A Monotonic Digitally Controlled Delay Element

Mohammad Maymandi-Nejad and Manoj Sachdev, Senior Member, IEEE

Abstract—A monotonic digitally controlled delay element (DCDE) is implemented in the 0.18 μm CMOS technology. In this paper, the design procedure of the new architecture and measurement results are reported. The delay of the DCDE changes monotonically with respect to the digital input vector. The monotonicity is one of the important features of this new architecture. Due to its monotonic behavior, the design of the DCDE is rather straightforward. The DCDE can be analyzed by a simple set of empirical equations with reasonable accuracy and can be made more tolerant to process, temperature, and supply voltage variations. The implemented delay element provides a delay resolution of as low as 2 ps and consumes 170 μW to 340 μW static power depending on the digital input vector.

Index Terms—CMOS integrated circuits, delay circuits, delay-locked loops, programmable delay, test.

I. INTRODUCTION

VARIABLE delay elements are often used to manipulate the rising or falling edges of the clock or other pulses in integrated circuits (ICs). In a variable delay element, the delay between the rising/falling edge of the output and that of the input can be varied. This delay can be changed by either analog or digital means. In analog means, an analog controlling voltage or current allows us to achieve desired delay [1]. On the other hand, discrete voltage [2] or capacitance [3] allows manipulation of the delay through digital means in digitally controlled delay elements (DCDEs).

DCDE have many applications in VLSI circuits. They are used in digital delay-locked loops (DLLs) [4], digital phase-locked loops (PLLs) [5], [6], digitally controlled oscillators (DCOs) [2], [7], and microprocessors and memory circuits [8], [9]. DCDEs are also finding applications in the testing of high-performance digital VLSI circuits. Recently, some new techniques have been proposed which help test and diagnose the dynamic behavior of high-performance digital circuits at a frequency much lower than the nominal operating frequency of the chip [10], [11]. In all the above-mentioned applications, the delay element is one of the crucial components and its precision directly affects the overall performance of the circuit.

There are several different architectures that have been used to implement a DCDE. In most of these architectures a switch network of nMOS transistors is placed at the source of the nMOS transistor (M1) in a CMOS inverter, as shown in Fig. 1. In this circuit only the delay of the falling edge of the output voltage can be controlled by the input vector. In order to control the delay of the output rising edge, another, similar switch network of pMOS transistors should be placed at the source of the pMOS transistor (M2) in the inverter. The number of nMOS transistors in the switch network depends on the desired number of different separate delays and the required delay resolution. Depending upon the digital input vector, the equivalent resistor of the switch network (or the current passing through it) changes and causes the delay of the inverter to change [6], [8].

One of the main drawbacks of these delay elements is that the delay of the circuit may not change monotonically with respect to the input vector. It makes the design of the circuit difficult, hence the circuit should be thoroughly simulated for all the possible combinations of the input vector. For example, in the case of the circuit used in [8], finding the sizes of the transistors in the switching network is a matter of optimal coding.

In this paper, we report a DCDE with a new architecture implemented in 0.18 μm bulk CMOS technology. The main feature of this new architecture is that the delay changes monotonically with respect to the input vector. This makes the design of the circuit straightforward. Moreover, the delay of the proposed DCDE can be analyzed with a set of simple empirical equations [12].

This paper is organized as follows. In Section II, we will briefly discuss two commonly used methods for implementing DCDEs, as well as the new architecture. The shortcomings of two different approaches will be highlighted and the main features of the new architecture will be discussed. In Section III, we outline the design steps taken in the designing of new DCDE. The measurement results of the fabricated circuit will be provided in Section IV. These results are compared with the simulation result and empirical equation. In Section V, it is explained how it is possible to make the behavior of the DCDE more tolerant to process, temperature, and supply voltage variations. Finally, conclusions will be presented in Section VI.

