed blocks within a communication transceiver and a serial link. Front-end tapered buffer chain, serial-to-parallel converters, clock and data recovery (CDR), multiplexers, and demultiplexers all use high-speed buffers and latches with a robust performance in the presence of noise [1] [2].

CMOS current-mode logic buffers were first introduced in [3] to implement a giga-hertz MOS adaptive pipeline technique. the CML circuits can operate with lower signal voltage and higher operating frequency at lower supply voltage than CMOS circuits can. However, CML buffers suffer from dissipating more static power than CMOS inverters. Recently, there have been efforts to alleviate this shortcoming [4]. Due to their superior performance, CML buffers are the best choice for high-speed applications. As a consequence, it is an essential need to have a systematic approach to optimally design CML buffers and CML buffer chains.

This paper presents a the systematic procedure of CML buffer design and introduces a new CMOS CML latch circuit. The paper is organized as follows. First, in section 2, the large-signal behavior of a differential circuit is extensively illustrated. This will prepare us to study the design of CMOS buffer chain. In section 3 we illustrate a new 0.18μm CMOS CML latch that is capable of working at 20GHz. Finally, section 4, provides the experimental results that verify the accuracy of our design approach.

3. TAPPERED CML BUFFER DESIGN

A current-mode logic (CML) buffer is based on the differential architecture. Fig. 1. (a) shows a basic differential architecture. The tail current, I_{SS}, provides an input-independent biasing for the circuit. The differential circuit is easily neutralized using a pair of capacitors (Fig. 1.(a)), C_D, that will diminish the deleterious effects of input-output coupling through the device overlap capacitance, C_{GD}.

As the differential input varies from $-\infty$ to $+\infty$, each output node of the differential pair varies from $V_{DD}-R_D I_{SS}$ to V_{DD}. Fig. 1 (b) shows the voltage variations of the output nodes in terms of the differential input [5].

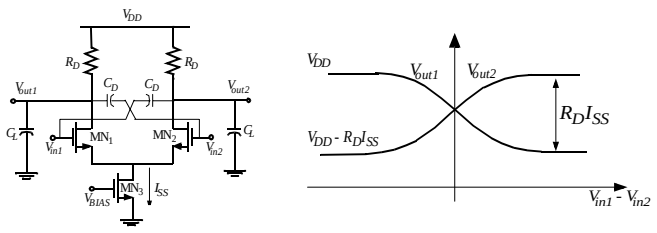

Fig. 1. (a) A neutralized CMOS differential pair. (b) Transfer characteristics.

From Fig. 1. (a) one can see that the maximum output differential voltage swing, V_{odm}, is only a function of the drain resistor and the tail current, provided that the current switching takes place. Clearly, the maximum output swing of a CML buffer is less than that of a CMOS inverter, which makes this class of buffers an ideal choice for low-power integrated circuit design.

The minimum value of the input common-mode level, $V_{in,CM_{min}}$ is achieved when the tail current begins to operate in saturation. The input common-mode level reaches its maximum value, $V_{in,CM_{max}}$ when the transistors MN_1 and MN_2 are either at pinch-off or at cutoff [5].

$$V_{GS,12}+(V_{GS3}+V_{THN}) \le V_{in,CM} \le min\left[V_{DD}-R_D\frac{I_{SS}}{2}+V_{THN}, V_{DD}\right] \quad (1)$$

where V_{GS12} is the common-mode overdrive voltage of transistors MN_1 and MN_2. Similarly, the output common-mode level varies from V_{DD} (when both MN_1 and MN_2 are off, and MN_3 is in the linear region) to $V_{DD}-R_D I_{SS}/2$ (when all transistors are in saturation). The voltage transition of the output common-mode level from V_{DD} to $V_{DD}-R_D I_{SS}/2$ is determined by the subthreshold current of MN_1 or MN_2.

