Tunable CMOS Delay Gate With Improved Matching Properties

Przemysław Mroszczyk, Student Member, IEEE, and Piotr Dudek, Senior Member, IEEE

Abstract—This paper presents the analysis and design of a tunable CMOS delay gate with improved matching properties as compared with the commonly used "current starved inverter" (CSI). The main difference between these structures lies in the location of the current limiting transistor on the inverter's output rather than on the side of the power rail. This improves the dynamic performance of the proposed "output split inverter" (OSI) circuit reducing the influence of the MOS transistor mismatch on the generated delay time variability. A test chip including two arrays consisting of 512 16-stage delay lines employing the CSI and OSI gates has been designed and fabricated in a standard 90 nm CMOS technology. The experimental results show that the proposed OSI circuit is about 10-50% more accurate than the conventional current starved inverter with no penalty in terms of the increased area, power consumption or complexity. Applications of the proposed circuit are in the design of time-to-digital converters (TDCs), delay locked loops, readout circuits for particle detection and time-based asynchronous computation systems.

Index Terms—CMOS, current starved inverter, delay line, fabrication mismatch.

I. INTRODUCTION

UNABLE CMOS delay gates and delay lines are important functional sub-blocks in a number of applications requiring generation of the controlled delay time intervals such as delay locked loops [1], time-to-digital converters [2], silicon pixel readout circuits for particle detection [3]-[6], asynchronous processor arrays [7], and neuromorphic circuits [8], [9]. Due to the parameter variability caused by the fabrication process, an array of identically designed delay gates or delay lines will always generate delay time intervals with randomly varying offsets, even under the same bias conditions. Such mismatch of the generated time intervals is usually reported as the dominant factor limiting the precise operation of the entire design. The majority of applications found in literature employ a typical structure of a delay gate based on the current starved inverter (CSI) circuit [Fig. 1(a)] and provide discussions and analyses concerning parameter variability caused by fabrication mismatch [3]-[6].

The optimization of circuit performance in terms of parameter variability usually focuses on reducing MOS device mismatch by minimizing the variability of physical parameters of the transistors. A commonly used approach is transistor size

The authors are with the School of Electrical and Electronic Engineering, The University of Manchester, Manchester, M13 9PL, U.K. (e-mail: przemysław. mroszczyk@postgrad.manchester.ac.uk; pdudek@manchester.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSI.2014.2312491

Fig. 1. Schematic diagrams of the n-MOS asymmetric delay gate circuits: (a) the commonly used current starved inverter (CSI), and (b) the proposed "output-split inverter" (OSI).

scaling based on the assumption that the local variation averages out when the gate area increases [12]–[21]. It has been shown that the parameter variability observed between identically designed MOS transistors relates to their gate size WL and the distance D between them. For a particular parameter P of the device, the variance σ_P^2 depends on the area and the distance according to the formula [13]:

$$\sigma_P^2 = \frac{A_P^2}{WL} + S_P^2 D^2 \tag{1}$$

where A_P and S_P are the process dependent parameters. Usually, for estimation of the short range parameter variation of closely laid out devices (i.e., the fabrication mismatch) only the first, area-dependent component in (1) can be considered. Also, proper selection of the W/L ratio can further improve the device matching by minimizing the influence of the absolute gate roughness on the variability of its area [16]. The matching between adjacent transistors can also be improved by employing proper layout drawing techniques accounting for the well-proximity effect, the STI stress and the metal coverage issues [15], [21]. Other practically used mismatch reduction techniques aim to minimize the impact of the parameter variability in MOS transistors on the operation of the entire circuit or its particular sub-blocs. The commonly used methods employ auto-zero calibration (offset compensation) techniques, bias point optimization, selective transistor optimization (e.g., enlarging the area of the transistors mostly contributing to the overall circuit performance), design redundancy and averaging, and post fabrication trimming [3]–[7]. In general, methods used at this level should be devised individually depending on the circuit and its application. When analyzing the statistical behavior of a circuit with MOS transistors, it is often important to estimate the relationship between the variability of the circuit parameter of interest and the variability of the electrical or physical parameters of the device. This is typically done using the method of moments [19]

Manuscript received September 13, 2013; revised December 01, 2013, January 08, 2014, and January 29, 2014; accepted February 17, 2014. Date of publication April 28, 2014; date of current version August 26, 2014. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) of the U.K.. This paper was recommended by Associate Editor M. Alioto.

which approximates the variance σ_F^2 of a function f of n uncorrelated random variables, $f(P_1, P_2, \ldots, P_n)$ with variances $\sigma_{P1}^2, \sigma_{P2}^2, \ldots, \sigma_{Pn}^2$ by the formula:

$$\sigma_F^2 = \left(\frac{\partial f}{\partial P_1}\right)^2 \sigma_{P1}^2 + \left(\frac{\partial f}{\partial P_2}\right)^2 \sigma_{P2}^2 + \ldots + \left(\frac{\partial f}{\partial P_n}\right)^2 \sigma_{Pn}^2$$
(2)

In this paper an alternative structure of a delay gate, the "output-split inverter" (OSI) presented in Fig. 1(b), is proposed. The only topological difference between the CSI and OSI circuits is the location of the current limiting transistor M_D on the drain rather than source side of the switching transistor M_1 , which separates or "splits" the output of the inverter. In principle, the operation of both circuits is similar and, for the rising input edge, the transistor M_D controls the current discharging the output capacitance C_1 regulating the output signal falling edge slope and hence discharge time. However, during the transient state, both circuits behave differently which has a direct impact on their performance under the process parameter variability. This paper extends our work presented in [11], and provides more detailed analysis and discussion as well as measurement results of the CSI and OSI delay lines fabricated in a standard 90 nm CMOS technology. The proposed OSI structure was previously used, for example, in the build of charge pump circuits [1], [8], [9], linearly tuned delay element [10], and asynchronous arrays for binary image processing [7]. Nevertheless, to the best of our knowledge, no previous work, except our papers [7] and [11] postulate the superior matching performance of this circuit.