II. DESIGN TECHNIQUES OF VARIABLE DELAY ELEMENTS

There are several popular techniques for designing a variable delay element. Fig. 2 illustrates a DCDE based on the cur-
rent-starved delay element. As can be seen in this figure, there are two inverters between input and output of this circuit. The charging and discharging currents of the output capacitance \( (C_{L1}) \) of the first inverter, composed of \( M_1 \) and \( M_2 \), are controlled by two sets of current-controlling nMOS \( (M_{n0}, M_{n1}, \ldots) \) and pMOS \( (M_{p0}, M_{p1}, \ldots) \) transistors at the source of \( M_1 \) and \( M_2 \), respectively. The current controlling transistors are sized in a binary fashion, i.e., the \( W/L \) ratio of \( M_{n1} \) is twice that of \( M_{n0} \) and so on. It allows us to achieve binary incremental delays. As can be seen, by applying a specific binary vector to the controlling transistors, a combination of transistors is turned on at the sources of \( M_1 \) and \( M_2 \) transistors. Such an arrangement controls the rise time and fall time, and hence the delay, of the output voltage of the first inverter.

Fig. 3 illustrates another technique for implementing a DCDE. In this circuit, a variable resistor is used to control the delay. A stack of \( n \) rows by \( m \) columns of nMOS transistors is used to make the variable resistor. This resistor subsequently controls the delay of \( M_1 \). In the circuit of Fig. 3, only the rising edge of the Out can be changed with the input vector. Another stack of pMOS transistors can be used at the source of the pMOS transistor, \( M_2 \), to have control over the delay of the falling edge.

One of the problems with the above mentioned DCDE architectures is the nonmonotonic delay behavior with ascending binary input pattern. As can be seen in the circuits of Figs. 2 and 3, the input vector changes the effective resistance of transistor(s) placed at the source of the nMOS or pMOS transistors of the first inverter. This not only changes the resistance at the source of \( M_1 \) or \( M_2 \), but also changes the parasitic capacitance associated with transistors at these nodes. This is because the parasitic capacitance at the drain of a MOSFET is different in the ON and OFF states. Therefore, there are two factors which depend on the input vector and affect the delay:

i) The resistance of the controlling transistor: By increasing/decreasing the effective ON resistance of the controlling transistor(s) at the source of \( M_1 \) (\( M_2 \)), the circuit delay can be increased/decreased.

ii) The effective parasitic capacitance of the controlling transistor: As the effective capacitance of the controlling transistors at the source of \( M_1 \) (\( M_2 \)) increases due to the input vector, the charge sharing effect causes the capacitance at the output of the current-starved inverter to be (dis)charged faster and the overall delay of the circuit decreases.

Since the \( W/L \) ratio of the controlling transistors should change in binary fashion, often channel length, \( L \), is increased to realize small \( W/L \) ratios. A longer transistor puts a higher resistance and a larger parasitic capacitance at the source of \( M_1 \) (\( M_2 \)). A larger resistance increases the delay; however, a larger parasitic capacitance decreases the delay. Fig. 4 illustrates a simple current-starved inverter and the transient simulation results. The circuit depicted in Fig. 4(a) has two delay controlling transistors \( (M_{n0}, M_{n1}) \) with different \( W/L \) ratios placed at the source of the nMOS transistor \( M_1 \). At any time, at least one of the delay controlling transistors is on. Hence, a total of three different delays can be realized. The transient behavior of this delay element is shown in Fig. 4(b). As can be seen when transistor \( M_{n0} \) is conducting, the delay between the input and the output is larger compared to when transistor \( M_{n1} \) is conducting. Note that transistor \( M_{n0} \) has a larger \( W/L \) ratio which means that the delay should be lower if only the equivalent resistance of \( M_{n0} \) is taken into account. As mentioned above, the parasitic capacitance at the source of \( M_1 \) in Fig. 4(a) is also important. Fig. 5 shows the voltage at the source of \( M_1 \). As can be seen when transistor \( M_1 \) turns on, the \( C_{L1} \) charge shares with the \( C_P \). If transistor \( M_{n0} \) is on, the voltage drop at the source of \( M_1 \) is less compared to the

![Fig. 2. Current-starved delay element.](image)

![Fig. 3. Another delay element.](image)
voltage drop when transistor $M_{n1}$ is on. This is because the parasitic capacitance at the source of $M_1$ due to transistor $M_{n0}$ is less than the parasitic capacitance due to $M_{n1}$ in spite of $M_{n1}$ having smaller $W/L$ ratio of the two transistors. Therefore, even though transistor $M_{n1}$ has a smaller $W/L$ that should lead to a larger delay, the delay is smaller because of the initial discharge of $C_{L1}$ into $C_P$.