To achieve the best performance in a CML buffer, a complete current switching must take place, and the current produced by the tail current needs to flow through the ON branch only. In a tapered buffer chain a CML buffer drives another buffer, which means that output terminals of the driving buffer stage are connected to the input terminals of the driven stage, as shown in Fig. 2. To satisfy the above performance requirement, the differential voltage swing of the first CML buffer must exceed $\Delta V_{in2,max}=\sqrt{2I_{SS2}/(\mu_n C_{ox}(W/L)_2)}$ of the following stage, or:

$$R_{D1}I_{SS1} \ge \sqrt{2I_{SS2}/\left(\mu_n C_{ox}\left(\frac{W}{L}\right)_2\right)} \quad (2)$$

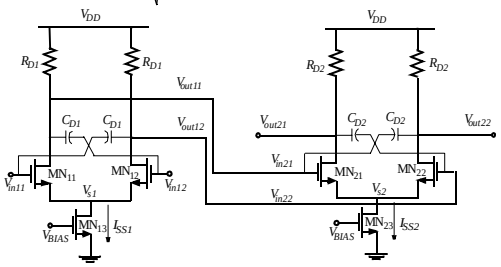

Fig. 2. Two CML buffers in cascade

Furthermore, the load resistors should be small in order to reduce the RC delay and increase the bandwidth. To guarantee a high-speed operation, NMOS transistors of the differential pair must operate only in the saturation. To satisfy this requirement for the circuit shown in Fig. 1, first, the input common-mode voltage must be within the interval specified in Eq. (1); and secondly,

$$V_{in_k,max}-V_{THN} \le V_{out,kj} \le V_{DD} \quad \text{for} \quad k=1,2 \text{ and } j=1,2 \quad (2)$$

which sets a maximum allowable level for the differential output swing as follows:

$$R_{Dk} I_{SSk} \le V_{THN} \qquad \text{for} \quad k = 1, 2 \qquad (3)$$

In addition, a high-speed CML output driver must drive a large off-chip load through the bondwire and package trace. The output driver must thus have a large current drive capability. This means that NMOS transistors of the second CML buffer in Fig. 2 must be large. A large transistor has a large gate-to-channel capacitance that seriously degrades the propagation delay and the voltage swing of the preceding predriver stage. To reduce the propagation delay of the predriver, a chain of tapered buffers is introduced between the first predriver stage and the second buffer. The minimum delay is obtained by dividing the delay equally over all stages. This is achieved by gradually scaling up all stages with a constant taper factor, *u*. On the other hand, the chip package interface at very high frequencies is appropriately modeled as a transmission line that is terminated by a load impedance, which is a series RC circuit (*cf.* Fig. 3). The series load resistance, Z_0, provides the high-frequency parallel matched termination to the bondwire. Fig. 3 shows the schematic of the output CML driver driven by N-1 tapered CML buffers along with the chip-package interface being modeled as the transmission line.

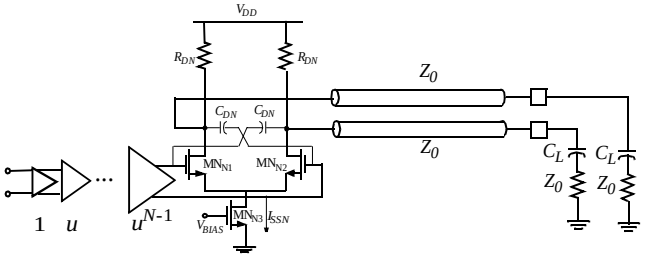

Fig. 3. An output CML buffer driving off-chip loads. The chip-package interface is electrically modeled using a lossless transmission line.

The chip bondwires exhibit high-Q inductances. Therefore it is safe to model the chip-package interface using a lossless transmission line. To avoid potentially disastrous transmission line effects such as slow ringing and propagation delays, the bondwires are terminated both at the source using a series termination ($R_{DN} = Z_0$), and at the destination using a parallel termination (Z_0). Given a well-defined output voltage swing ($R_D I_{SS}$) and with R_D being determined by the matched termination, the tail current I_{SSN} is easily calculated. For instance, an output differential voltage swing of 0.4V for a 50Ω line driver requires a bias current of 8mA. Now, using a set of constraints, we present design guidelines to design a tapered CML buffer chain and determine appropriate values for the circuit components of the CML buffer.