Section II of this paper analyses the operation of both gates based on Spice simulation results, Section III discusses the impact of parameter variability on the circuit performance and presents a simplified analytical model for circuit optimization, Section IV presents the test chip implementation, Section V discusses the obtained experimental results and Section VI concludes the paper.

II. CIRCUIT OPERATION

The implementation of both delay gates presented in Fig. 1 uses the structure of a logic inverter (transistors M_1 and M_2) and employs the idea of delaying the output slope by discharging the capacitance C_1 with the drain current of M_D controlled by the bias voltage V_D . Assuming that the transistors M_1 and M_2 work as ideal switches, the operation of both circuits is very similar. Depending on the transition of the input signal, the slew of the output signal is either controlled by the current limiting transistor M_D (for the rising input edge) or is determined by the strength of M_2 pulling the output node up to V_{DD} (for the falling input edge). However, a more detailed analysis of these gates reveals substantial differences in their operation which are of high importance in terms of the process parameter variability and its influence on circuit performance. The simulation results showing the transitions of V_{IN} , V_{OUT} and V_{SD} signals of the asymmetric CSI and OSI delay gates from Fig. 1, for the rising input slope (when the output load C_1 discharges through the current limiting transistor M_D), are presented in Fig. 2. In the simulations MOS transistor models from a standard 90 nm CMOS technology were used assuming the same sizes for the current limiting transistors $W_D/L_D = 1 \ \mu m/0.5 \ \mu m$ and for the switching transistors $W_{1,2}/L_{1,2} = 1 \ \mu m/80 \ nm$. Capacitances C_1 and C_2 are always present due to the junction ca-



Fig. 2. Simulation results of the circuits from Fig. 1: (a) CSI gate, and (b) OSI gate for the rising edge of the input voltage V_{IN} , $V_D = 300$ mV, $V_{DD} = 1.0$ V and $t_R = 100$ ps (MOS transistor models from a 90 nm CMOS technology were used assuming $W_D/L_D = 1 \ \mu m/0.5 \ \mu m$, $W_{1,2}/L_{1,2} = 1 \ \mu m/80$ nm; $C_1 = 1$ fF, $C_2 = 0$ plus inherent source/drain MOS capacitances).

pacitances of MOS transistors. An additional capacitance of 1 fF has been added to represent the external load of the output node. In the full custom layout design, both transistors M_D and M_1 may share the same diffusion stripe therefore there is no additional capacitance attached to the V_{SD} node apart from the geometry-dependent one associated with the drain and source regions of the MOS transistors, already included in their models. In the simulations the bias voltage $V_D = 300 \text{ mV}$ was used for both gates and the generated delay time T_D was measured between 50% of V_{DD} of the input (V_{IN}) and output (V_{OUT}) signals.

The main difference in the operation of both delay gates can be observed during the transient state when the rising input edge turns M_1 on and M_2 off. In the case of the proposed OSI gate, capacitance C_2 (initially charged close to V_D through M_D when M_1 is off for $V_{IN} = 0$) quickly discharges to zero whereas capacitance C_1 is gradually discharged from V_{DD} to zero by the drain current of M_D [Fig. 2(b)]. In the CSI circuit, however, capacitance C_2 is initially discharged since M_D is always on, and the rising input edge causes the transistor M_1 to short its drain and source nodes such that the voltages V_{OUT} and V_{SD} converge close to the common value V_{CM} , before the load capacitance starts discharging [Fig. 2(a)]. As a result, in the CSI circuit capacitance C_1 will discharge quicker and the output voltage V_{OUT} will drop faster to 50% of V_{DD} for the same current of M_D as compared to the OSI one. The variability of the slew of the output signal is mostly dependent on the parameter variability of M_D and the output load, but in the CSI circuit the discharge time will also be affected by the variability of V_{CM} resulting from the mismatch between transistors M_1 and M_2 . Therefore, the commonly used CSI circuit tends to generate shorter and more variable delay intervals than the proposed OSI structure where capacitance C_1 of the output node always discharges from the constant V_{DD} voltage. In the following, we provide a simplified analysis presenting only the first-order behavior addressing the major differences between circuits in Fig. 1 and their operation under the presence of the process parameter variability.

A. Current Starved Inverter (CSI)

A simplified analysis of the CSI circuit showing the transitions of voltages V_{IN} , V_{OUT} and V_{SD} is presented in Fig. 3. The timeline can be divided into three phases: the initial phase $(t < t_1)$, the switching phase $(t_1 < t < t_2)$, and the discharge phase $(t > t_2)$.



Fig. 3. The behavior of the CSI delay gate in a transient state for the rising input edge.



Fig. 4. The behavior of the OSI delay gate in a transient state for the rising input edge.

