ORIGINAL ARTICLE

A novel high-performance time-balanced wide fan-in CMOS circuit

CrossMark

Sherif M. Sharroush

Dept. of Electrical Engineering, Fac. of Engineering, Port Said, Port Said Univ., Egypt

Abstract   There is no doubt that static complementary CMOS logic is one of the most dominant logic-circuit families available. However, CMOS circuits with wide fan-in suffer from a relatively poor performance that is apparent in increased area, large time delay, and large power consumption. This is typically the case with CMOS circuits containing NMOS or PMOS stacks (i.e. branches containing a relatively large number of serially connected transistors). In this paper, a novel circuit that depends on applying the input signals in the form of pulses with a certain width will be presented as an alternative to stack circuits. The proposed scheme will be investigated quantitatively with the effect of the pulse width on the performance of the proposed scheme taken into account. The proposed scheme will be compared with the conventional CMOS logic from the points of view of area, high-to-low propagation delay, and average power consumption. The parameter variations and second-order effects will also be taken into account. Simulation results verify the correct operation of the proposed scheme and that the percentage reduction in the average propagation delay is 15.8% and 61.25% in cases of four and eight inputs, respectively, adopting the 45 nm CMOS technology with V_{DD} = 1 V.

1. Introduction

There are various and well known circuit families that can be used in realizing logic gates: the static complementary CMOS, the pass-transistor logic, the dynamic MOS logic, the dynamic-cascode voltage swing logic, the pseudo-NMOS logic, and the current-mode logic [1]. Among the various CMOS logic-circuit families, the static complementary CMOS proves the most important and robust one in realizing logic gates. However, its performance, like the other families, degrades with increas-

ing the number of the inputs [2]. The word "performance" here includes the area, the low-to-high and the high-to-low propagation delays, the power consumption, and the noise margin. So, ingenious circuit techniques must be adopted in order to improve the performance of such circuits or alternative circuits can instead be used.

Specifically, CMOS circuits such as NAND or NOR gates with wide fan-in contain branches with a large number of serially connected NMOS or PMOS transistors, respectively. As a result, their performance degrades. In this paper, a novel circuit that is based on applying the input signals in the form of pulses with a suitable width will be proposed as a wide fan-in NAND gate, hence the name "time balanced." The proposed circuit has a smaller area, time delay, and power-delay product compared

to the conventional CMOS logic when the number of the inputs exceeds a certain limit to be determined in this paper.

The remainder of this paper is organized as follows: Section 2 provides a quick survey of the previous solutions to the problem at hand (the degradation of the performance of the logic families with increasing the fan-in). The proposed solution is presented qualitatively in Section 3 with the quantitative analysis presented in Section 4. The impacts of second-order effects and process variations on the proposed scheme are discussed in Sections 5 and 6, respectively. The proposed scheme is verified by simulation in Section 7. Finally, the paper is concluded in Section 8.

2. Previous work

The previous work related to the problem at hand can be classified into three fronts: *reordering*, *resizing*, and *synthesis*. Reordering techniques can be achieved on the input and the transistor levels in order to reduce the power consumption in CMOS circuits. The reordering of the inputs does not modify the circuit schematic of the gate; however, it merely changes the order of the inputs. On the other hand, the reordering of the transistors modifies both the order of the inputs and the order at which the transistors are serially connected. Lowering the power consumption by reordering schemes, however, is usually associated with a delay penalty. This is because reordering usually causes movement of the inputs that arrive lately farther away from the output of the gate, thus resulting in an increase in the delay. In [3], Prasad and Roy proposed an algorithm for reordering the multi-pass transistors. In [4], an algorithm that includes transitions at the internal nodes of a complex CMOS gate to derive the optimal configuration was presented.

On the other side, some forms of transistor resizing techniques depend on resizing the transistors in the gate such that minimum power consumption is achieved, however, with no increase in the delay. These techniques depend on evaluating the delay in the several paths of the circuit and determining those paths with delays that are lower than that in the critical path (i.e. paths with a positive slack), then reducing the sizes of the transistors in these paths. The process repeats until either the slack becomes equal to zero or the transistors reach the minimum possible size.

Added to the resizing techniques is the following: The lowermost transistor is fabricated with the largest size with the aspect ratios of the upper NMOS transistors decreasing as we move away from bottom to top. Several sizing schemes including the linear and exponential sizing or a combination of the two [5] can be adopted. The effect of sizing on the performance of CMOS circuits was investigated in [6–11]. A combination of both the input reordering and the transistor resizing approaches was presented by Tan in [12]. Finally, some synthesis techniques that depend on creating novel circuits that have the same output as the conventional stack but with improved performance were presented in [13,14]. The scheme presented in this paper lies in the third category.

In the next section, the proposed scheme will be presented.

3. The proposed scheme

Refer to Fig. 1(a) for illustration of the proposed *n*-input NAND gate. First, the *dis* signal will be activated to turn on

M_N and thus discharge any remnant charge on C_L. Then, the *dis* signal will be deactivated and the inputs, A_1, A_2, ..., A_n, will be applied, thus turning on the related NMOS transistors, M_{N1}, M_{N2}, ..., and M_{Nn}. Assume that the input signals are in the form of pulses with a certain width, T. The charging current of C_L depends on the number of the activated inputs and the level to which C_L settles as is determined by both the number of the activated inputs and the pulse width, T. The main idea is simply as follows: T is chosen such that the voltage across C_L in case of $n - 1$ activated inputs, V_{n-1}, will be smaller than the threshold voltage of the inverter, V_{thinv}, and thus the inverter output will be at logic "1." On the other hand, if all the n inputs are activated, then the voltage across C_L, V_n, will be larger than V_{thinv} and thus the inverter output will be at logic "0." Of course, if the number of the activated inputs is smaller than $n - 1$, the voltage across C_L, V_{CL}, will be smaller than V_{thinv} and thus the inverter output will be at logic "1" as it must be. A buffer consisting of two cascaded inverters can be used at the output to obtain a full-swing output. Note that if the power-supply voltage feeding the circuit was applied as a pulse with a certain width and the NMOS transistors were kept always activated, the capacitor, C_L, would discharge to ground upon turning off this pulse through the activated NMOS transistors.

Alternatively, a sense amplifier can be used to latch the output data by comparing the voltage across C_L by a reference voltage, V_{ref}, which is ideally the arithmetic average of V_n and V_{n-1}, i.e.

$$V_{ref} = \frac{V_n + V_{n-1}}{2}. \tag{1}$$

Refer to Fig. 1(b) for the sense-amplifier based scheme. It must be noted that the reference voltage, V_{ref}, must be activated after the deactivation of the input signals. This is to ensure obtaining a correct output. If it were not for this synchronization, the output will be at logic "0" in all cases because initially the voltage across C_L is smaller than V_{ref}. If the number of the inputs is large such that the worst-case voltage difference across C_L, $V_n - V_{n-1}$, is smaller than the acceptable limit for reliable operation, then the scheme of Fig. 1(a) can be extended to any number of inputs using an OR gate as shown in Fig. 1(c). The OR gate can be implemented using a static CMOS NOR gate and an inverter. According to the circuit of Fig. 1(c), there is no need to use a buffer at the output of each *n*/2 inputs' stage as the static CMOS inverter after the two-input NOR gate provides a rail-to-rail output swing. The corresponding alternatives to PMOS stacks are shown in Fig. 2(a), (b), and (c), respectively.

4. Circuit design issues

In this section, the first version of the proposed scheme shown in Fig. 1(a) will be investigated quantitatively assuming that two cascaded inverters are added at the output to obtain a rail-to-rail swing. Obviously, the robustness of the proposed scheme depends to a large extent on the difference between the two values of the voltage, V_{CL}, in cases of all-activated and all-except-one activated inputs. This difference represents the smallest difference and thus this analysis represents the worst-case scenario. This difference also represents the valid range within which the threshold voltage of the first inverter

can be chosen. An expression for this range will be derived. An expression for the optimum value of the pulse width, T_{opt}, (at which the voltage difference is maximum) will also be derived along with the corresponding maximum value of the voltage difference. In addition, the allowable range of T will also be discussed. The proposed scheme will be compared with the
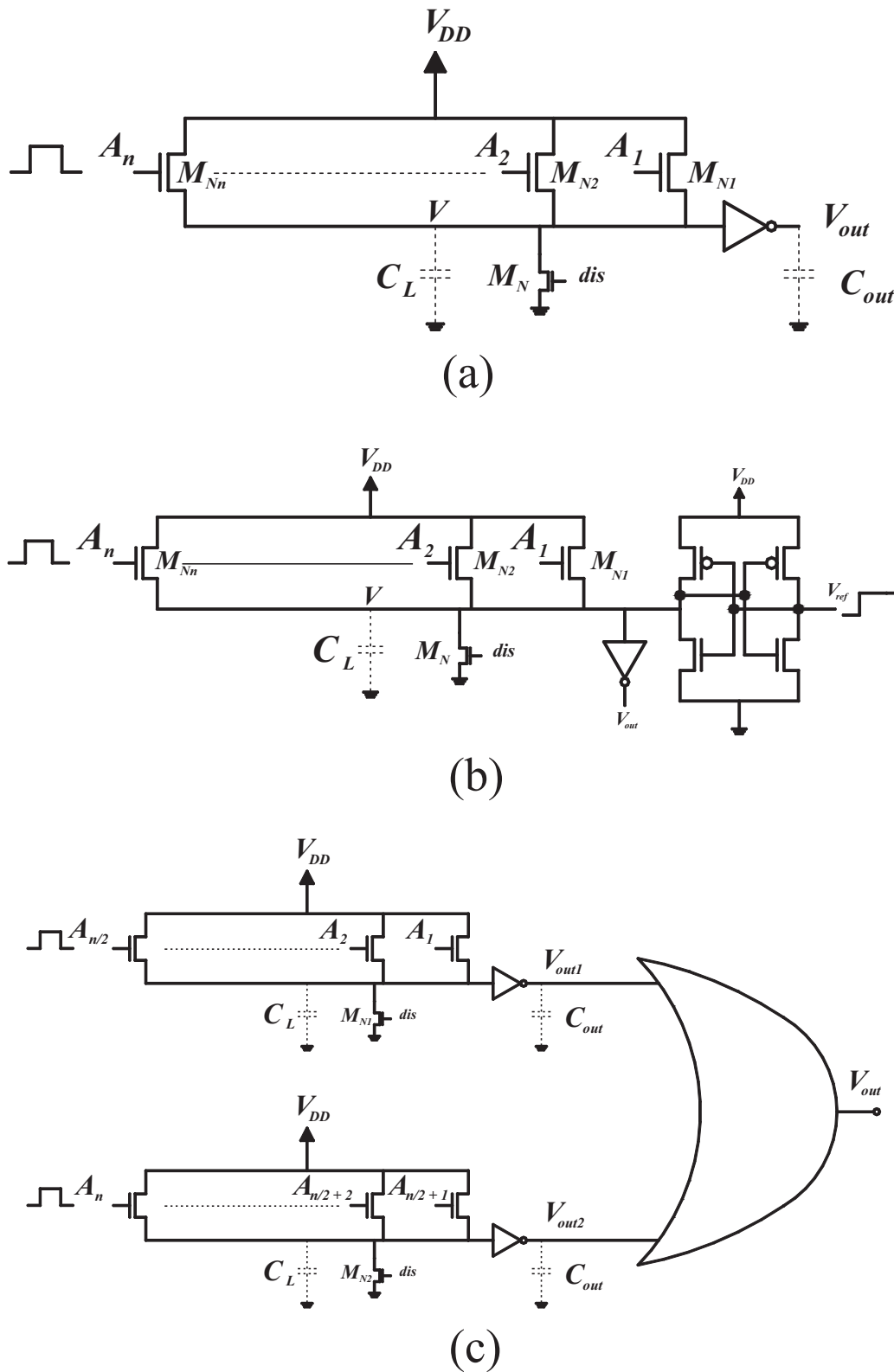

Figure 1 (a and b) The two proposed alternatives to NMOS stacks. (c) A circuit represents the extension of the scheme in (a) for any number of inputs using an OR gate.

conventional CMOS logic from the points of view of area, propagation delay, power consumption, and power-delay product.