The propagation delay is computed using the open-circuit time constant method [6]. For instance, the delay of the simple low-voltage differential stage of Fig. 1 (a) is $0.69 R_D C_L$. Various HSPICE simulations on high-speed CML buffers show that the delay obtained by the open-circuit time-constant method is within 8% of the actual simulation.

Minimizing the overall propagation delay of CML buffer increases the overall operation frequency of the buffer significantly. For a slowly varying input signal, increasing the small-signal voltage gain will further decrease the output transient variations and the output transition time. In a chain of tapered CML buffers, to attain a constant voltage swing, transistor sizes are scaled up while the drain resistances are scaled down with a constant scaling factor. This will lead us to the fact that small-signal voltage gains of all constituting stages of the buffer chain are identical.

$$\sqrt{\mu_n C_{ox} \left(\frac{W}{L}\right)_1 I_{SS1}}\, R_{D1} = \sqrt{\mu_n C_{ox} \left(\frac{W}{L}\right)_2 I_{SS2}}\, R_{D2} = \cdots = \sqrt{\mu_n C_{ox} \frac{W}{L} I_{SS}}\; R_D$$

As a consequence, Eq. (5) provides us with a lower bound for the maximum small-signal voltage gain at equilibrium, that is:

$$\left(A_{v,eq} = \sqrt{\mu_n C_{ox} \frac{W}{L} I_{SS}}\; R_D\right) \ge \sqrt{2} \qquad (4)$$

The drain resistor, R_{DN}, of the last output CML buffer is determined by the series impedance matching to bondwire's characteristic impedance. Subsequently, I_{SSN} of the last driver stage is calculated using the output differential voltage swing and R_D. The only remaining parameter in the last CML driver left is the (W/L) of the source-coupled transistor pair, which is obtained from the common-mode characteristic of the last CML buffer. If the common-mode input voltage lies in the allowable range given by Eq. (1), then the tail current is equally divided between the two branches of the differential stage, i.e.,

$$\left(V_{in_k,CM} - V_{sk} - V_{THN} \ge V_{in_k,CM} - V_{BIAS} - 2V_{THN}\right) = \sqrt{I_{SSk} \Big/ \left(\mu_n C_{ox} \left(\frac{W}{L}\right)_k\right)}$$

$$\text{for } k = 1, 2, \ldots, N \qquad (5)$$

where $V_{in_k,CM}$ is the common-mode input voltage of the k[th] driver in the buffer chain. $V_{in_k,CM}$ is specified by the output common-mode voltage of the previous stage. Given a ***tapered buffer chain*** with a constant differential voltage swing, the maximum (W/L) of the transistor pair of the k[th] CML buffer is then calculated by solving Eq. (6):

$$V_{DD} - R_D \frac{I_{SS}}{2} - V_{BIAS} - 2V_{THN} = \sqrt{I_{SSk} \Big/ \left(\mu_n C_{ox} \left(\frac{W}{L}\right)_k\right)} \qquad (6)$$

In the above equation $R_D I_{SS}$ is the constant differential output swing of a tapered CML buffer chain.

As mentioned above, in a chain of tapered CML buffers, the minimum delay is obtained by dividing the delay equally over all stages. However, the question is how many buffer stages are required to achieve the optimum delay. To answer this question, the propagation delay of an arbitrarily chosen CML stage in a buffer chain is first derived. Fig. 4 shows the k[th] stage in a chain of N tapered stages driving another CML stage along with the capacitors that contribute to the delay calculation.

The common node s_{k+1} shown in Fig. 4 undergoes a smaller variation compared to the voltage variations of the input terminals particularly in a matched differential pair. In fact, it is easily shown that for a maximum differential input variation of $\Delta V_{in,max}$, the maximum variation of the common node is $\Delta V_{in,max}/\sqrt{2}$. Therefore, the equivalent capacitance seen at the common node s_{k+1} is approximately $C_{s,k+1} = \sqrt{2}\, C_{DBS,k+1}$ rather than $C_{DBS,k+1}$.