With the above discussion, it becomes apparent that monotonic delay behavior of the delay element cannot be ensured with ascending input vector. This situation is further complicated as the number of delay controlling transistors increases. Therefore, it becomes difficult to predict the circuit delay for a given input vector. Hence, during the design phase the circuit should be simulated for all the possible input combinations. The design of high-resolution delay element becomes a nontrivial task due to the lack of a one-to-one relationship between transistor sizes and corresponding delay. If, in the design phase, the desired delays are not met, it is not very clear whether the size of a transistor in the nMOS or pMOS network should be increased or decreased.

A new architecture, which eliminates the above-mentioned non-monotonic delay behavior, is proposed in [12]. Fig. 6 shows the new delay element. As can be seen in this figure, the delay of a current-starved inverter, $M_8$–$M_{11}$, is controlled by the current passing through $M_8$ and $M_{11}$. Transistor $M_8$ controls the fall time of the output of this inverter while $M_{11}$ controls the rise time. The current passing through $M_8$ is determined by $M_5$ and the controlling current ($I$) passing through it. Meanwhile, the current passing through $M_{11}$ is determined by the controlling current ($I$) and transistors $M_5$–$M_7$. The delay controlling pMOS transistors $M_1, M_2, M_3, \ldots$ should be sized in a binary fashion. The input vector turns these pMOS transistors on or off. In this way, the current passing through $M_5$ ($I$) will be determined by the input vector. This controlling current ($I$) will later be mirrored to $M_8$ and $M_{11}$ and controls the delay of the first inverter. Note that transistor $M_4$ is always on. This circuit can implement $2^N$ different delays where $N$ is the number of pMOS controlling transistors. It is clear that the parasitic capacitances at the source of $M_8$ and $M_{10}$ are the same for all the input vector combinations. Therefore, when the input vector changes, only the (dis)charging current of the first inverter changes, and the charge sharing remains the same. This causes the delay of the circuit to change monotonically with respect to the input vector, which makes the design of this circuit straightforward compared to the other delay elements. The design steps of this circuit will be discussed in the next section.

Another point which is worth mentioning is that both the rising and falling edge delays can be varied by this circuit. This has come at the expense of three more transistors ($M_6, M_7$, and $M_{11}$), while in the conventional delay elements, the number of added transistors for this purpose is more. Note that transistors $M_6, M_7$, and $M_{11}$ do not need to be very large, while the delay-controlling transistors in conventional delay elements are large and consume extra area due to their binary sizing scheme.

### III. DESIGN OF THE NEW DCDE

The circuit shown in Fig. 7 has been designed and fabricated. In this circuit, only the delay of the rising edge of the output, Out, can be changed. Compared to the circuit shown in Fig. 6, this circuit has four inverters, including the current-starved inverter between input, In, and output, Out. The two extra inverters help in driving a large capacitive load (2 pf). The design procedure of the proposed DCDE is explained in [12]. Salient design steps are as follows:

i) The sizes of transistors $M_{11}$ to $M_{16}$ are determined by the load capacitance (2 pf in this case). Transistor $M_8$ should

![Fig. 4. (a) Simple current-starved delay element with two digital inputs and (b) its output voltage.](image)

![Fig. 5. The voltage at the source of $M_1$ in Fig. 4 for three different input vectors.](image)
be much smaller than \(M_0\) such that the discharging current is controlled by \(M_8\). The ratio of transistors \(M_{10}\) and \(M_0\) should be \(\mu_{\text{el}}/\mu_{\text{hh}}\) where \(\mu_{\text{el}}\) and \(\mu_{\text{hh}}\) represent electron and hole mobilities. Transistor \(M_7\) can be the same size as \(M_8\), since these two transistors make a current mirror. However, these transistors may have different sizes to reduce the static power consumption of the DCDE, as explained in [12].

ii) The number of pMOS controlling transistors \((N)\) can be obtained from the desired number of different delays \((m)\) of DCDE such that \(m = 2^N\). Moreover, the circuit must contain one more pMOS transistor \((M_6)\) which is always on. In our case, we have selected 5 + 1 pMOS controlling transistors, which provide us 32 different delays.