In the following analysis, i_D, v_{GS}, v_{DS}, and v_{BS} represent the drain current, the gate-to-source voltage, the drain-to-source voltage, and the body-to-source voltage, respectively. The
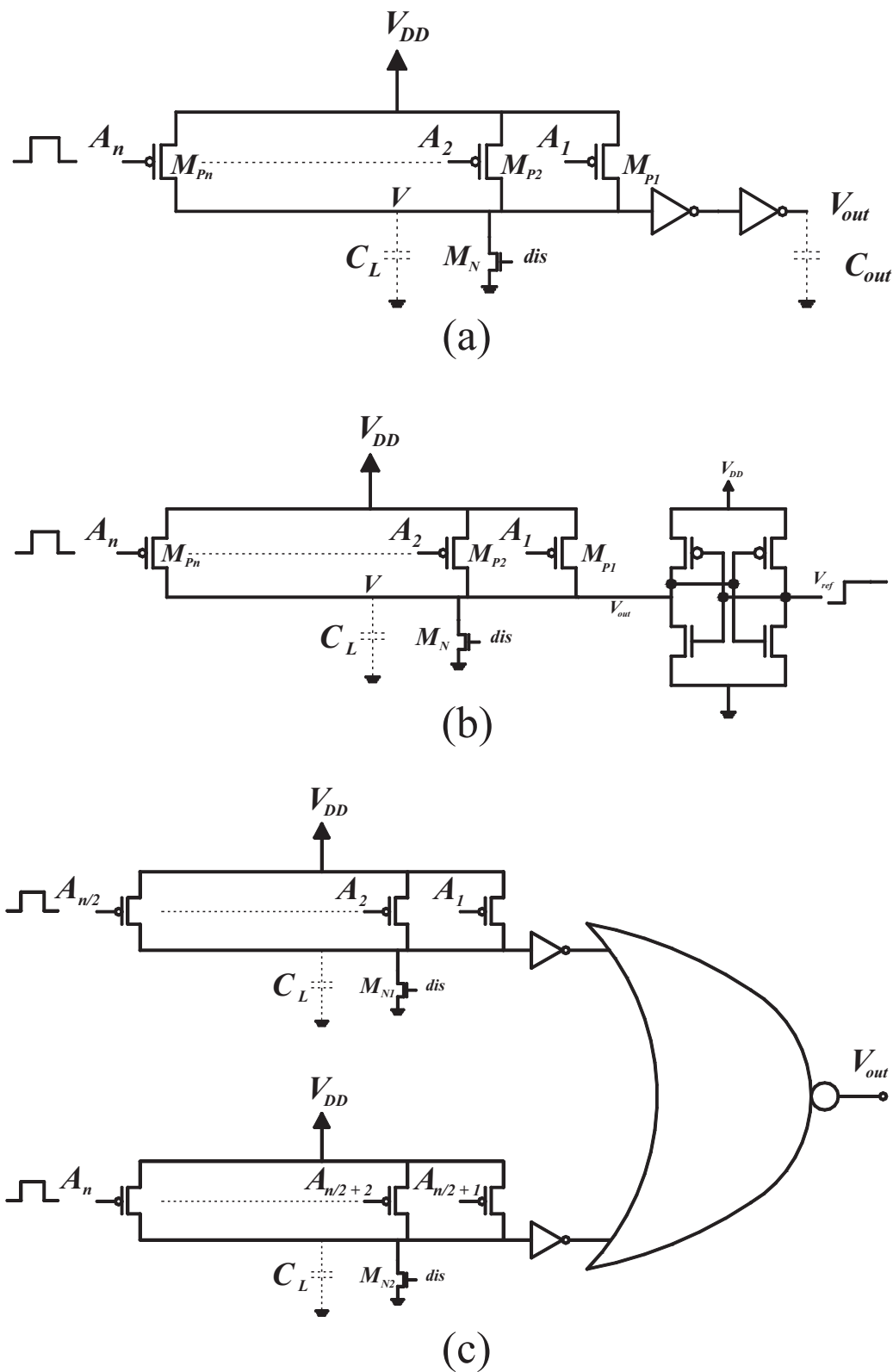

Figure 2 (a and b) The two proposed alternatives to PMOS stacks. (c) A circuit represents the extension of the scheme in (a) for any number of inputs using a NOR gate.

small letters represent the variables as functions of time while the capital ones represent certain values for these variables. Capital subscripts are used in the two cases. Unless otherwise specified, the following values will be adopted [15,16]: W (channel width) = L (channel length) = 45 nm, n (number of inputs) = 8, V_{DD} (power-supply voltage) = 1 V, V_{thn0} (threshold voltage of NMOS transistors at $V_{BS} = V_{DS} = 0$ V) = 0.25 V, V_{thp0} (threshold voltage of PMOS transistors at $V_{SB} = V_{SD} = 0$ V) = −0.32 V, k_n' (process-transconductance parameter of NMOS devices) = 638 µA/V^2, k_p' (process-transconductance parameter of PMOS devices) = 249 µA/V^2, γ (body-effect coefficient) = 0.4, λ_n (channel-length modulation effect parameter of NMOS devices) = λ_p (channel-length modulation effect parameter of PMOS devices) = 0.1 V^{-1}, α (a factor representing the short-channel effects) = 1.3, α_s (switching activity) = 1, C_{ox} (gate-oxide capacitance per unit area) = 0.0172 F/m^2, T (width of input pulses) = 15 ps, v_{sat} (free-electron saturation velocity) = 10^5 m/s [17], and f_s (frequency of switching) = 1 GHz.

Adopting the convention that the PMOS transistor has twice the area of the NMOS one to compensate for the mobility difference and assuming that the parasitic capacitance at each terminal is proportional to the aspect ratio of the associated transistors [18], then C_L can be expressed as $(4 + n)C$, where C is the parasitic capacitance associated with each terminal of a minimum-sized device and will be taken equal to 1 fF. In the following analysis, all the NMOS devices will be assumed minimum-sized while the PMOS ones have an aspect ratio equal to 2 unless otherwise specified.

4.1. Allowable range of V_{thinv}

In the following analysis, the short-channel MOSFET model will be adopted. According to this model, the $i_D - v_{GS} - v_{DS}$ relationship in the saturation region is [17]

$$i_D = WC_{ox}v_{sat}(v_{GS} - V_{thn})(1 + \lambda v_{DS}). \tag{2}$$

The saturation region according to this model occurs as long as $v_{DS} \geqslant v_{DSsat}$ where

$$v_{DSsat} = (1 - k)(v_{GS} - V_{thn}). \tag{3}$$

where k is a parameter that models the velocity-saturation effect. The value of k depends on the MOS technology and

increases with the overdrive voltage, though it may be regarded as constant under certain conditions [19]. For deep-submicron devices, k varies between 0 and 1 [20]. The threshold-voltage variation with the body effect will be approximated by the following relationship [18]:

$$V_{thn} = V_{thn0} - \gamma V_{BS}, \tag{4}$$

where V_{thn0} is the threshold voltage for zero source-to-body voltage and γ is the linearized body-effect coefficient (assuming that the source-to-body voltage of the transistors are small such that this effect can be linearized [18]). Taking into account that the body terminals of the NMOS transistors are connected to the most-negative terminal which is 0 V results in

$$V_{thn} = V_{thn0} + \gamma V_S. \tag{5}$$

Neglecting the channel-length modulation effect results in the following equation for the voltage, V_{CL}, across the parasitic capacitance, C_L, in case of n activated inputs:

$$nWC_{ox}v_{sat}(V_{DD} - V_{thn} - V_{CL}) = C_L \frac{dV_{CL}}{dt}. \tag{6}$$

Substituting for V_{thn} from Eq. (5) into Eq. (6) results in

$$nWC_{ox}v_{sat}[V_{DD} - V_{thn0} - (1 + \gamma)V_{CL}] = C_L \frac{dV_{CL}}{dt}. \tag{7}$$

Solving this equation by separation of variables and taking into account that C_L was treated here as initially discharged results in

$$V_{CL}(t) = \frac{(V_{DD} - V_{thn0})}{(1 + \gamma)}\left[1 - e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}t}{C_L}}\right]. \tag{8}$$

Assuming that the width of the input pulses is T, then in cases of all-activated inputs and all-except-one activated inputs, the two voltages across C_L (at $t = T$), V_n and V_{n-1}, will be

$$V_n = \frac{(V_{DD} - V_{thn0})}{(1 + \gamma)}\left[1 - e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}}\right] \tag{9}$$

and

$$V_{n-1} = \frac{(V_{DD} - V_{thn0})}{(1 + \gamma)}\left[1 - e^{\frac{-(1+\gamma)(n-1)WC_{ox}v_{sat}T}{C_L}}\right], \tag{10}$$
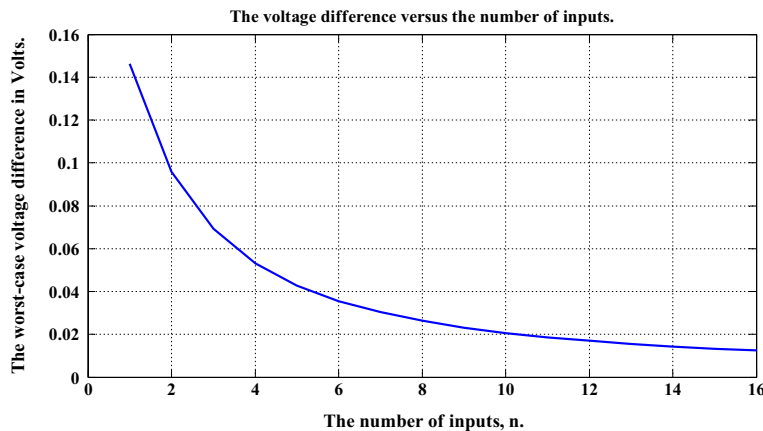
respectively. The difference between these two voltages is