The 50% delay of the k[th] stage is as follows:

$$t_{d,k} = 0.69 R_{D,k} (C_{DB,k} \textcircled{S} C_{s,k} + C_{GS,k+1} \textcircled{S} C_{s,k+1}) \qquad (7)$$

where Ⓢ represents the series connection of electrical elements. The total propagation delay of the buffer chain is readily calculated:

$$t_d = \sum_{k=1}^{N} t_{d,k} = 0.69 N R_{D1} (C_{DB1} \textcircled{S} C_{s1} + X^{1/N} C_{GS1} \textcircled{S} C_{s1}) \qquad (8)$$

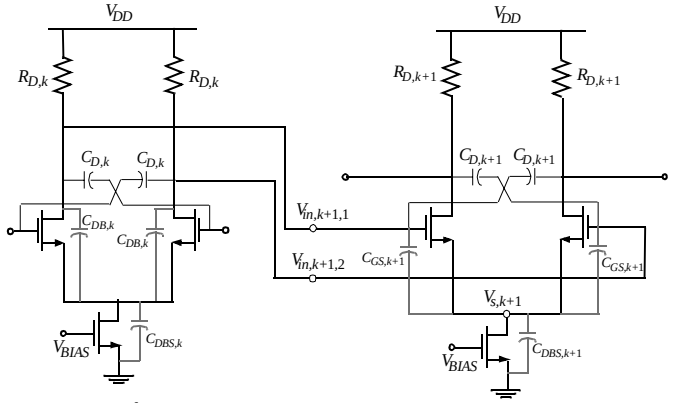

Fig. 4. The k[th] and (k+1)[st] stages of a tapered CML buffer along with the parasitic capacitances

Interestingly, the functional dependence between delay and the number of stages (or taper factor) is similar to the one in a CMOS

buffer chain [7]. To be more specific, consider a chain of tapered CML buffers driving a lossless transmission line with a characteristic impedance of Z_0. Suppose that the gate aspect-ratio of the transistor pair of the last CML line driver is X times larger than that of the first predriver stage. It is easily proved that if $C_{DB1} = \gamma C_{s1}$ and $C_{GS1} = \eta C_{s1}$; then it is easily proved that the optimum number of stages will be the numerical solution to the following equation:

$$X^{1/N_{opt}} = \exp\left[\frac{\frac{\gamma/\eta}{1+\gamma} + \left(\frac{1}{1+\eta}\right) X^{1/N_{opt}}}{\left(\frac{1}{1+\eta}\right) X^{1/N_{opt}}}\right] \qquad (9)$$

or in the special case, if $C_{DB1} << C_{GS1}$ then, $N_{opt} = \ln(X)$ which is well-known result.

To further increase the bandwidth (reduce the delay), the intermediate stages use inductive peaking as demonstrated in Fig. 5.

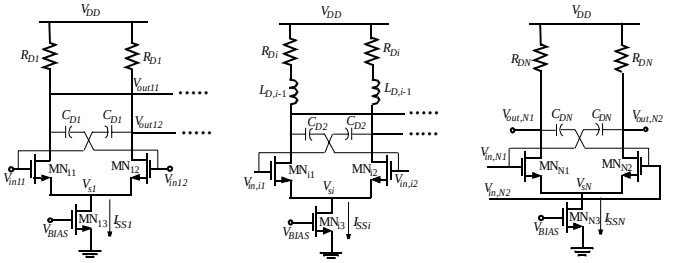

Fig. 5. Multiple stage CML buffers along with the inductive peaking

The addition of the inductor in series with the drain resistor delays the current flow through the branch containing the resistor, making more current available for charging the device capacitors, and reducing the rise and fall times. From another perspective, the addition of an inductance in series with the load capacitance introduces a zero in the transfer function of the CML stage which helps offset the roll-off due to parasitic capacitances. Inductive peaking can increase the bandwidth to about 1.72 times larger than the unpeaked case [6]. Inductance values are scaled with the same taper factor as the drain resistors are.