iii) Assuming transistors \(M_1\) to \(M_5\) are not present, transistor \(M_6\) is sized to get the maximum desired delay (e.g., 2.42 ns in our case).

iv) After sizing \(M_6\), we put another pMOS transistor (e.g., \(M_0\)) in parallel to \(M_6\) to obtain the minimum desired delay (2.06 ns). Note that \(M_0\) is not shown in the figure since this transistor is subsequently fragmented into \(N\) (in our case) smaller transistors.

v) Transistor \(M_0\) is now fragmented into \(N = 5\) transistors, \(M_1\) to \(M_5\), in a binary fashion. That is,

\[
(W/L)_{M_i} = \frac{2^{i-1}}{2^5-1} (W/L)_{M_6}, \quad i = 1, \ldots, 5.
\]

Table I shows the size of all the transistors up to the second inverter, which are important in the delay of the circuit.

**IV. EXPERIMENTAL RESULTS AND DISCUSSION**

The DCDE shown in Fig. 7, with its transistor sizes illustrated in Table I, is implemented in 0.18 \(\mu\)m CMOS technology. Fig. 8 shows the die microphotograph. In this section, the experimental results of the fabricated circuit is reported. Moreover, the delays obtained from the empirical equations will be compared to measurement.

A. Experimental Results

The DCDE circuit occupies an area of approximately 100 \(\times\) 50 \(\mu\)m\(^2\) and consumes a static power of 170 \(\mu\)W to 340 \(\mu\)W (depending on the input vector) from a 1.8 V supply voltage.

Fig. 9 illustrates the measured delay and the controlling current \((I)\) of the experimental prototype circuit versus the input vector given in binary sequence. The solid curves in the graph depict the schematic simulation results. The dots represent the measured values of delay and the controlling current. The measured data shows the monotonic delay behavior of the proposed DCDE. For higher values of the input vector, the proposed DCDE is able to realize delay increments of as low as 2 ps. On the other hand, for smaller values of input vectors, these two
is also observed. The measured current is less than what it is expected from the post-layout simulation and experiment have.

However, we implemented A for the input vector m). Even though the controlling current obtained from schematic simulation, post-layout simulation, and measurement.

Fig. 10. Controlling current with respect to input vector obtained from schematic simulations, post-layout simulations, and measurement.

Moreover, the controlling current obtained from schematic simulation in general is greater than the controlling current obtained from post-layout simulation and experiment.

As can be seen from Fig. 9, there are discrepancies between the simulation and measured data. Firstly, the delay range of the measured data (320 ps) is smaller compared to the simulated range (359 ps). Hence, a delay error of approximately 10% exists between measurement and simulation. A corresponding error between measured and simulated range values of I is also observed. The measured I ranges from 95 to 188 µA, while the simulated values range from 91 to 219 µA. This shows that part of the delay discrepancy is caused by the corresponding error in the controlling current. Secondly, Fig. 9 also illustrates noticeable delay differences between measured and simulated values at certain input vectors. Two such vectors are shown in the figure (abcde = 11011, abcde = 01111). If we look at the measured delay and current curves, it becomes apparent that for input vector abcde = 11011, the current is less than what it should be, hence the corresponding delay is more than expected. This can be better observed in Fig. 10, where the controlling current obtained from schematic simulation, post-layout simulation, and experiment is more clearly shown. As can be seen, the controlling current obtained from schematic simulation increases uniformly with respect to the input vector, but the current obtained from post-layout simulation and experiment have a break at the input vector 11011. This discrepancy can be explained by looking at the sizes of controlling pMOS transistors shown in Table I. According to this table, the size of transistors M5 and M4 are 2.1/20 and 4.2/20, respectively, which means we have increased the width of M4 to make its W/L ratio twice that of M5. The W/L ratio of transistor M3 should be twice that of M4. It means that M3 should have a W/L ratio of 8.4/20 (µm). However, we implemented M3 (W/L = 4.2 µm/10 µm) by decreasing its length by half compared to that of M4 while keeping its width the same as M4. This was done to satisfy the DRC rule which states that no point in an n-well should be more than 5 µm away from the well contact. This is the cause of differences between the simulation and measurement for the input vector abcde = 11011. In order to verify this assumption, we consider the circuit shown in Fig. 11. In this figure, the current passing through transistor M1, with a W/L = 4.2/10 (µm), is compared with the current passing through transistor M2, with W/L = 8.4/20 (µm). Even though the W/L ratios of the two transistors are the same, the currents passing through them are not equal. To further verify this assumption, we changed the transistor M3 with W/L = 4.2/10 into a transistor with W/L = 8.4/20 (µm). The post-layout simulations show that with this modification the controlling current will increase to 108.4 µA instead of 107.4 µA for the input vector abcde = 11011. Inspecting Fig. 9, similar problem can be observed for the case of input vector abcde = 01111. This is the input vector which turns on transistor M3. This shows the importance of layout design. Hence, in order to achieve very accurate delays, the controlling pMOS transistors should be laid out carefully.