Figure 3   The worst-case voltage difference at C_L versus n.

$$\Delta V = V_n - V_{n-1}$$

$$= \frac{(V_{DD} - V_{thn0})}{(1 + \gamma)} \left[e^{\frac{-(1+\gamma)(n-1)WC_{ox}v_{sat}T}{C_L}} - e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}} \right]. \qquad (11)$$

Refer to Fig. 3 for the plot of the worst-case voltage difference across C_L versus the number of the inputs which shows a

monotonic decrease as expected. This voltage difference represents the valid range within which the threshold voltage of the first inverter can be chosen. So, the larger this voltage difference, the more robust the scheme will be to the effect of the process variations. To determine the value of T at which ΔV is maximum, let it be T_{opt}, simply differentiate ΔV with respect
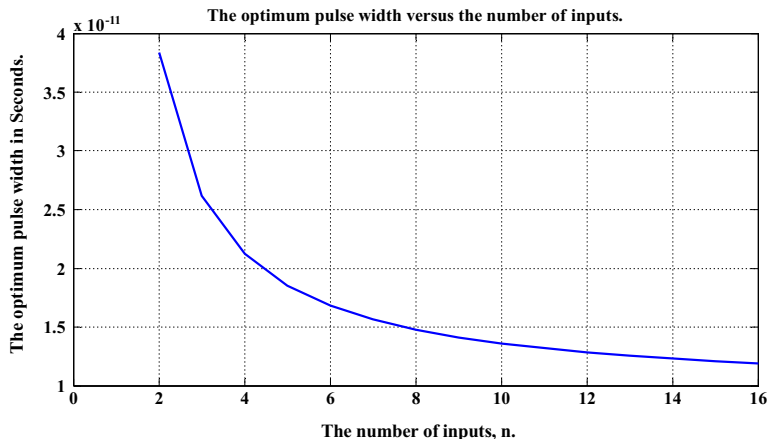

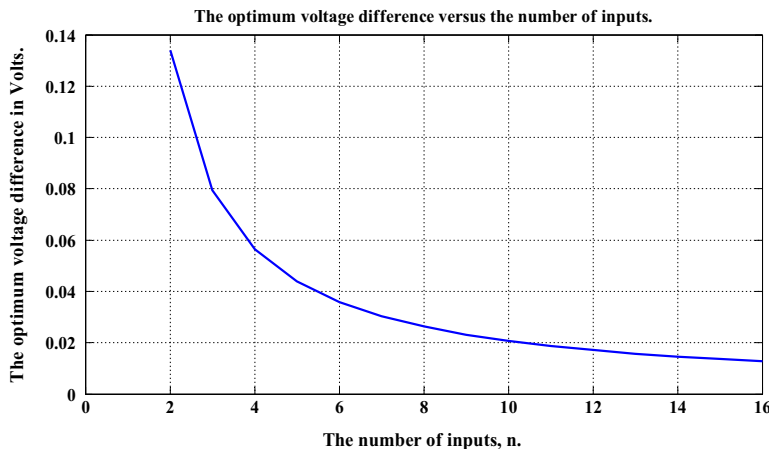**Figure 4**   The optimum pulse width versus n.

Figure 5   The optimum voltage difference versus n.
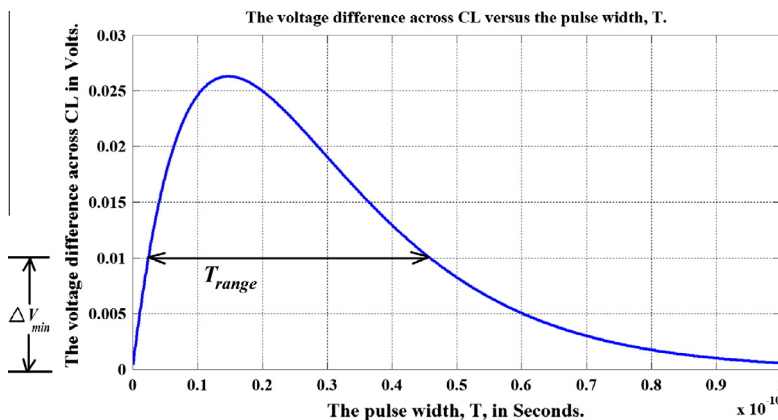

Figure 6   The relationship between the voltage difference, ΔV, and the pulse width, T.

to T and equate the first derivative to zero. So,

$$\frac{d(\Delta V)}{dt} = \frac{d}{dt}\left[\frac{(V_{DD} - V_{thn0})}{(1 + \gamma)}\left[e^{\frac{-(1+\gamma)(n-1)WC_{ox}v_{sat}T}{C_L}} - e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}}\right]\right]$$
$$= 0,$$

(12)

from which we obtain

$$T_{opt} = \frac{C_L \ln[n/(n-1)]}{WC_{ox}v_{sat}(1+\gamma)}.$$

(13)

Refer to Fig. 4 for the plot of T_{opt} versus n. As shown, T_{opt} decreases with increasing n as expected. The maximum value of ΔV, let it be $(\Delta V)_{max}$, can be determined by substituting for T_{opt} from Eq. (13) into Eq. (11). So,

$$(\Delta V)_{max} = \frac{(V_{DD} - V_{thn0})}{(1+\gamma)(n-1)}\left(\frac{n-1}{n}\right)^n.$$

(14)

Note that the maximum voltage difference does not depend on C_L. Refer to Fig. 5 for the plot of $(\Delta V)_{max}$ versus n which shows a monotonic decrease.

4.2. Allowable range of T

The plot of $\Delta V = V_n - V_{n-1}$ versus T is shown in Fig. 6. The optimum behavior of ΔV versus T can be expected from the plots of V_n and V_{n-1} along with the plot of ΔV versus time as shown in Figs. 7 and 8, respectively. In fact, the NMOS transistors charging C_L act as a variable resistor whose resistance depends on the number of the activated inputs. In case of all activated inputs, this resistance is smaller than that in case of all-except-one activated inputs. So, C_L charges faster in the first case but in the two cases, the steady-state voltages reach the same value which is independent of the number of the activated inputs. Hence, the voltage difference, ΔV, approaches 0 V at steady state as shown in Fig. 8.

Assuming that the minimum acceptable value for ΔV is $(\Delta V)_{min}$ such that the scheme still operates satisfactorily in spite of the effect of the process variations. Substituting by ΔV into Eq. (11) by $(\Delta V)_{min}$ results in

$$(\Delta V)_{min} = \frac{(V_{DD} - V_{thn0})}{(1+\gamma)}\left[e^{\frac{-(1+\gamma)(n-1)WC_{ox}v_{sat}T}{C_L}} - e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}}\right],$$
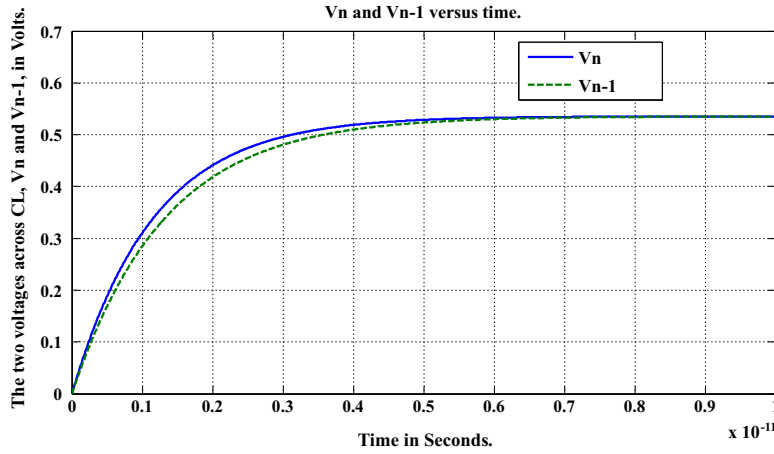
(15)

Figure 7   The two voltages across C_L, V_n and V_{n-1}, versus time.
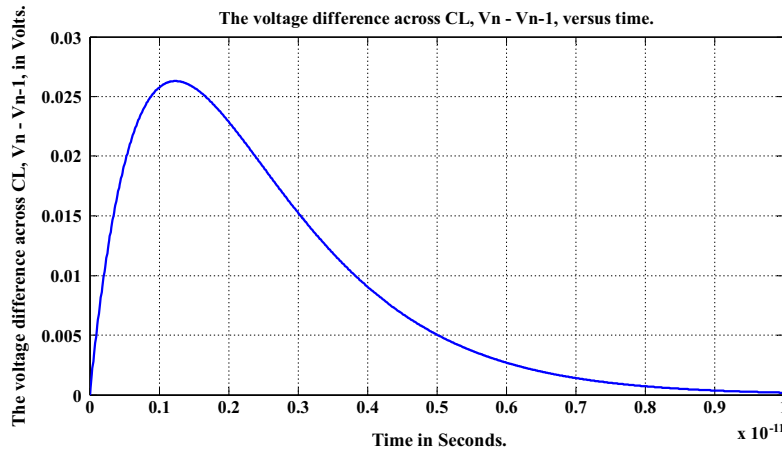

Figure 8   The voltage difference across C_L, V_n - V_{n-1}, versus time.
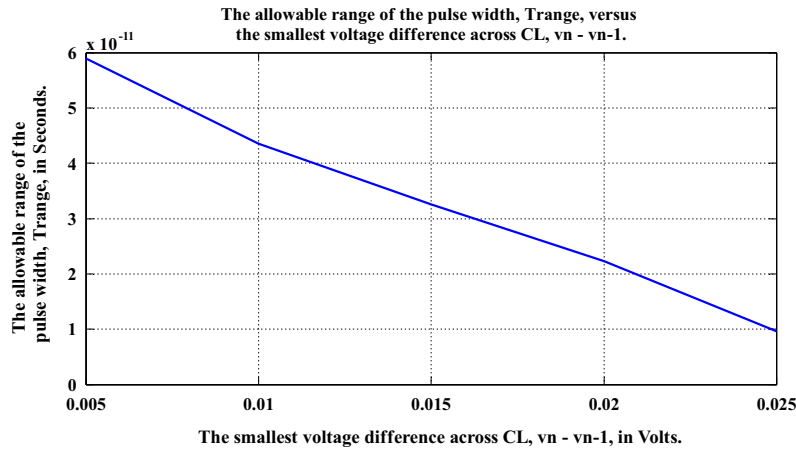
Figure 9   The allowable range for the pulse width versus the smallest voltage difference across C_L for $n = 8$.

which is a transcendental equation of T. So, T cannot be found explicitly in terms of $(\Delta V)_{min}$. However, as obvious from Fig. 6, as $(\Delta V)_{min}$ decreases, the allowable range of T, T_{range}, increases. The relationship between $(\Delta V)_{min}$ and T_{range} is shown in Fig. 9 for $n = 8$ (which is obtained point by point).