3. ULTRA HIGH-SPEED LATCH DESIGN

A current-mode logic (CML) latch consists of an input tracking stage, M_{N1} and M_{N2}, utilized to sense and track the data variation and a cross-coupled regenerative pair, M_{N3} and M_{N4}, being employed to store the data. Fig. 6 demonstrates a CMOS CML latch circuitry.

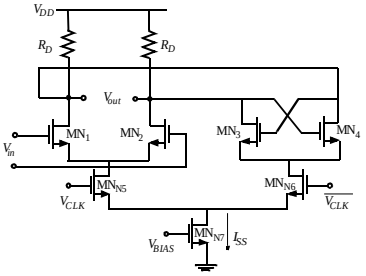

Fig. 6. the circuit schematic of a CMOS CML buffer.

The track and latch modes are determined by the clock signal inputs to a second differential pair, M_{N5} and M_{N6}. When the signal V_{CLK} is "HIGH", the tail current ISS entirely flows to the tracking circuit, M_{N5} and M_{N6}, thereby allowing V_{out} to track V_{in}. In the latch-mode, the signal V_{CLK} goes low, the tracking stage is disabled, whereas the latch pair is enabled storing the logic state at the output.

Like CML buffers, a CML latch operates with relatively small voltage swings which is $4V_{THN}$ peak-to-peak differential-mode. Fig. 6 allows us to implement high-speed latch circuit. However, there are several shortcomings involved in the design of the regenerative latch in Fig. 6, that lead to a complete operation failure at very high-frequencies ($\geq 10GHz$). The primary limitation is that a single tail current is used for both tracking and latch circuits. Consequently, the bias operations of tracking and latch circuits are tightly related. This will severely limit the allowable transistor sizes for a reliable latch operation. At ultra high-frequencies ($\geq 10GHz$) the parasitic capacitances of transistors, M_{N1} and M_{N2}, degrade the required minimum gain for a proper tracking operation (Eq. (4)). Therefore, the tail current must be sufficiently high to achieve a wider range of linearity and a larger transconductance. On the other hand, the latch circuit does not need a large bias current at ultra high-frequencies.

To address the aforementioned problems, the regenerative CML latch is modified so that the latch circuit and the tracking circuit use two distinct tail currents. Fig. 7 shows the new CML latch circuit.

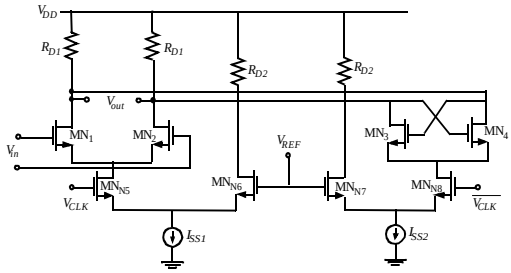

Fig. 7. The circuit schematic of the new CMOS CML latch circuit.

As observed in Fig. 7, the tracking stage and the latch stage are now separately optimized for a correct latch operation at ultra high-frequencies. Note that it is important the source coupled pair transistors have high gain. This is obviously achieved with larger W/L for each transistor of the cross-coupled pair. However, this technique greatly limits the driving capability. Therefore the CML latch is followed by a CML buffer to recover the logic level.

4. EXPERIMENTAL RESULTS

In this section the performance of the CML buffer is evaluated by performing experiments on single stage as well as multiple stages of the buffer. Experiments are set up to show the performance of the new CML latch depicted in Fig. 7 at 20GHz data-rate. First, the accuracy of Eq. (9) is verified by running HSPICE simulation on a chain of CML buffers. Then, the performance of the circuit in Fig. 7 is compared with the conventional CML latch shown in Fig. 6.

4.1. Tapered CML buffer experiment

Similar to a CMOS tapered buffer, a single CML buffer might not be sufficient to drive an off-chip load. There are, however, more design trade-offs involved in the design of a CML tapered buffer than in a CMOS tapered buffer. A superior high-frequency performance in a CML buffer is guaranteed only if the design guidelines explained thoroughly in Section 2 to be taken into consideration.