In the initial phase, the capacitance C_2 is discharged to zero through the current limiting transistor M_D (for $V_D > 0$) and capacitance C_1 is charged to V_{DD} through M_2 . In the switching phase, the rising edge of V_{IN} turns M_1 on and M_2 off. As a result, V_{OUT} and V_{SD} converge closer to the common level V_{CM} denoting the starting point for the discharge phase. In a closer view, the transistor M_1 turns on first (when $V_{IN} > V_{thn}$) quickly increasing the conductance between nodes V_{OUT} and V_{SD} . In the same time the drain-source conductance of M_2 decreases practically disconnecting the output node from the power rail. Due to the current limiting transistor M_D , the total current flowing through M_1 and M_2 is reduced and the observed output voltage drop (from V_{DD} to V_{CM}) can practically be attributed to the charge sharing between capacitances C_1 and C_2 . During the discharge phase the high logic level on the input keeps transistor M_1 fully turned on which connects capacitances C_1 and C_2 in parallel. The discharge rate of these capacitances depends mainly on the bias voltage V_D controlling the current limiting transistor M_D .

B. Output Split Inverter (OSI)

The analysis of the OSI circuit showing the transitions of voltages V_{IN} , V_{OUT} and V_{SD} is presented in Fig. 4. In the following, only the first-order effects will be discussed indicating the major differences in the operation between the CSI and OSI structures.

In the initial phase, voltage V_{IN} equals zero assuring that capacitance C_1 is charged to V_{DD} through the transistor M_2 . The current limiting transistor M_D operates in weak inversion with its gate-source voltage high enough above zero to conduct the small off current of M_1 , therefore the capacitance C_2 remains



Fig. 5. Transient mismatch Monte Carlo simulations (5000 runs) of the delay gates from Fig. 1 tuned to generate equal mean delays $T_D \approx 0.46$ ns ($V_D = 280 \text{ mV}$ for the CSI and $V_D = 300 \text{ mV}$ for the OSI gate): (a) input and output signal transitions, (b) detailed view of the output signals crossing the 50% V_{DD} threshold.

charged closely to the gate bias voltage V_D . In the switching phase the rising edge of the input signal turns M_1 on (when $V_{IN} > V_{thn}$), which quickly discharges C_2 to zero and, after that, turns M_2 off (when $V_{IN} > V_{DD} - |V_{thp}|$) and C_1 starts discharging with rate dependent mainly on the drain current of M_D controlled by the bias voltage V_D . Further operation of this gate is practically the same as in the case of the CSI one.

III. MISMATCH ANALYSIS

In literature mismatch analysis and optimization of the timing parameters of the current starved inverter circuit is usually done assuming that the transistors M_1 and M_2 work as ideal switches and the precision of the generated time delay T_D depends mainly on the current limiting transistor M_D [1]–[6]. More elaborate analysis of this circuit, accounting for the inter- and intra-die parameter fluctuations, providing guide-lines towards mismatch modelling and circuit optimization, was presented in [5]. The majority of works conclude that the precision of the time T_D can be improved by enlarging the current limiting transistor but the corresponding precision-area trade off should also be considered. For example, the techniques employing redundant gate insertion for post fabrication trimming were successfully used in some specific applications but at the expense of increased circuit area and complexity [4].

The effects of parameter variation of MOS transistors in the realizations of the CSI and OSI delay gates are presented in Fig. 5. In the simulations, the same circuit realizations as before (Fig. 2) were used, but with the mismatch Monte Carlo MOS transistor models and bias voltages V_D tuned for both gates to ensure the same mean value of the generated delays T_D . It can be seen that the random variability of the generated delay is larger in the CSI gate [Fig. 5(b)]. The detailed simulation results accounting for the variability of the generated delay T_D caused by mismatch of individual transistors in the delay gates are presented in Table I. In particular, the effects of the input signal slope variability was verified using additional buffer BUFF consisting of two inverters (designed using the same transistors as M_1 and M_2) driving the input of the delay gates. It can be concluded that the variability of the generated delay time T_D in both gates depends mainly on the variability of the current limiting transistor M_D [11]. In the following, we provide a simplified analysis of the switching and discharge phases for both delay gates explaining the influence of the MOS parameter variability on the precision of the generated delays.

 TABLE I

 MISMATCH MONTE CARLO SIMULATION RESULTS OF THE CSI AND OSI GATES



Fig. 6. The transient state of the CSI circuit showing the influence of the MOS parameter variability on the V_{CM} voltage and the generated delay T_D .

A. Mismatch in the CSI Circuit

The analysis of the CSI circuit showing the influence of the process parameter fluctuation on the variability of the V_{CM} voltage and the generated delay time T_D is presented in Fig. 6.

During the switching phase $(t_1 < t < t_2)$ both voltages V_{SD} and V_{OUT} converge towards the common level V_{CM} which, in a crude approximation, can be estimated from the charge sharing between C_1 and C_2 (this will be further explained in Section IV). Therefore the variability of the voltage V_{CM} will be affected by the variability of these capacitances but also by the variability of all the MOS transistors, especially M_D , and other effects such as off (leakage) current of M_2 , and the capacitive coupling between the input and output. The discharge phase will mainly be affected by the variability of the current limiting transistor M_D . For example, due to the random variation of the threshold voltage V_{thn} of M_D , this transistor may be slightly "faster" (higher drain current for lower values of V_{thn}) or slightly "slower" (lower drain current for higher values of V_{thn}) than a regular one. For the faster M_D , the corresponding discharge slope will be steeper and also the V_{CM} voltage will be lower (due to the higher current of M_D discharging the output node during the switching phase), whereas for the slower M_D , the V_{CM} voltage will reach a higher value and the discharge phase will take a longer time. As a result, it can be observed that not only the parameter variability of the current limiting transistor M_D but also the variability of V_{CM} voltage (ΔV_{CM}) will affect the precision ΔT_D of the generated time delay T_D .