4.3. The logic swing of the first inverter

From the qualitative discussion of the proposed scheme in Section 3, it is obvious that the best robustness to process variations is achieved when the threshold voltage of the first inverter, V_{thinv}, lies midway between V_n and V_{n-1}, that is,

$$V_{thinv} = \frac{V_n + V_{n-1}}{2}. \tag{16}$$

Now, the logic swing at the output of the first inverter, LS, can be determined from the voltage-transfer characteristics (VTC) of this inverter which is shown qualitatively in Fig. 10.

It is obvious that the logic swing at the inverter output can be maximized by increasing the slope of the VTC in the transition region. The slope of the VTC in the transition region is given by [2]

$$slope = -(g_{mN} + g_{mP})(r_{oN}//r_{oP}),$$

where g_{mN} and g_{mP} represent the transconductances of the constituting NMOS and PMOS transistors of the first inverter
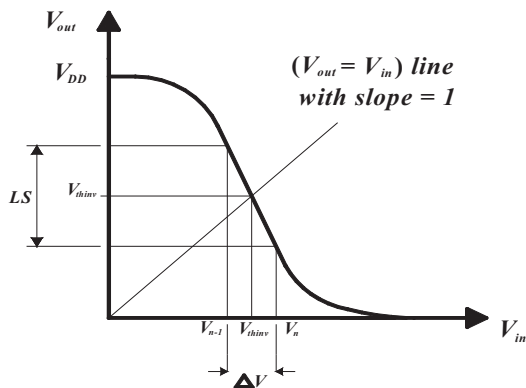
and r_{oN} and r_{oP} represent the output resistances of the constituting NMOS and PMOS transistors, respectively. The logic swing at the output of the first inverter, LS, is given by

$$LS = -(g_{mN} + g_{mP})(r_{oN}//r_{oP})\Delta V. \tag{17}$$

In order to increase LS, the two output resistances, r_{oN} and r_{oP}, must be increased which can be achieved by increasing the channel lengths of both transistors. In order to obtain the best performance, the channel width must also be increased which translates to a larger area. The parasitic capacitances associated with these transistors increase with increasing their dimensions. This is obviously a tradeoff between the robustness of the proposed scheme on one side and the area and speed on the other.

4.4. Area considerations

In comparing the areas of the proposed scheme with the conventional CMOS n-input NAND gate shown in Fig. 11, we will adopt the convention that the area of a certain transistor is equal
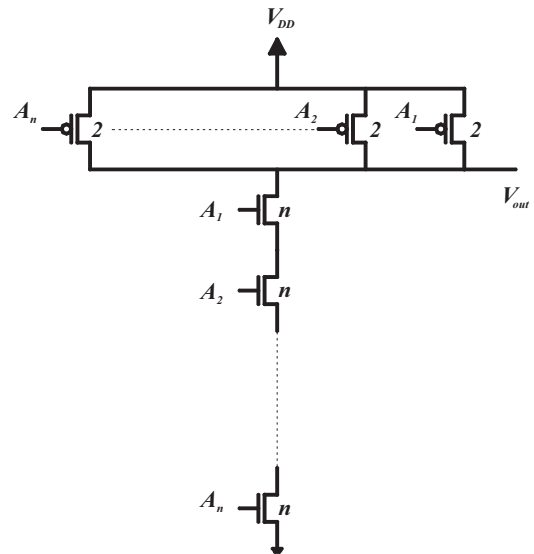

Figure 10   The voltage-transfer characteristics (VTC) of the first inverter. LS represents the logic swing at the inverter output.

Figure 11   The conventional CMOS n-input NAND gate with the sizing illustrated.

to its channel area [2]. Adopting the convention that the PMOS transistor has twice the area of the NMOS one to compensate for the mobility difference and adopting the conventional sizing strategy of increasing the aspect ratio of the transistors in the stack with n transistors by n in order to compensate for the delay increase [2], then the areas of the conventional and proposed schemes, A_c and A_p, can be approximated by

$$A_c = (n^2 + 2n)WL, \tag{18}$$

and

$$A_p = (10 + n)WL, \tag{19}$$

respectively. Refer to Fig. 12 for the plots of A_c and A_p versus n for $W = L = 45$ nm. It can be concluded from this rough estimation of the area that the proposed scheme has an area advantage when n exceeds 3.

4.5. High-to-low propagation delay

We will in this subsection compare the high-to-low propagation delays of the conventional and the proposed schemes, t_{PHLc} and t_{PHLp}, respectively. t_{PHLc} was derived in [21] for the case of $n + 1$ series-connected NMOS transistors and was found to be

$$t_{PHLc} = \frac{C_{outc}[V_{thn} + \alpha(V_{DD} - V_{thn})][1 + WC_{ox}R_{tota\ln}(1 + \gamma)v_{sat}]}{WC_{ox}(V_{DD} - V_{thn0})v_{sat}}$$
$$+ 2.3(n+1)R_1 C_{outc} \tag{20}$$

where $R_{tota\ln}$ is given by

$$R_{tota\ln} = \frac{n}{k'_n \left(\frac{W}{L}\right)\left(\frac{V_{DD}}{2} - V_{thn}\right)}, \tag{21}$$

and R is the equivalent resistance of each of the NMOS transistors. Modifying the last two equations to be valid for the case of n series-connected NMOS transistors results in the high-to-low propagation delay of the conventional stack with n inputs being given by

$$t_{PHLc} = \frac{C_{outc}[V_{thn} + \alpha(V_{DD} - V_{thn})][1 + WC_{ox}R_{tota\ln-1}(1 + \gamma)v_{sat}]}{WC_{ox}(V_{DD} - V_{thn})v_{sat}}$$
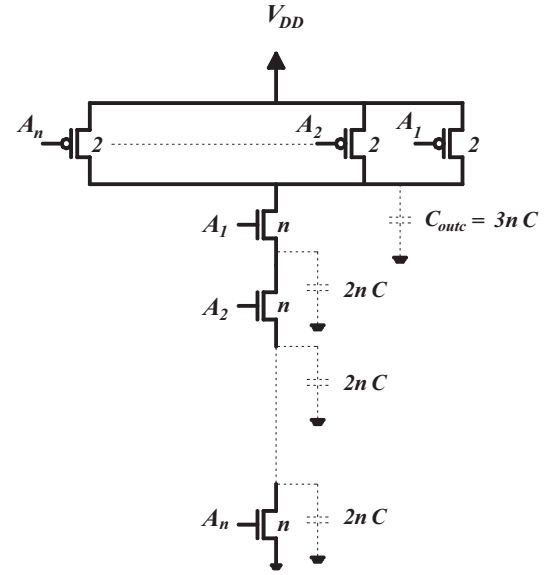$$+ 2.3nRC_{outc}, \tag{22}$$

Figure 13 The circuit diagram of the conventional CMOS n-input NAND gate with the sizing and the internal capacitances affecting the propagation delay illustrated.

where $R_{tota\ln-1}$ is the equivalent resistance of the $n-1$ lowermost transistors in the stack and is given by

$$R_{tota\ln-1} = \frac{n-1}{k'_n \left(\frac{W}{L}\right)\left(\frac{V_{DD}}{2} - V_{thn}\right)}. \tag{23}$$

In Eq. (20), C_{outc} is the parasitic capacitance at the output node of the conventional stack. When adopting the previously described convention for evaluating the parasitic capacitances and the conventional sizing strategy of multiplying the aspect ratio of the series-connected transistors by their number in order to get the same performance as the inverter, we get $C_{outc} = 3n$ fF (refer to Fig. 13).

For the scheme of Fig. 1(a), t_{PHLp} contains four subcomponents; the time required to charge C_L, $t_{PHLp1} = T$, the high-to-low propagation delay of the first inverter, t_{PHLp2}, the low-to-high propagation delay of the second inverter, t_{PHLp3}, and the high-to-low propagation delay of the third inverter, t_{PHLp4}. We have neglected the time required to initially discharge C_L from V_n to 0 V, which is very small compared to the other
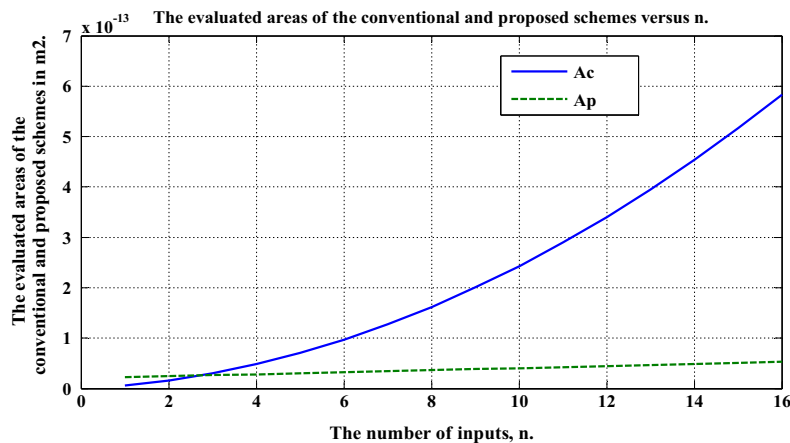

Figure 12 The relationship between A_c and A_p versus n.

subcomponents (we will return to this point in Section 7). The high-to-low and the low-to-high propagation delays of the inverter can be written as [22]

$$t_{PHL} = \frac{2C_{outp}}{k'_n \left(\frac{W}{L}\right)_n (V_{DD} - V_{thn})} \left[\frac{V_{thn}}{V_{DD} - V_{thn}} + \frac{1}{2} \ln \left(\frac{3V_{DD} - 4V_{thn}}{V_{DD}} \right) \right]$$ (24)

and

$$t_{PLH} = \frac{2C_{outp}}{k'_p \left(\frac{W}{L}\right)_p (V_{DD} - |V_{thp}|)} \left[\frac{|V_{thp}|}{V_{DD} - |V_{thp}|} + \frac{1}{2} \ln \left(\frac{3V_{DD} - 4|V_{thp}|}{V_{DD}} \right) \right],$$ (25)

respectively, where C_{outp} is the parasitic capacitance at the inverter output. So, the high-to-low propagation delay of the proposed scheme is

$$t_{PHLp} = T + \frac{2C_{outp1}}{k'_n \left(\frac{W}{L}\right)_n (V_{DD} - V_{thn})} \left[\frac{V_{thn}}{V_{DD} - V_{thn}} + \frac{1}{2} \ln \left(\frac{3V_{DD} - 4V_{thn}}{V_{DD}} \right) \right]$$
$$+ \frac{2C_{outp2}}{k'_p \left(\frac{W}{L}\right)_p (V_{DD} - |V_{thp}|)} \left[\frac{|V_{thp}|}{V_{DD} - |V_{thp}|} + \frac{1}{2} \ln \left(\frac{3V_{DD} - 4|V_{thp}|}{V_{DD}} \right) \right]$$
$$+ \frac{2C_{outp3}}{k'_n \left(\frac{W}{L}\right)_n (V_{DD} - V_{thn})} \left[\frac{V_{thn}}{V_{DD} - V_{thn}} + \frac{1}{2} \ln \left(\frac{3V_{DD} - 4V_{thn}}{V_{DD}} \right) \right],$$ (26)

where C_{outp1}, C_{outp2}, and C_{outp3} are the parasitic capacitances at the outputs of the three inverters in their order. Refer to Fig. 14 for the plot of t_{PHLc} versus n. The high-to-low propagation delay of the proposed scheme is 80.646 ps which is relatively independent of n. Thus, the proposed scheme is faster than the conventional stack for all values of n.