Fig. 8 plots propagation delay as a function of number of CML stages for different values of X, where X is the ratio between the off-chip load impedance and the load impedance of the first predriver stage. In practice, X is between 30-100.

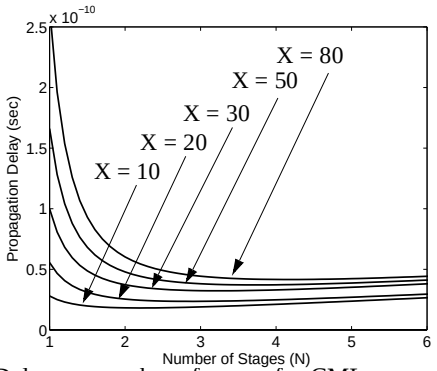

Fig. 8. Delay vs. number of stages for CML tapered buffer chain.

The delay variation in terms of the number of stages for CML tapered buffer and CMOS tapered buffer are almost identical. However, the total propagation delay of a CML buffer chain for a given value of X is less than that of CMOS buffer chain, which is in accordance with what is expected. Remember that 50% propagation delay of a CMOS inverter is inversely proportional to NMOS and PMOS transconductance parameters and directly proportional to the load capacitance [1]. According to (8), the propagation delay of a CML buffer is directly proportional to the load capacitance (similar to a CMOS inverter) and the drain resistance. A larger threshold voltage and a lower drift velocity associated with a PMOS transistor cause the propagation delay of a CMOS inverter to be larger than that of a CML buffer that uses the same transistor size.

4.2. Inductive peaking

The inductive peaking was proposed as an efficient and simple circuit technique to speed up the buffer's response. Figures 9 (a) and (b) demonstrate the differential output voltage of a CML buffer without and with the inductive peaking, respectively. The inductance value is 4nH and signals are running at 5GHz. The output voltages of CML buffer in the presence of inductance will have larger amplitude and as a result faster rise and fall times.

4.3. CML Latch

The proposed CML latch circuit in Fig. 7 has a superior performance compared to the one shown in Fig. 6 at ultra high-frequencies ($\geq 10GHz$) for the input data-rate. Figures 10 and 11 demonstrate outputs of the latch circuits at 20GHz data-rate. The conventional CML latch fails to operate as a latch at this frequency, whereas the new CML latch is capable of working as a latch at this frequency.

6. Conclusions

In this paper we investigated important problems involved in the design of a CML buffers and latches. A new design procedure to systematically design a chain of tapered CML buffers was proposed. We proved that the differential architecture of a CML buffer makes it functionally robust in the presence of environmental noise sources (e.g., crosstalk, power/ground noise). A new 20GHz regenerative latch circuit will be introduced. Experimental results show a higher performance for the new latch architecture compared to other existing latch circuits. It was also shown, both through the experiments and by using efficient analytical models, why CML buffers are better than CMOS inverters in high-speed low-voltage applications.

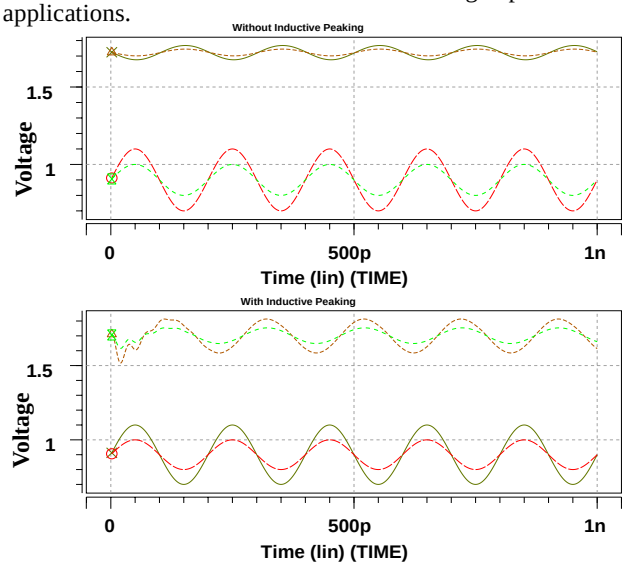

Fig. 9. (a) Input and output waveforms of a CML buffer without inductive peaking. (b) Input and output waveforms of a CML buffer with inductive peaking