B. Mismatch in the OSI Circuit

The analysis of the OSI circuit showing the influence of the MOS parameter fluctuations on the variability of the generated time delay T_D is presented in Fig. 7. Due to the current limiting transistor M_D "splitting" the output of the inverting stage, the rising edge of V_{IN} may not force an immediate transition of V_{OUT} as was observed in the case of the CSI circuit. While the capacitance C_2 quickly discharges to zero, M_2 still pulls the output node up to V_{DD} postponing the discharge phase roughly



Fig. 7. The transient state of the OSI circuit showing the influence of the MOS parameter variability on the generated delay time T_D .

until V_{IN} crosses the threshold $V_{DD} - |V_{thp}|$ switching transistor M_2 off. The variability ΔT_D in the generated delay time will mainly depend on the parameter mismatch in transistors M_D and M_2 . Similarly as before, the slightly faster transistor M_D will force the discharge phase earlier and will discharge the output capacitance C_1 faster. Additionally, for the slightly slower transistor M_2 with the higher threshold voltage $|V_{thp}|$ the discharge phase may begin earlier than for the slightly faster one, further increasing the variability of the generated delay time T_D . The influence of the variability of the threshold voltage ΔV_{thp} of M_2 is usually suppressed by a sharp slope of V_{IN} . Also, the discharge phase always begins for the same output voltage $V_{OUT} = V_{DD}$ irrespective of the variability in C_1 and C_2 . Because of this, the starting point of the discharge phase is more stable (C_1 is always charged to the constant voltage V_{DD}) and the discharge time of C_1 is longer for the same current of M_D as compared to the CSI structure where $V_{CM} < V_{DD}$. This makes the generated delay time of the OSI circuit less prone to mismatch.

C. Simplified Analytical Model

In the proposed simplified analytical model only the dynamic behavior of the CSI and OSI delay gates during the discharge phase (with the initial conditions determined by the switching phase) will be considered. It is assumed that the current limiting transistor M_D operates in the saturation and strong inversion regions within the generated delay time interval T_D . This assumption holds for typical applications where the output signal triggers the next stage (e.g., the next gate in a delay line) at the 50% signal level, which is higher than a typical value of the saturation voltage of M_D (usually $V_{DSAT} \ll V_{DD}/2$). Also, the gate-source voltage of M_D (equal to V_D when M_1 is fully turned on) is usually higher than the threshold voltage V_{thn} in order to avoid the increased impact of parameter mismatch on the circuit operation when M_D is in the subthreshold region. The schematic diagrams of the simplified CSI and OSI delay gates representing the state of each circuit after the switching phase are shown in Fig. 8.

In both cases the current limiting transistor M_D was replaced with an ideal current source i_D . The transistors M_1 and M_2 were replaced with switches where the non-ideal behavior of these devices, in the case of the CSI gate, can be seen as an additional factor affecting the V_{CM} voltage. The assumption of charge sharing between C_1 and C_2 , introduced in Section II as the primary reason for the output voltage drop, was verified in the simulations and remains valid almost within the entire tuning range except for the very short delays for $V_D \gg$



Fig. 8. Schematic diagrams of the simplified delay gates representing the state of the circuits after the switching phase for: (a) CSI delay gate ($V_{OUT} = V_{CM}$), and (b) OSI delay gates ($V_{OUT} = V_{DD}$).

 V_{thn} when M_D starts pulling the node V_{SD} and V_{OUT} closer down to zero during the switching phase. In particular for the gate design implemented in the test chip, the inherent geometry dependent drain/source capacitances are approximately equal to 2.5 fF for the V_{OUT} node and 1.5 fF for the V_{SD} node. The values were calculated assuming charge sharing effect with the resulting V_{CM} voltage obtained from simulations for different load capacitances (within range of 0...1 fF) and different bias voltages V_D generating delays within range of 0.6...9 ns. It is important to note that other effects observed during the switching phase, such as non-zero off current of M_2 , non-zero drain current of M_D and the capacitive coupling between the input and output nodes, also affect the V_{CM} voltage, therefore the charge sharing model provides only a crude approximation, suitable for rough estimation of V_{CM} . Nevertheless, it is sufficient for the purpose of this analysis. Another important issue concerns the separation between the switching and discharge phases. In practice, the discharge phase may be triggered earlier, as soon as the conductance of M_1 becomes dominant. For the purpose of this analysis the discharge phase is assumed to begin when the input voltage crosses 50% of V_{DD} and terminate (generating time delay T_D) when the output voltage crosses the same threshold (the capacitances continue to discharge to 0 V after that).