4.6. Average power consumption

In this subsection, the average power consumption of the conventional and proposed schemes will be compared for a circuit with n inputs. For the conventional CMOS n-input NAND gate, refer to Fig. 13 in which the sizing of each transistor and the parasitic capacitances at each node are shown.

In our estimation, the short-circuit and leakage power consumption components will be neglected for the conventional stack. So, the only component that will be taken into account is the dynamic-switching power consumption associated with charging the parasitic capacitances indicated to V_{DD}. The total dynamic-switching power consumption of an IC is given by [23]

$$P_{switching} = f_s V_{DD} \sum_{i=1}^{P} \alpha_{si} C_{Li} V_{swingi}$$ (27)

where P is the total number of nodes within a CMOS circuit, C_{Li} is the equivalent parasitic capacitance of the ith node, α_{si} is the switching activity of the ith node, and V_{swingi} is the voltage swing on the ith node assuming that the power supply providing the charge is V_{DD}. Now, for the stack of Fig. 13, $C_{outc} = 3nC$ charges in all the input combinations except the one in which all the inputs are activated which corresponds to all off PMOS devices. So, there are $2^n - 1$ input combinations that C_{outc} will charge in. For the upper parasitic capacitance with value $2nC$, it charges for all the input combinations which have A_1 equal to 1 except the case of all-activated inputs in which all the PMOS devices are off. The last number of input combinations is obviously $2^{n-1} - 1$. The same procedure can be applied for the lower parasitic capacitances with value $2nC$ including the lowermost one which charges only when all the input combinations are 1 except the lowermost one, A_n. The latter case occurs in only one input combination. Combining all these terms and dividing by the number of the input combinations, 2^n, results in the average power consumption of the conventional stack being given by

$$P_{avgc} = \frac{\alpha_s f_s V_{DD}^2}{2^n} [3nC(2^n - 1) + 2nC(2^{n-1} - 1) + 2nC(2^{n-2} - 1)$$
$$+ \ldots + 2nC(2^{n-(n-2)} - 1) + 2nC(2^{n-(n-1)} - 1)]$$

$$P_{avgc} = \frac{\alpha_s f_s V_{DD}^2}{2^n} [3nC(2^n - 1) + 2nC[2^{n-1} + 2^{n-2} + \ldots + 2^2 + 2^1$$
$$- (n - 1)]]$$

$$\therefore P_{avgc} = \frac{\alpha_s f_s V_{DD}^2}{2^n} [3nC(2^n - 1) + 2nC(2^n - n - 1)].$$ (28)

Now, for the proposed scheme of Fig. 1(a), the power consumption includes the dynamic-switching power consumption associated with charging C_L and the parasitic capacitances at the outputs of the three inverters, the dc power consumption of the first inverter, and the short-circuit power consumption associated with the three inverters. The voltage across C_L in case of n activated inputs is given by the following equation:

The high-to-low propagation delay of the conventional stack versus the number of inputs.
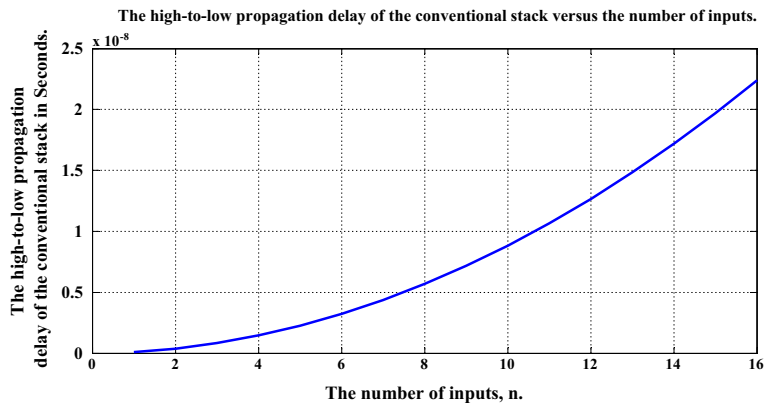
Figure 14   The plot of the high-to-low propagation delay of the conventional stack versus the number of inputs.

$$V_n = \frac{(V_{DD} - V_{thn0})}{(1+\gamma)}\left[1 - e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}}\right]. \tag{29}$$

In case of only one activated input, the voltage across C_L charges to V_1 which can be found by substituting n by 1 into Eq. (29). So, the associated switching power consumption is

$$P_{p1} = \alpha_s f_s C_L V_{DD} V_1.$$

The case of only one activated input occurs for n input combinations. Stated another way, the case of only one activated input occurs for a number of input combinations equal to the combination of n taken one at a time which is $_nC_1$. Similarly, when there are only two activated inputs, C_L charges to V_2 with an associated switching power consumption given by

$$P_{p2} = \alpha_s f_s C_L V_{DD} V_2. \tag{30}$$

The case of any two activated inputs occurs for a number of input combinations equal to $_nC_2$ which is the combination of n taken 2 at a time as the order of the activated inputs is not important [24]. $_nC_m$ is given by [24] (where n is a positive integer and m is a nonnegative integer)

$$_nC_m = \frac{n!}{m!(n-m)!}. \tag{31}$$

Repeating this procedure with the other voltages results in the average switching-power consumption of the proposed scheme being given by

$$P_{CL} = \frac{\alpha_s f_s C_L V_{DD}}{2^n}[(0)_nC_0 + (V_1)_nC_1 + (V_2)_nC_2 + \ldots$$
$$+ (V_{n-1})_nC_{n-1} + (V_n)_nC_n]. \tag{32}$$

The first and last two terms correspond to the two cases of no and all-activated inputs, respectively. The average switching-power consumption associated with the three inverters is

$$P_{p3} = \alpha_s f_s V_{DD}^2 (C_{outp1} + C_{outp2} + C_{outp3}). \tag{33}$$

Note that Eq. (33) overestimates the switching-power consumption of the first inverter as its output is not rail-to-rail swing. Concerning the dc power consumption of the first inverter, P_{DC}, it does not flow for all of the input combinations; rather, it flows when both the NMOS and PMOS devices of the first inverter become activated. However, we will simplify

its estimation and assume that it flows in all the input combinations. Also, a value of $V_{DD}/2$ is assumed at its input. If this inverter is assumed to be matched, then its output will also be at $V_{DD}/2$. For these two reasons, the value of the dc power consumption is overestimated. P_{DC} can be written as

$$P_{DC} = \frac{V_{DD}}{2}k'_n\left(\frac{W}{L}\right)_n\left(\frac{V_{DD}}{2} - V_{thn}\right)^2\left(1 + \lambda\frac{V_{DD}}{2}\right), \tag{34}$$

where the NMOS transistor of the first inverter is certainly in saturation as its v_{GS} and v_{DS} are both equal to $V_{DD}/2$. Now, the short-circuit power consumption of an inverter is given by [1]

$$P_{sc} = \frac{\alpha_s K \tau f_s (V_{DD} - 2V_{thn})^3}{12}, \tag{35}$$

assuming a matched inverter where τ is the rise or fall time (assuming that they are equal) of the input waveform, K is the device-transconductance parameter. Multiplying P_{sc} of Eq. (35) by 3 to account for the short-circuit power consumption of the three inverters and adding the result to that in Eqs. (32)–(34) result in the average power consumption of the proposed scheme being equal to

$$P_{avgp} = \frac{\alpha_s f_s C_L V_{DD}}{2^n}[(0)_nC_0 + (V_1)_nC_1 + (V_2)_nC_2 + \ldots$$
$$+ (V_{n-1})_nC_{n-1} + (V_n)_nC_n] + \alpha_s f_s V_{DD}^2(C_{outp1}$$
$$+ C_{outp2} + C_{outp3}) + \frac{\alpha_s K \tau f_s (V_{DD} - 2V_{thn})^3}{4}$$
$$+ \frac{V_{DD}}{2}k'_n\left(\frac{W}{L}\right)_n\left(\frac{V_{DD}}{2} - V_{thn}\right)^2\left(1 + \lambda\frac{V_{DD}}{2}\right) \tag{36}$$

An important note is in order here. Due to the need to reduce the dynamic-switching power consumption, V_{DD} reduces. However, in order not to degrade the performance, V_{thn} also reduces but the ratio, V_{DD}/V_{thn}, reduces with technology scaling [21]. The expected result is that the short-circuit power consumption reduces with technology scaling. It seems to be a power advantage for the proposed scheme. Refer to Fig. 15 for the plots of P_{avgc} and P_{avgp} versus n for $\tau = 1$ ns. It is obvious from this figure that P_{avgp} is larger than P_{avgc} for practical values of n. This can be attributed to the static-dc as well as the short-circuit power consumption. It is also
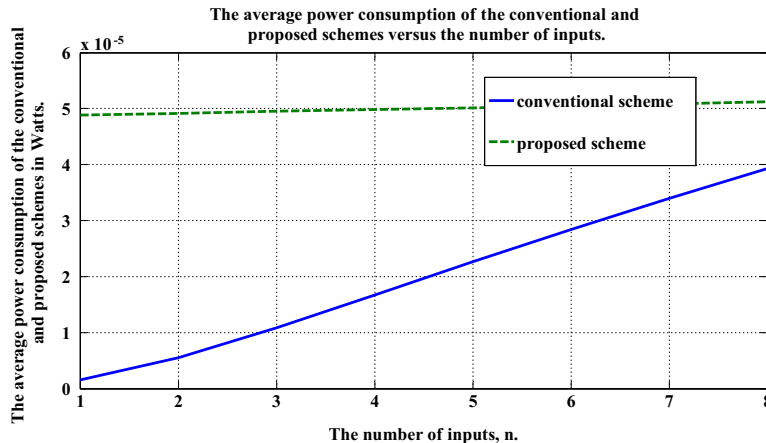

Figure 15 The plots of P_{avgc} and P_{avgp} versus n.

obvious that P_{avgc} and P_{avgp} show a monotonic increase versus n. Finally, refer to Fig. 16 for the plots of the power-delay products of the conventional and proposed schemes versus n which shows the superiority of the proposed scheme when n exceeds 2.