B. Empirical Equation Validation

One of the advantages of this DCDE is the existence of theoretical as well as simple empirical equations that can be used for the analysis of this delay element. The delay of the DCDE shown in Fig. 7 can be found from the following empirical equations [12]:

\[
t_d = t_{d0} + \frac{A_1}{V_g - V_{41}^2} \tag{5}
\]

\[
V_g = V_2 + A_2 \sqrt{I} \tag{6}
\]

\[
I = I_0 + \bar{u}I_1 + \bar{a}I_2 + \bar{r}I_3 + \bar{d}I_4 + \bar{e}I_5. \tag{7}
\]
In the above equations, $A_1, V_1, A_2, V_2,$ and $I_0$ to $I_5$ are constants. $I_0$ to $I_5$ are determined from the simulation (or measurement) of the circuit for six different input vectors. The details of how to obtain these parameters are discussed in [12]. The second term in (5) accounts for the delay of the first two inverters and $t_d(1)$ is the delay due to the rest of the inverters in the circuit, i.e., inverters three and four in Fig. 7.

Based on the measurement of the $V_g$ (gate voltage of $M_8$), the controlling current $I$, and the delay $t_d$ of the fabricated delay element, the values shown in Table II are obtained for the above-mentioned parameters. Fig. 12 shows the delay of the DCDE obtained from the empirical equation and measurement. As can be seen the two curves are in good agreement. The maximum error between the two curves is less than 2%.

V. ROBUSTNESS ENHANCEMENT

One of the main issues in the design of a delay element is the impact of supply voltage, temperature, and process variations on its delay. Compared to conventional delay element architectures, like the ones shown in Figs. 3 and 4, the proposed delay element can be made less sensitive to process and other environmental variations with a minor change.

As explained in previous sections, the delay of the proposed delay element is controlled by the current passing through the controlling transistor of the current-starved inverter (transistors $M_8$ and $M_{11}$ in Fig. 6). The delay of the delay element is controlled by the delay of the current-starved inverter. In order to keep the delay independent of process, supply voltage and temperature variations, the controlling current $I$ (in Fig. 6) should be independent of these parameters. In the proposed architecture, one may make the controlling current independent of the above mentioned parameters by using current sources instead of the controlling pMOS transistors ($M_1$ to $M_4$ in Fig. 6) as shown in Fig. 13. The circuit of Fig. 13, transistors $M_2$ to $M_4$, act as current sources and the 4-bit digital input $(abc1)$ controls the delay. The input vector turns on/off transistors $M_{2C}$ to $M_{4C}$, which are in series with the current source transistors ($M_2$ to $M_4$). The gate voltages of these current source transistors are controlled by transistor $M_1$ and the reference current $I_{ref}$. Comparing the circuit of Fig. 13 with that of Fig. 6, it is clear that the pMOS controlling transistors in Fig. 6 are replaced with current sources. This technique helps reduce the impact of supply voltage, temperature, and process variations. For the sake of simplicity, we call the circuit architecture shown in Fig. 6 “type 1” and the architecture shown in Fig. 13 “type 2”. Fig. 14 compares the delay of the type 1 and type 2 delay elements. In order to do a fair comparison, the DCDE in Fig. 13 is designed such that it provides exactly the same delay values as the type 1 delay element. Fig. 14(a) illustrates the delay of the two circuits for two different supply voltages (i.e., 1.8 V and 1.7 V). As can be seen, the delays of the two circuits are exactly identical for $V_{ddl} = 1.8$ V. When the supply voltage is reduced to 1.7 V, the delay of the two circuits has changed, however, the type 2 circuit is less sensitive to supply voltage variations. Fig. 14(b) compares the impact of temperature on the delay of the two circuits. As can be seen, the impact of temperature on the delay of type 2 DCDE is less compared to that of type 1. The impact of process variation is also examined on the behavior of the delay element. Fig. 15 shows the delay of the two types of the delay element for the input value of $abc1 = 0111$ at different process corners at room temperature. Clearly, the type 2 DCDE is less sensitive to process variations compared to type 1. These simulation