D. Variability Analysis

In the proposed approach, the delay time T_D will be derived for the simplified CSI and OSI circuits from Fig. 8 assuming the discharge scheme presented in Fig. 9. The discharge phase of the current starved inverter (CSI) delay gate starts for the output voltage $V_{OUT} = V_{CM} = V_{DD} \times C_1/(C_1 + C_2)$. Assuming the ideal operation of the switches, the generated delay time $T_D = T_{CSI}$ depends only on the current i_D discharging the capacitances $C_1 + C_2$ from the initial voltage V_{CM} to V_P (typically equal $V_{DD}/2$, terminating the generated time interval T_D , Fig. 9). In the case of the proposed OSI circuit, the discharge phase always starts for the output voltage $V_{OUT} = V_{DD}$ and terminates when the current i_D discharges the capacitance C_1 down to V_P (the capacitance C_2 is already shorted to the ground and does not participate in the discharge phase). The generated delay time of the CSI and OSI gates is:

$$T_{CSI} = (C_1 + C_2) \frac{V_{CM} - V_P}{i_D},$$
(3)

$$T_{OSI} = C_1 \frac{V_{DD} - V_P}{i_D} \tag{4}$$



Fig. 9. The transient state of the CSI and OSI circuit models generating the delay time interval T_D measured at 50% input and output signal level and assuming the discharge of the respective capacitances by the i_D current.

One of the advantages of the proposed OSI circuit is its capability of generating longer delay time intervals T_D for the same bias conditions. This is mainly caused by the fact that C_1 is usually larger than C_2 due to the additional load of the node V_{OUT} , and the voltage V_{CM} is usually lower than V_{DD} . In particular, inserting $V_{CM} = V_{DD} \times C_1/(C_1 + C_2)$ and $V_P = V_{DD}/2$ to (3), and assuming $V_{CM} \ge V_P$ (to ensure discharging of the load capacitance from V_{CM} to V_P) in (3) and (4), the ratio T_{CSI}/T_{OSI} will simplify to the following formula:

$$\frac{T_{CSI}}{T_{OSI}} = \frac{C_1 - C_2}{C_1}$$
(5)

It is important to note that (5) was derived assuming $C_1 >$ C_2 (which is usually the case due to the additional load capacitance) and switching at 50% of the maximum signal level $(V_P = V_{DD}/2)$. It can be observed that, in such case, the OSI gate will generate a longer delay time interval $(T_{OSI} > T_{CSI})$. The proposed charge sharing approach can be further extended to include other systematic effects affecting V_{CM} voltage. For example, for a high coupling between the input and output nodes (due to, e.g., high gate-source and gate-drain capacitances in MOS transistors) the additional charge injected to nodes V_{OUT} and V_{SD} from the input slope will slightly increase the V_{CM} voltage which, from the perspective of the proposed model, can be seen as an increase of the capacitance C_1 in (5). Assuming charge sharing between C_1 and C_2 any other factor affecting the charge stored on these capacitances can be theoretically accounted for by modifying the value of C_1 or C_2 . In the extreme case, when the charge injected to the output node causes $V_{CM} \approx V_{DD}$, the operation of both CSI and OSI delay gates will become similar due to the fact that both gates will start the discharge phase for the same initial condition $V_{OUT} = V_{DD}$.

The variability of the generated time delay T_D of the CSI and OSI circuits can be estimated by applying (2) to the calculated delays T_{CSI} and T_{OSI} . For both circuits it is assumed that the variability of the delay time T_D results mainly from the variability of the parameters of the current limiting transistor M_D and the capacitive load of the output node, therefore only the variability of the current i_D and capacitances C_1 and C_2 will be accounted for in the following calculations. Despite its limitations, the proposed approach covers all major contributors to the T_D time variability including all MOS transistors (e.g., the capacitive load of the next stage in a delay line will depend on the variability of M_1 and M_2) and the interconnecting tracks. The normalized delay variances derived for the CSI and OSI circuits ((3) and (4)) assuming $V_{CM} = V_{DD} \times C_1/(C_1 + C_2)$, $V_P = V_{DD}/2$ and $V_{CM} \ge V_P$ are equal respectively to:

$$\frac{\sigma_{TCSI}^2}{T_{CSI}^2} = \frac{\sigma_{ID}^2}{i_D^2} + \frac{\sigma_{C1}^2 + \sigma_{C2}^2}{(C_1 - C_2)^2},\tag{6}$$

$$\frac{\sigma_{TOSI}^2}{T_{OSI}^2} = \frac{\sigma_{ID}^2}{i_D^2} + \frac{\sigma_{C1}^2}{C_1^2}$$
(7)

where σ_{ID}^2 , σ_{C1}^2 and σ_{C2}^2 are the variances of the current i_D and the capacitances C_1 and C_2 . Assuming the simplest square law model of the transistor M_D operating in strong inversion and saturation in (6) and (7), the following equations can be obtained:

$$\frac{\sigma_{TCSI}^2}{T_{CSI}^2} = \frac{4\beta \sigma_{Vth}^2 T_{CSI}}{V_{DD}(C_1 - C_2)} + \frac{\sigma_\beta^2}{\beta^2} + \frac{\sigma_{C1}^2 + \sigma_{C2}^2}{(C_1 - C_2)^2}$$
(8)

$$\frac{\sigma_{TOSI}^2}{T_{OSI}^2} = \frac{4\beta \sigma_{Vth}^2 T_{OSI}}{V_{DD}C_1} + \frac{\sigma_{\beta}^2}{\beta^2} + \frac{\sigma_{C1}^2}{C_1^2}$$
(9)

Equations (8) and (9) show a linear dependence of the normalized delay variance in terms of the generated delay time. For a typical circuit implementation (when $C_1 > C_2$) the slope and the y-intercept components in (8) are greater than the respective ones in (9) due to the dependency of the delay time of the CSI circuit on both C_1 and C_2 . As a result, the delay variance of the CSI gate is higher than the one of the OSI gate for the same generated delay. It can be concluded that the OSI gate generates delay time intervals that are longer, as shown in (5), and less affected by parameter variability [(8) and (9)].