5. Second-order effects

In this section, the second-order effects will be taken into account. Among these effects are the channel-length modulation, body-effect, drain-induced barrier lowering (*DIBL*), short-channel effects, and narrow-channel effects.

5.1. The channel-length modulation effect

If this effect were taken into account, the drain-current equation of the MOS transistors in the input paths would be modified to include the term $(1 + \lambda_n v_{DS})$. The result, of course, is to increase the drain current for the same terminal voltages. However, the voltage across C_L upon charging will be (according to the simulation results) around 350 mV for four inputs. Using a power-supply voltage of 1 V results in a drain-to-source voltage of 0.65 V. Adopting $\lambda_n = 0.1\ \text{V}^{-1}$ results in a term of value

between 1.1 (when C_L is initially discharged and thus $V_{DS} = V_{DD} = 1\ \text{V}$) and 1.065 (when C_L is charged to 0.35 V and thus $V_{DS} = 0.65\ \text{V}$) multiplied by the adopted drain-current equation. Thus, this effect can safely be neglected with no loss of good accuracy.

5.2. The body effect

This effect was already taken into account in the analysis performed in Section 4. However, note that the body-effect coefficient, γ, is a fabrication-process parameter and is given by [22]

$$\gamma = \frac{\sqrt{2qN_A\varepsilon_s}}{C_{ox}} \tag{37}$$

where q is the electronic charge, N_A is the doping concentration of the p-type substrate, ε_s is the electric permittivity of silicon (1.04×10^{-12} F/cm), and C_{ox} is the gate-oxide capacitance per unit area. It is shown in [25] that in order for the MOSFET transistor device to operate properly in spite of CMOS technology scaling, the doping of the substrate, N_A, must be increased. However, the gate-oxide thickness, t_{ox}, decreases in order to reduce short-channel effects [23] with
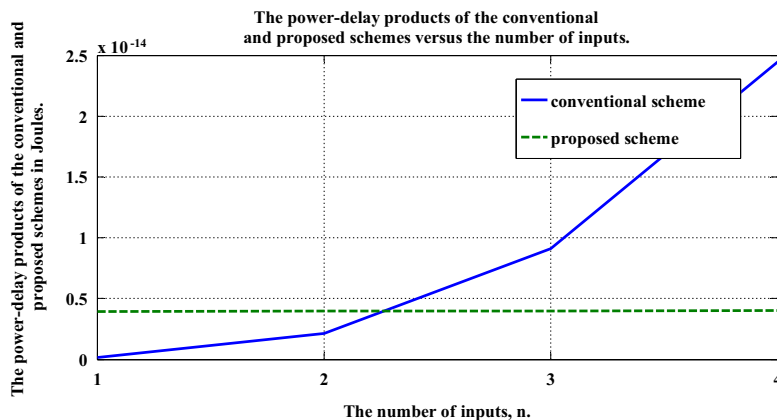

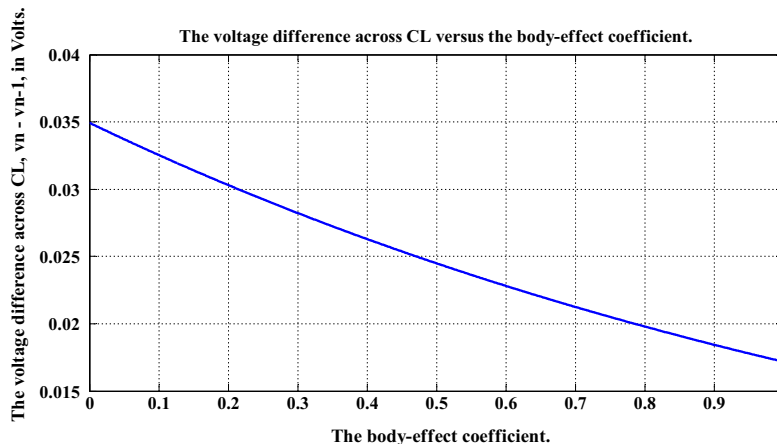**Figure 16**   The plots of the power-delay products of the conventional and proposed schemes versus n.

Figure 17   The voltage difference across C_L versus the body-effect coefficient.

the result that the gate-oxide capacitance per unit area, C_{ox}, increases. The increase in C_{ox} more than compensates for the increase of N_A and the net result is that the body-effect parameter, γ, decreases with technology scaling. The effect of the weakening of the body effect on the threshold voltage can be shown in Fig. 17 for the plot of the voltage difference, $\Delta V = V_n - V_{n-1}$, versus γ. It is apparent that the voltage difference increases with decreasing γ which seems to be an advantage gained with technology scaling.

5.3. The drain-induced barrier lowering (DIBL) effect

As the magnitude of the reverse bias voltage across the drain-to-body pn junction is increased, the depth of the junction depletion layer increases. A deeper depletion layer around the drain contributes a larger amount of depletion charge to the channel. An increased drain-to-body reverse bias voltage, therefore, enhances the short-channel effects and lowers the magnitude of the threshold voltage of the MOSFET transistor. The threshold-voltage degradation caused by an increased or decreased drain bias voltage of an N-channel or P-channel MOSFET, respectively, is commonly referred to as drain-induced barrier-lowering (DIBL) [26]. The threshold-voltage variation due to the DIBL effect can be expressed as

$$V_{thn} = V_{thn0} - \eta V_{DS}, \tag{38}$$

where V_{thn0} is the value of the threshold voltage at $V_{DS} = V_{BS} = 0$ and η is the DIBL coefficient. If this expression for V_{thn} were adopted in the previous analysis, the voltage, $V_{CL}(t)$, can be expressed as

$$V_{CL}(t) = \frac{[V_{DD}(1+\eta) - V_{thn0}]}{(1+\gamma+\eta)}\left[1 - e^{\frac{-(1+\gamma+\eta)nWC_{ox}v_{sat}t}{C_L}}\right]. \tag{39}$$

Refer to Fig. 18 for the relationship between the voltage difference across C_L and η. η is typically on the order of 0.1. For this value of η, the percentage variation of ΔV is 5.7%.

5.4. The short-channel effects

As the channel length of a MOSFET is reduced with technology scaling, the depletion regions around the source and drain terminals become closer with the result that the total depth of the source and drain depletion regions becomes comparable to the effective channel length in deep-submicrometer devices [23]. Thus, more charge is contributed to the depletion region beneath the gate area by the source-to-substrate and the drain-to-substrate depletion layers in short-channel devices. The threshold voltage thus lowers with decreasing gate length in what is known as V_{thn}-roll-off [27,28]. Refer to Fig. 19 for the relationship between ΔV and V_{thn0}. ΔV changes by a percentage of 6.13% for V_{thn0} ranging between 0.2 and 0.25 V.

5.5. The narrow-channel effects

Due to the reduction of the channel width, more gate charge is required to invert the channel because a larger percentage of the gate-induced space charge is lost in fringing fields [29]. This results in monotonically increasing V_{thn0} with decreasing the channel width. It seems that the increase of V_{thn0} somewhat compensates for the reduction associated with decreasing the channel length.

6. Effect of process variations

In this section, the effect of the process variations on the voltage difference at the input of the first inverter will be investigated quantitatively. The effect of the variations of each of V_{thn0} (the threshold voltage of the charging NMOS devices), W (the channel width of the charging NMOS devices), and T will be investigated one at a time. Assume that the variation of V_{thn} is ΔV_{thn}. So, substituting V_{thn} in Eq. (11) by $V_{thn} + \Delta V_{thn}$ results in the voltage difference being given by

$$\begin{aligned}\Delta V &= V_n - V_{n-1} \\ &= \frac{(V_{DD} - V_{thn0} - \Delta V_{thn0})}{(1+\gamma)}\left[e^{\frac{-(1+\gamma)(n-1)WC_{ox}v_{sat}T}{C_L}} - e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}}\right].\end{aligned} \tag{40}$$
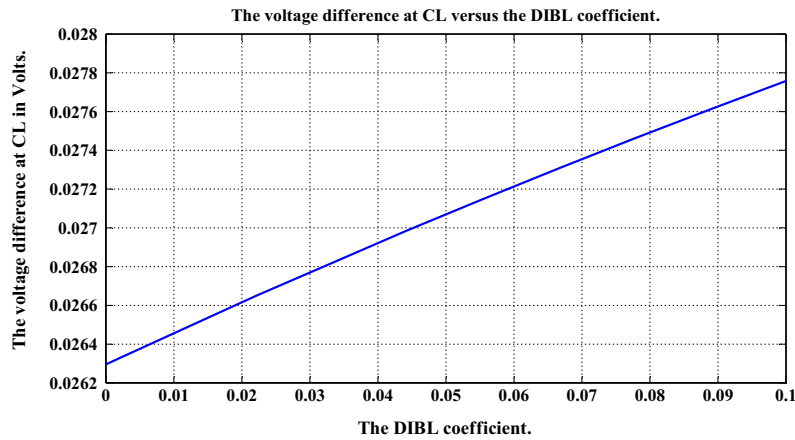
The variation in ΔV due to that in V_{thn} is thus

$$\Delta(\Delta V)_{\Delta V thn} = \frac{-\Delta V_{thn0}}{(1+\gamma)}\left[e^{\frac{-(1+\gamma)(n-1)WC_{ox}v_{sat}T}{C_L}} - e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}}\right]. \tag{41}$$

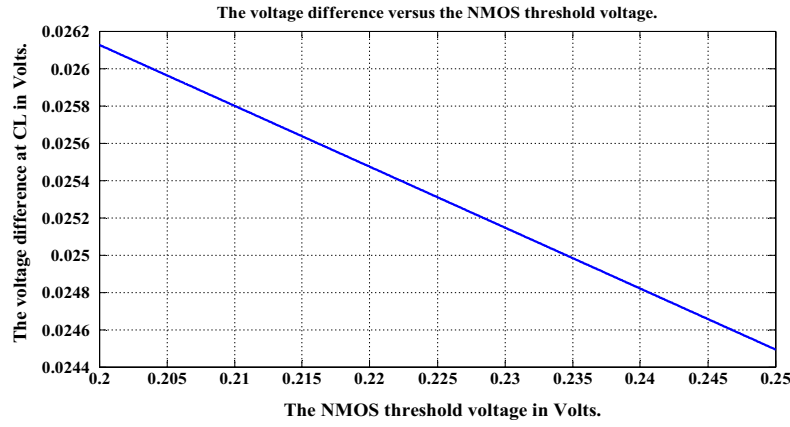**Figure 18** The voltage difference across C_L versus the DIBL coefficient, η.