<table>
<thead>
<tr>
<th>$A_1$ (mV)$^2$</th>
<th>$A_2$ (mV/\mu A$^{1/2}$)</th>
<th>$V_1$ (mV)</th>
<th>$V_2$ (mV)</th>
<th>$I_0$ (\mu A)</th>
<th>$I_1$ (\mu A)</th>
<th>$I_2$ (\mu A)</th>
<th>$I_3$ (\mu A)</th>
<th>$I_4$ (\mu A)</th>
<th>$I_5$ (\mu A)</th>
<th>$t_d$ (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7086</td>
<td>35.4</td>
<td>-26.79</td>
<td>423</td>
<td>94.5</td>
<td>3.8</td>
<td>7.4</td>
<td>9.5</td>
<td>26.1</td>
<td>47.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>
In this paper, we reported a digitally controllable delay element (DCDE) using a new architecture. The DCDE is implemented in 0.18 µm CMOS technology. Compared to other popular methods for implementing DCDEs, the design of the new DCDE is simple and its delay changes monotonically with respect to the digital input vector. Moreover, its delay behavior can be described by a simple empirical equation [12]. However, this DCDE consumes 170 to 340 µW of static power, which is larger than the static power consumed by other DCDEs discussed above. For many applications where precise timing generation is required, such as delay fault testing [10], [11] and timing verniers [13], higher static power may be tolerated. The implemented DCDE can provide delay steps of as low as 2 ps. Moreover, the architecture of the proposed DCDE allows the designer to make the delay less sensitive to supply voltage, temperature, and process variations.

VI. CONCLUSION

results shows that the proposed delay element architecture can be made more tolerant to process and environmental variations. This is due to the circuit architecture which easily allows one to replace the controlling transistors with current sources. Needless to say, if cascode current sources are used instead of simple current sources, the robustness of the DCDE can be further increased.

REFERENCES

Mohammad Maymandi-Nejad received the B.Sc. degree from Ferdowsi University of Mashhad, Iran, in 1990, and the M.Sc. degree from Khajeh-Nassir Tossi University of Technology, Iran, in 1993, both in electrical engineering. He received the Ph.D. degree in electrical engineering in 2005 from the University of Waterloo, Canada.

From 1994 to 2001, he was an instructor with the Department of Electrical Engineering, Ferdowsi University of Mashhad. During this period, he was involved in teaching and research and he also conducted several industrial projects in the field of automation and computer interfacing. His field of interest includes low-voltage, low-power analog integrated circuits.

Manoj Sachdev (M’87–SM’97) received the B.E. degree (with Honors) in electronics and communication engineering from the University of Roorkee, India, and the Ph.D. degree from Brunel University, London, U.K.

He is a Professor in the Electrical and Computer Engineering Department at the University of Waterloo, Canada. He was with Semiconductor Complex Limited, Chandigarh, India, from 1984 to 1989, where he designed CMOS integrated circuits. From 1989 to 1992, he worked in the ASIC division of SGS-Thomson, Agrate (Milan), Italy. From 1992 to 1998, he was with Philips Research Laboratories, Eindhoven, The Netherlands, where he researched various aspects of VLSI testing and manufacturing. His research interests include low-power and high-performance digital circuit design, mixed-signal circuit design, test and manufacturing issues of integrated circuits. He has written a book and two book chapters on testing, and has contributed to more than 100 articles in conferences and journals. He holds more than 15 granted and several pending U.S. patents in the area of VLSI design and test.

Dr. Sachdev received the Best Paper Award at the European Design and Test Conference, 1997, and an honorable mention award for his paper at the International Test Conference, 1998.