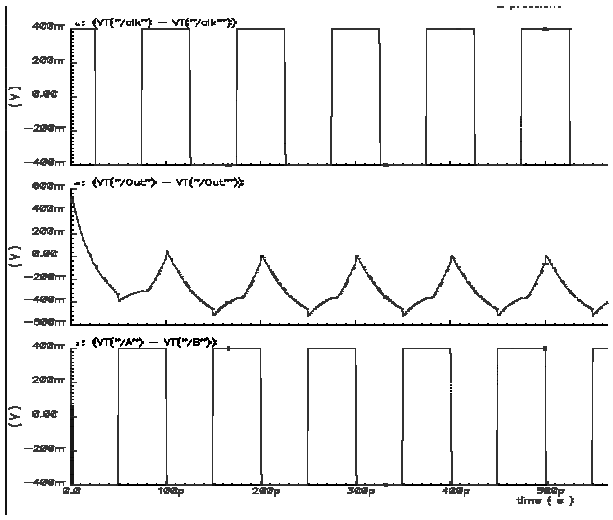

Fig. 10. The 20GHz conventional CML latch. (a) The input data at 20GHz. (b) The half-rate clock at 10GHz

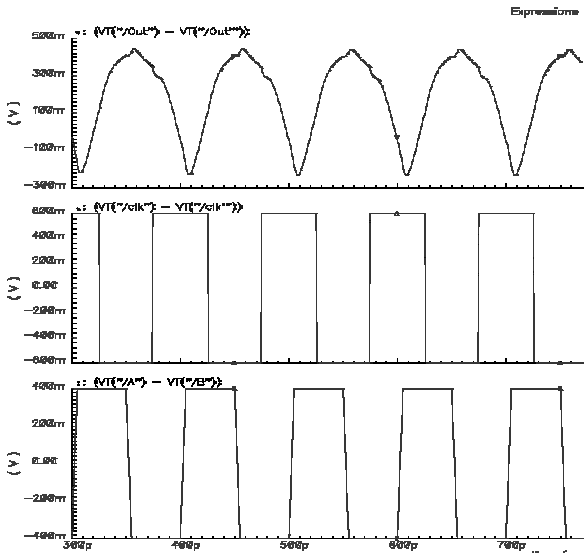

Fig. 11. The 20GHz new CML latch. (a) The input data at 20GHz. (b) The half-rate clock at 10GHz

References

[1] J. Rabaey, *Digital Integrated Circuits: A Design Perspective*, Prentice-Hall, 1996.

[2] B. Razavi, "Prospects of CMOS Technology for High-Speed Optical Communication Circuits," *IEEE J. Solid-State Circuits*, vol. 37, No. 9, pp. 1135-1145, Sept. 2002.

[3] M. Mizuno, M. Yamashina, K. Furuta, H. Igura, H. Abiko, K. Okabe, A. Ono, H. Yamada, "A GHz MOS adaptive pipeline technique using MOS current-mode logic," *IEEE J. Solid-State Circuits*, vol. 31, No. 6, pp. 784-791, June 1996.

[4] M. H. Anis, M. I. Elmasry, "Self-timed MOS current mode logic for digital applications," *IEEE Int'l Symp. on Circuits and Systems*, vol. 5, pp. 113-116, May 2002.

[5] B. Razavi, *Design of Analog CMOS Integrated Circuits*, pp. 101-134, McGraw-Hill, 2001.

[6] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998.

[7] N. Hedenstierna, K. O. Jeppson, "CMOS Circuit Speed and Buffer Optimization," *IEEE Trans. Computer-Aided Design*, vol. CAD-6, No. 2, pp. 270-281, March 1987.