Figure 19 The relationship between ΔV and V_{thn0}.

Repeating the analysis with respect to the variations in W and T and assuming that the variations of these two parameters, ΔW and ΔT, are small, then the following approximation can be used (where a is a constant):

$$e^{-a(x+\Delta x)} = e^{-ax}(1 - a\Delta x). \tag{42}$$

The variations in ΔV due to the variations in these two parameters are thus given by

$$\Delta(\Delta V)_{\Delta W} = \frac{\Delta W(V_{DD} - V_{thn0})}{(1+\gamma)} \left[-e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}} \left(\frac{(1+\gamma)nC_{ox}v_{sat}T}{C_L} \right) \right.$$
$$\left. -e^{\frac{-(1+\gamma)(n-1)WC_{ox}v_{sat}T}{C_L}} \left(\frac{(1+\gamma)(n-1)C_{ox}v_{sat}T}{C_L} \right) \right], \tag{43}$$

and

$$\Delta(\Delta V)_{\Delta T} = \frac{\Delta T(V_{DD} - V_{thn0})}{(1+\gamma)} \left[-e^{\frac{-(1+\gamma)nWC_{ox}v_{sat}T}{C_L}} \left(\frac{(1+\gamma)nWC_{ox}v_{sat}}{C_L} \right) \right.$$
$$\left. -e^{\frac{-(1+\gamma)(n-1)WC_{ox}v_{sat}T}{C_L}} \left(\frac{(1+\gamma)(n-1)WC_{ox}v_{sat}}{C_L} \right) \right], \tag{44}$$

respectively.

Assuming that the variations in V_{thn}, W, and T, are uncorrelated, then the total variation of ΔV, $\Delta(\Delta V)$, can be expressed as the sum of the products of each of the sensitivities of ΔV by the corresponding change [30]. Thus, we get

$$\Delta(\Delta V) = \Delta V_{thn}\left(\frac{\partial \Delta V}{\partial V_{thn}}\right) + \Delta W\left(\frac{\partial \Delta V}{\partial W}\right) + \Delta T\left(\frac{\partial \Delta V}{\partial T}\right). \tag{45}$$

Refer to Figs. 20–22 for the absolute variations of ΔV versus the percentage variations in V_{thn}, W, and T, respectively. Figs. 21 and 22 show the plots of the absolute variations of ΔV due to that in W and T evaluated approximately and exactly (without the approximation of Eq. (42)). It is obvious from these figures that the absolute variations in ΔV corresponding to a 100% variation in V_{thn}, W, and T, are 8.8 mV, 6.8 mV, and 6.8 mV, respectively. The total absolute variation of ΔV is only 7.2 mV for a 20% variation in each of V_{thn}, W, and T.

7. Simulation results

The proposed scheme will be verified in this section by simulation using the 45 nm CMOS technology with $V_{DD} = 1$ V for four and eight inputs. The aspect ratios of all the transistors are taken equal to 2. The 50% point convention will be adopted in evaluating the propagation delays. According to the simulation, the time required to discharge C_L from V_n to 0 V in case of four inputs is 20.8 ps. This represents only 3.8% of the worst-case propagation delay and thus can safely be neglected as stated in Section 4. Refer to Fig. 23 for the plot of the worst-case voltage difference versus time. According to
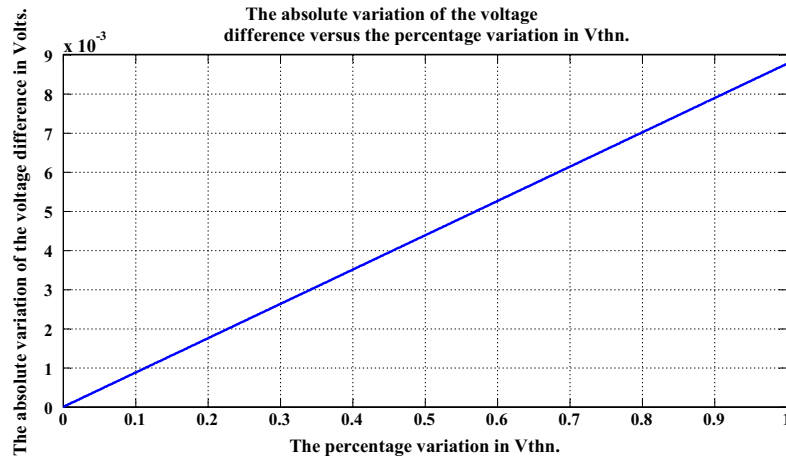

Figure 20 The absolute value of the variation of the voltage difference versus the percentage variation in V_{thn}.
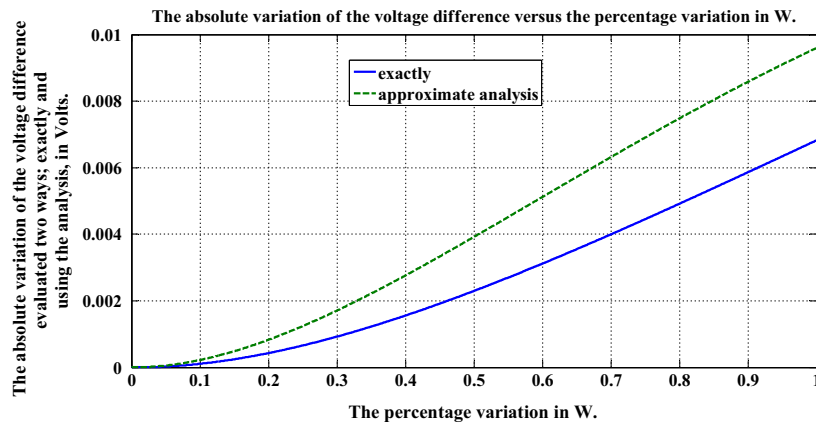
Figure 21 The absolute value of the variation of the voltage difference versus the percentage variation in W.
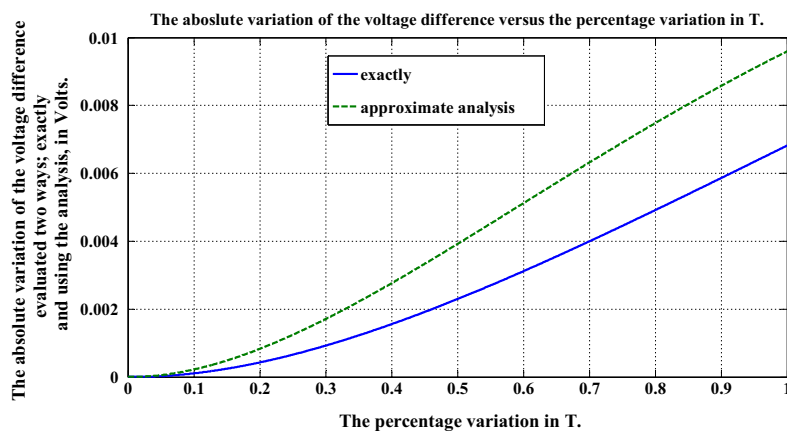

Figure 22 The absolute value of the variation of the voltage difference versus the percentage variation in T.
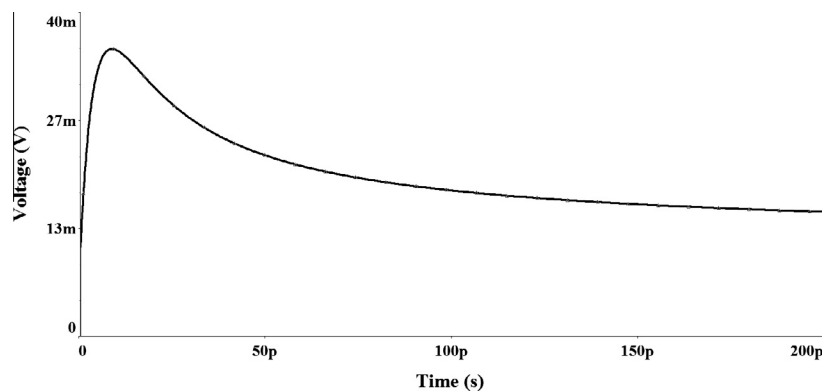

Figure 23 The simulation results illustrating the worst-case voltage difference across C_L versus time.

this figure, the maximum value of this difference is 35.5 mV and occurs at a time of 7.8 ps. Figs. 24 and 25 show the simulation results of the conventional and proposed schemes for the two cases of logic "1" and logic "0" at the output for four inputs using two cascaded inverters. The threshold voltage of the first inverter is adjusted to lie between the values of V_n and V_{n-1}. The threshold voltages of the NMOS transistors

of the two cascaded inverters of the added buffer are put equal to 50 mV in order to speed-up the emergence of the output in case of logic "1" at the output. For eight inputs, the proposed alternative can be achieved using two copies of the circuits used with the case of four inputs and a two-input OR gate as shown in Fig. 1 (c). The two-input OR gate can in turn be achieved using a two-input static-complementary CMOS
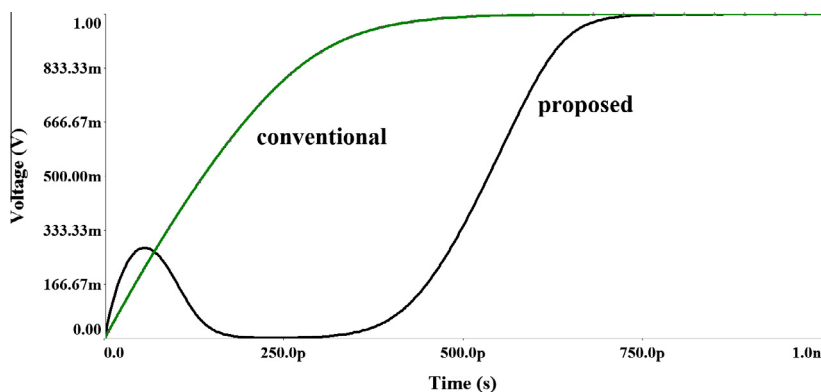
Figure 24 Simulation results showing the output voltage versus time for the conventional and proposed schemes in case of logic "1" at the output for *four* inputs.