The proposed analysis can also be applied to designs of delay lines where a certain number of delay gates are connected in series creating a chain. For N gates with delay T_D each, connected in a chain, the delay of the entire line is equal to $T_N = NT_D$, i.e., it will increase linearly with the number of stages. Assuming that delays T_D generated by different stages are normally distributed and independent random variables with variance σ_{TD}^2 , the total delay variance of a line will be equal to $\sigma_{TN}^2 = N\sigma_{TD}^2$, and hence the normalized delay variance of the line consisting of N such stages can be calculated as:

$$\frac{\sigma_{TN}^2}{T_N^2} = \frac{1}{N} \frac{\sigma_{TD}^2}{T_D^2}$$
(10)

It should be noted that, in some applications, the symmetric delay gates with two current limiting transistors on both pull-up and pull-down sides may be used [3]–[6]. Their operation is practically the same as the operation of the discussed asymmetric circuits but the delaying of the output signal occurs on both falling and rising slopes controlled either by transistor M_{DN} or M_{DP} (e.g., for the falling output slope transistor M_{DN} controls the delay time whereas M_{DP} does not participate in the discharge phase). It is important to note that the approximately linear dependency of the normalized delay variance on the generated delay can be observed in practice (Fig. 16), for short delays, when the current limiting transistors operate in strong inversion.

IV. IMPLEMENTATION

In order to compare the operation and statistical parameters of the CSI and OSI delay gates, two arrays of delay lines employing these structures were fabricated in a standard 90 nm CMOS technology. The architecture of the entire test system

COLUM SELECT SHIFT REGISTER 32 × 16 OSI DELAY CELLS OATA CLA N VO BUFFERS OUT

Fig. 10. The test system with the CSI and OSI delay line arrays implemented on the chip.



Fig. 11. The schematic diagram of the delay cell including the 16-stage CSI/OSI delay line, I/O buffers, and the AND gate used in the row/column addressing.

implemented on the chip including the arrays of 512 delay lines each and the additional control logic, is shown in Fig. 10.

Each array consists of 32×16 delay cells selected using the row/column addressing maintained by the boundary shift register on the left and top sides of the arrays. The schematic diagram of the delay cell including a 16-stage delay line and additional control and I/O logic is presented in Fig. 11. A particular cell from any of the arrays can be selected by shifting the programming sequence into the register using signals DATA and CLK, which sets the appropriate column and row lines to the high logic state. An additional AND gate, implemented in each delay cell, will detect this condition and connect the output of the delay line to the output line through a tri-state buffer. In order to reduce the capacitive load of this buffer, the output line is shared only for the lines from the same column. The COL signal is then used to enable another tri-state buffer in the I/O block which connects the selected column to the global output. The input signal is buffered at the input of each delay cell and also individually for each column in the I/O block. In order to assure the uniform propagation times of the input and output signals, the same numbers of buffers were used for each delay line irrespective of its position in the array. Each delay cell includes a 16-stage delay line implemented using the symmetric variant of the CSI or OSI delay gate, presented in Fig. 12. In this implementation there are two current limiting transistors M_{DN} and M_{DP} of the size $W/L = 1 \ \mu m/0.5 \ \mu m$ controlled by the bias voltages V_{DN} and V_{DP} respectively. The width of the switching transistors $M_{1,2}$ is the same as the current limiting ones.

The layouts of the designed cells with delay lines consisting of 16 CSI and OSI gates used in the experiments are presented in Fig. 13. The chip micrograph showing the region where the delay line arrays were implemented is presented in Fig. 14. The area of the presented test system is $160 \ \mu m \times 1140 \ \mu m$.

V. EXPERIMENTAL RESULTS

The measurement results of the fabricated chip were obtained in a laboratory environment using an Agilent 54641D oscilloscope and a KCPSM3 (Xilinx *PicoBlaze*) controller implemented on a Spartan 3 FPGA development board. The block diagram showing the setup used in the experiments is presented in



Fig. 12. Schematic diagrams of the delay gates with two complementary current limiting transistors M_{DN} and M_{DP} : (a) CSI variant, (b) OSI variant.



Fig. 13. Layouts of the delay cells with lines consisting of 16 serially connected (a) CSI gates and (b) OSI gates (the size of each delay line is $3.7 \,\mu m \times 27 \,\mu m$).



Fig. 14. Chip micrograph showing the location the delay cell arrays.

Fig. 15. The KCPSM3 system communicates with a PC application (e.g., Matlab) via an RS-232 serial port, executes received commands and, based on that, provides communication with the chip by programming the boundary shift register and generating the square wave signal of 5 μ s period to drive a selected delay line. The delay time was measured on rising edges from 50% to 50% of the fixed level representing high logic state of input and output signals. The data acquisition setup of the oscilloscope assumed full bandwidth and averaging based on 64 samples to suppress the time jitter. Both the KCPSM3 system and the oscilloscope were working in a loop controlled by a Matlab script selecting delay lines in turn and collecting the measured delay times.