Figure 25 Simulation results showing the output voltage versus time for the conventional and proposed schemes in case of logic "0" at the output for *four* inputs.
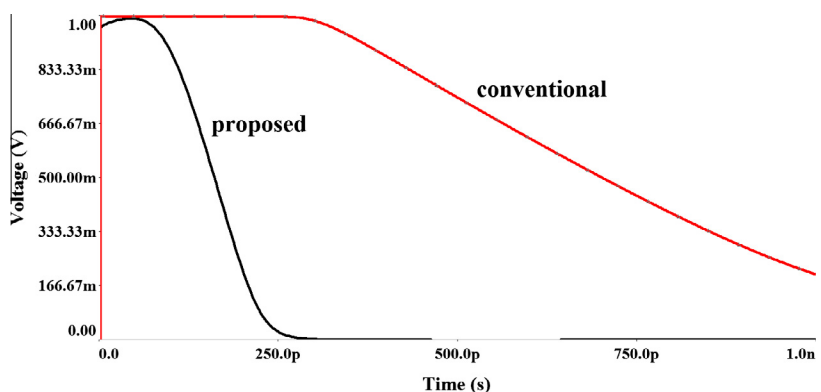

Figure 26 Simulation results showing the output voltage versus time for the conventional and proposed schemes in case of logic "1" at the output for *eight* inputs.

NOR gate and an inverter. Figs. 26 and 27 are the counterparts to Figs. 24 and 25 for *eight* inputs.

Refer to Table 1 for the high-to-low and the low-to-high propagation delays of the conventional and proposed schemes,

t_{PHLc}, t_{PHLp}, t_{PLHc}, and t_{PLHp}, along with their average, t_{avgc} and t_{avgp}, and the percentage reduction for these two cases: *four* inputs and *eight* inputs.
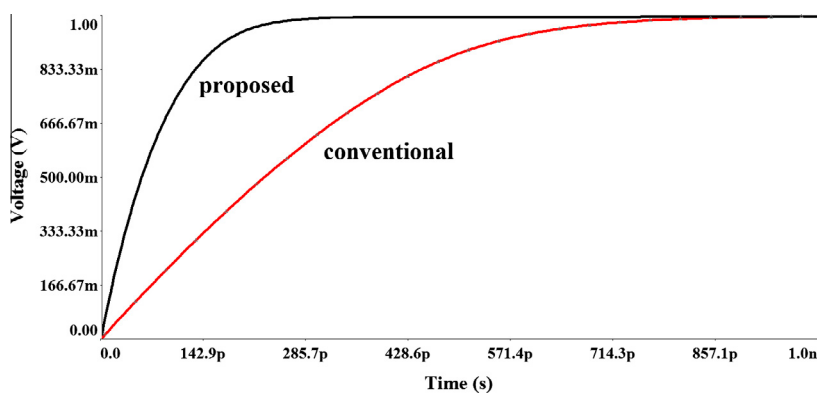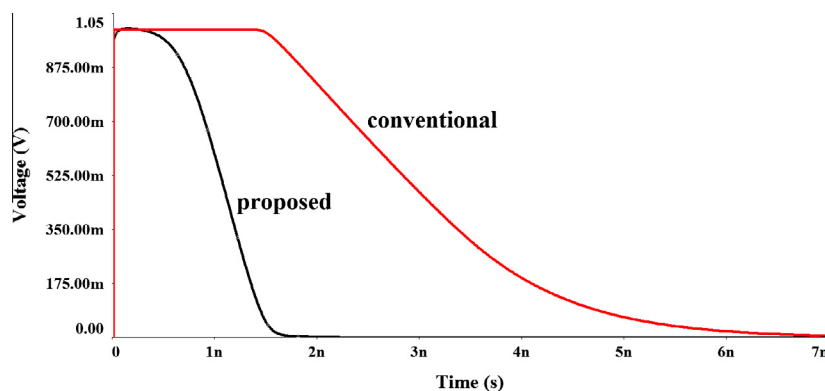
Figure 27 Simulation results showing the output voltage versus time for the conventional and proposed schemes in case of logic "0" at the output for *eight* inputs.

Table 1 The simulation results of the conventional and proposed schemes (all in ps) in case of *four* and *eight* inputs.

Four inputs		Eight inputs	
t_{PHLc}	700.86	t_{PHLc}	2900
t_{PLHc}	136.5	t_{PLHc}	227.2
t_{avgc}	418.68	t_{avgc}	1563.6
t_{PHLp}	158	t_{PHLp}	1156
t_{PLHp}	547.16	t_{PLHp}	55.5
t_{avgp}	352.58	t_{avgp}	605.75
% reduction in average propagation delay	15.8%	% reduction in average propagation delay	61.25%

8. Conclusions

In this paper, a novel alternative to CMOS stacks was presented. This alternative depends on applying the inputs with a certain pulse width. The percentage reduction in the average propagation delay was found to be 15.8% and 61.25% for the cases of four and eight inputs, respectively, for the 45 nm CMOS technology. It could be concluded from the rough estimation of the area that the proposed scheme has an area advantage over the conventional stack when the number of inputs exceeds 3. The proposed scheme has a smaller propagation delay for all numbers of inputs and a larger power consumption for practical number of inputs due to the static-dc and short circuit power consumption. However, the power-delay product of the proposed scheme is smaller than that of the conventional one when the number of inputs exceeds 2. The second-order effects were taken into account and it was shown that these effects do not affect the proposed scheme significantly.

References

[1] J.E. Ayers, Digital Integrated Circuits: Analysis and Design, CRC Press, Boca Raton, USA, 2005.
[2] A.S. Sedra, K.C. Smith, Microelectronic Circuits, seventh ed., Oxford University Press, New York, 2015.
[3] S.C. Prasad, K. Roy, Transistor reordering for power minimization under delay constraint, ACM Trans. Design Autom. Electr. Syst. 1 (2) (1996) 280–300.
[4] E. Musoll, J. Cortadella, Optimizing CMOS circuits for low power using transistor reordering, in: Proceedings of European Design and Test Conference, Paris, 1996, pp. 219–223.
[5] L. Ding, P. Mazumder, On optimal tapering of FET chains in high-speed CMOS circuits, IEEE Trans. Circuit Syst. 48 (12) (2001).
[6] M. Shoji, N.J. Warren, Apparatus for increasing the speed of a circuit having a string of IGFETs, U.S. Patent: 4430583, Feb. 7, 1984.
[7] B.S. Cherkauer, E.G. Friedman, The effects of channel width tapering on the power dissipation of serially connected MOSFETs, in: IEEE International Symposium on Circuits and Systems, 3–6 May 1993, Chicago, IL, vol. 3, 1993, pp. 2110–2113.
[8] R.H. Krambeck, C.M. Lee, H.F.S. Law, High-speed compact circuits with CMOS, IEEE J. Solid-State Circuits SC-17 (June) (1982) 614–619.
[9] S. Choudhary, S. Qureshi, Power aware channel width tapering of serially connected MOSFETs, in: International Conference on Microelectronics, 29–31 Dec 2007, Cairo, 2007, pp. 399–402.
[10] B.S. Cherkauer, E.G. Friedman, Channel width tapering of serially connected MOSFET's with emphasis on power dissipation, IEEE Trans. Very Large Scale Integr. VLSI Syst. 2 (1) (1994) 100–114.
[11] J. Yuan, C. Svensson, Principle of CMOS circuit power-delay optimization with transistor sizing, in: IEEE International Symposium on Circuits and Systems, 12–15 May 1996, Atlanta, GA, vol. 1, 1996, pp. 637–640.
[12] C. Tan, J. Allen, Minimization of power in VLSI circuits using transistor sizing, input ordering, and statistical power estimation, in: Proceedings of International Workshop Low-Power Design, 1994, pp. 75–80.

[13] X. Kavousianos, D. Nikolos, Novel single and double output TSC Berger code checkers, in: 16th Proceedings of VLSI Test Symposium, 26–30 Apr 1998, Monterey, CA, 1998, pp. 348–353.

[14] C. Metra, M. Favalli, B. Ricco, Tree checkers for applications with low power-delay requirements, in: Proceedings of International Symposium on Defect and Fault Tolerance VLSI Systems, 1996, Boston, MA, 1996, pp. 213–220.

[15] W. Kuzmicz, Leakage physics and modeling – exercises, available from the IDESA project web site <http://www.idesa-training.org/Docs/Leakage_exercises_final.pdf>.

[16] M.V. Dunga, X. Xi, J. He, W. Liu, K.M. Cao, X. Jin, J.J. Ou, M. Chan, A.M. Niknejad, C. Hu, BSIM4.6.0 MOSFET Model: User's Manual, University of California, Berkeley, 1986.

[17] D.A. Neamen, Semiconductor Physics and Devices: Basic Principles, fourth ed., McGraw-Hill, 2012.

[18] N.H.E. Weste, D.M. Harris, CMOS VLSI Design: A Circuits and Systems Perspective, fourth ed., Addison-Wesley, Massachusetts, USA, 2011.

[19] K.-Y. Toh, P.-K. Ko, R.G. Meyer, An engineering model for short-channel MOS devices, IEEE J. Solid-State Circuits 23 (4) (1988) 950–958.

[20] A. Hamoui, Current, delay, and power analysis of submicron CMOS circuits, Master Thesis, McGill University, Montréal, 1998.

[21] S.M. Sharroush, Design techniques for high performance MOS digital integrated circuits, Doctor of Philosophy Thesis, Port Said University, Port Said, Egypt, 2011.

[22] A.S. Sedra, K.C. Smith, Microelectronic Circuits, fourth ed., Oxford University Press, New York, 1998.

[23] V. Kursen, E.B. Friedman, Multi-Voltage CMOS Circuit Design, John Wiley & Sons Ltd, Britain, 2006.

[24] W. Chase, F. Bown, General Statistics, fourth ed., Wiley, 1999.

[25] J.P. Uyemura, CMOS Logic Circuit Design, Kluwer Academic Publishers, New York, 2002.

[26] Y.S. Abdalla, Design of high speed MUX/DMUX using a new all-time-on single-ended CMOS logic, Doctor of Philosophy Thesis, Waterloo, Ontario, Canada, 2006.

[27] Y. Tsividis, Operation and Modeling of the MOS Transistor, second ed., McGraw-Hill, Boston, 1999.

[28] Y. Cheng, C. Hu, MOSFET Modeling & BSIM3 User's Guide, Kluwer Academic Publishers, Boston, 1999.

[29] R.S. Muller, T.I. Kamins, Device Electronics for Integrated Circuits, second ed., John Wiley, New York, 1986.

[30] J.S. Rad, Design and analysis of robust variability-aware SRAM to predict optimum access-time to achieve yield enhancement in future nano-scaled CMOS, Doctor of Philosophy Thesis, University of California, Santa Cruz, USA, 2012.