A. Calibration

In the test system a particular delay line can be tested by programming the boundary shift register with a sequence addressing the corresponding cell in the array. The output signal



Fig. 15. Block diagram of the setup used for the delay time measurements.

of the selected delay line goes through two tri-state buffers (one in the cell and one in the I/O block) output buffer, the digital I/O cell in the power ring including additional buffers and level shifters (to match the external and core logic standards) and through a buffer driving the capacitance of the test probe. Similarly the input signal is provided through the digital I/O cell and a series of buffers until it reaches the selected delay line. Due to the fact that it is not possible to measure the generated delays T_D directly at the input and output nodes of the selected line, the delay time T_{MEAS} measured between the input and output slopes of the signals outside the chip will include additional offset time T_{OFF} introduced by the buffers between the test points on the PCB, such that $T_{MEAS} = T_D + T_{OFF}$. The offset time also slightly depends on the location of a particular delay line in the array due to different lengths of I/O paths. This variability, however, is only a small fraction of the entire length of the path between the array and the probes on the PCB, therefore it can be neglected in this analysis. In order to estimate the offset time T_{OFF} , an additional column with OSI delay lines with different number of stages was implemented. Based on the measurement results of the delays T_{MEAS} for four different OSI 16-stage and 1-stage delay lines for the bias voltages $V_{DN} = V_{DD}$ and $V_{DP} = 0$, the obtained mean values are $T_{MEAS-16} = 11.73$ ns (for 16-stage lines) and $T_{MEAS-1} =$ 9.50 ns (for 1-stage lines). Assuming that the 16-stage line generates delays 16 times longer than the 1-stage one, the calculated offset time is $T_{OFF} = 9.35$ ns. It should be noted that the input driver and the output buffer of the delay line slightly affect the propagation times T_D of mostly the first and last delay stage, and hence the delay of the entire line, however, this is negligible in the offset time estimation since $T_{OFF} \gg T_D$. The delay time offset T_{OFF} was subtracted from the raw data obtained from the measurements prior to any statistical computations and analyses presented in this section.

B. Measurements

In order to compare the performance of the CSI and OSI delay lines, the normalized delay variance will be calculated based on the measurement result obtained from the entire array containing 512 CSI and 512 OSI delay lines for symmetric bias voltages $V_{DN} = 150...430$ mV and $V_{DP} = 850...570$ mV $(V_{DP} = V_{DD} - V_{DN})$. The core supply voltage of the chip is $V_{DD} = 1.0$ V. The diagram showing the normalized delay variance as a measure of the relative time variability versus mean delay time, computed based on the obtained results of the CSI and OSI delay gates accounting for the offset time $T_{OFF} =$ 9.35 ns, is presented in Fig. 16.

The experimental results show that the normalized delay variance of the proposed OSI circuit is about 10%–50% lower than in the case of the CSI structure. In other words, the relative variability of the delays is lower in the OSI array when the mean



Fig. 16. Normalized delay variance vs. mean delay time obtained from the measurements of the CSI and OSI delay lines.

value of the delay is the same for both arrays. Moreover, the obtained improvement in the accuracy of the circuit operation has no penalty in terms of area, power or circuit complexity. The only observed limitation of the proposed OSI delay gate is a slightly narrower tuning range when compared to the CSI circuit. This can be explained assuming that, for the higher currents of M_D , the V_{CM} voltage can be pulled down even below V_P resulting in very short delays (Section II).

In both CSI and OSI circuits, the generated delay time T_D increases when lowering the bias voltage V_{DN} (or increasing the voltage $V_{DP} = V_{DD} - V_{DN}$). From (8) and (9), the variability of this current depends on the operating point of the transistor, and increases with the generated delay time, also increasing the corresponding relative time variability. This means that the precision of the circuit array degrades when generating longer delays. Therefore, the tuning range of the fabricated test arrays, where the delay lines consist of only 16 gates, will most probably be restricted to 20-30 ns, where the current limiting transistors operate in strong inversion. Obviously, in order to generate longer delays of the same precision, longer delay lines should be used. The normalized delay variance versus mean delay time limited to the 30 ns tuning range, is shown is Fig. 17.

The visual representation of the generated delay times across the CSI and OSI arrays on the same gray scale map is shown in Fig. 18. The corresponding histograms of the generated delays are presented in Fig. 19. Both arrays were tuned to generate delay time intervals of approximately 11 ns. In the experiment it was difficult to tune both arrays precisely and the mean values extracted from the measurements of the delay times were $T_{DCSI} = 11.463$ ns and $T_{DOSI} = 10.633$ ns for the CSI and OSI arrays respectively. For comparison, the difference of $T_{DCSI} - T_{DOSI}$ was added to the results obtained from the OSI array to align both histograms and set the range of the gray scale map to cover the corner cases from both data sets. In the map a black color represents the fastest line of $T_{DMIN} =$ 10.885 ns delay, and the white color represents the slowest line of $T_{DMAX} = 12.070$ ns delay (after the alignment). It can be observed that the image representing delay time in the OSI array has lower contrast and looks more uniform in comparison to the



Fig. 17. Normalized delay variance vs. mean delay time obtained from the measurements for the tuning range limited to 30 ns.





30

20

10

11.0

11.2 11.4 11.6 11.8

12.0

10

11.0

11.2 11.4 11.6 11.8

Fig. 18. Visual representation of the delay time variability on the same gray scale measured for: (a) CSI array and (b) OSI array (aligned results, see text for



12.0

image obtained from the CSI array. Also the corresponding distribution of the generated delays is narrower for the OSI array indicating less effect of the physical parameter variability on the circuit performance.

The operation of the 16-stage CSI and OSI delay lines designed in standard 180 nm, 90 nm and 65 nm CMOS technologies was verified in simulations of the respective post layout circuits including RC parasitics and using mismatch Monte Carlo MOS transistor models provided by the foundry. The size of the transistors in the delay gates was the same as shown in Fig. 12. Only the channel length of the switching transistors M_1 and M_2 was defined by the minimum feature size of a particular technology. The ratio of the delay variance of the OSI and CSI lines $(\sigma_{TDOSI}/\sigma_{TDCSI})^2$ versus mean delay time T_D obtained from the simulations in three different technology nodes and measurements of the test chip in the same tuning range 10 ns-200



Fig. 20. (a) Ratio of the delay variances of the OSI and CSI circuits vs. mean delay time obtained from the post layout simulations of the circuits designed in 180 nm, 90 nm and 65 nm standard CMOS technologies and from the measurements, (b) Generated time variability caused by the fabrication mismatch and noise (time jitter) measured for CSI and OSI arrays within 50 ns tuning range.

ns is presented in Fig. 20(a). It can be observed that the operation of the proposed OSI structure is less affected by the fabrication mismatch in three different technology nodes. The delay variance ratio remains within interval of 60%–75% for 65 nm technology and 50%–90% for 180 nm and 90 nm technologies (both in simulations and measurements).

Apart from the fabrication mismatch, the generated delay time intervals are also affected by noise (from the power supply, bias voltages and thermal effects in the circuit) which causes random variation of the output signal slope (jitter). In previous experiments jitter was removed by applying averaging over 64 samples, which was sufficient to obtain stable delay measurement readouts. The influence of noise on the generated delay time was calculated based on 1000 measurements of the delay of a particular line working under constant bias conditions. The measurement results showing the calculated standard deviation of the delay times as a result of fabrication mismatch and jitter within 50 ns tuning range are presented in Fig. 20(b). For the time jitter estimation the standard deviation and the mean value of the generated delay time was measured for the fastest and the slowest delay line in both CSI and OSI arrays, and the respective average values were calculated. The obtained results show that the delay time variability is dominated by the fabrication mismatch and remains almost one order of magnitude higher than the measured time jitter. It was observed that the measured time jitter was also affected by the noise of the setup (oscilloscope, probes and IO buffers) introducing a baseline noise level of $\sigma_{TD} = 30$ ps for delay $T_D = 3.44$ ns measured in the system without the chip and with the corresponding input and output pins shorted in the socket on the PCB. The same measurement repeated with averaging over 64 samples (set up in the oscilloscope) reduced the baseline noise level to $\sigma_{TD} = 4.8$ ps. Therefore, the measurements of the time delay variability caused by mismatch (done with averaging) are practically not affected by this noise but the time jitter measurements in Fig. 20(b) can, to some extent, be overestimated due to the baseline noise level of the setup.

In the experiments four other chips from the same fabrication run were tested. The results presented above were obtained from the measurements of chip (#1). The measured statistical parameters of the CIS and OSI delay lines (tuned to generate delays around 11 ns) for chips #1–#5 are presented in Table II. For all the chips (#1–#5) the offset time T_{OFF} was measured individually. The ratios of the standard deviation to the generated mean

 TABLE II

 MEASUREMENT RESULTS OF FIVE DIFFERENT CHIPS FROM THE SAME BATCH

Chip	T [ma]	CS	SI	OSI		
	<i>IOFF</i> [IIS]	T _{DMEAN} [ns]	σ_{TD} [ps]	T _{DMEAN} [ns]	σ_{TD} [ps]	
#1	9.35	10.63	203	11.46	154	
#2	9.15	11.05	231	11.70	151	
#3	9.24	10.91	216	11.01	151	
#4	9.20	11.16	224	11.31	169	
#5	9.28	11.37	220	11.45	155	

TABLE III Delay Time Variability of Five Different Chips

	Chip #1		Chip #2		Chip #3		Chip #4		Chip #5	
T _{DMEAN}	σ_{TD}/T	D [%]	σ_{TD}/\tilde{T}	, D [%]	σ_{TD}/T	D [%]	σ_{TD}/T	Γ _D [%]	σ_{TD}/T	I _D [%]
	CSI	OSI	CSI	OSI	CSI	OSI	CSI	OSI	CSI	OSI
20 ns	2.60	1.90	2.75	1.80	2.70	1.80	2.80	2.10	2.75	1.90
30 ns	3.33	2.43	3.43	2.37	3.47	2.30	3.57	2.63	3.47	2.47
40 ns	3.75	2.75	3.90	2.73	4.03	2.70	4.13	2.98	4.00	2.83
50 ns	4.42	3.08	4.24	3.02	4.48	3.00	4.58	3.32	4.46	3.18

delay time of the CSI and OSI delay lines in each chip for fixed delays 20, 30, 40, 50 ns are presented in Table III.

VI. CONCLUSIONS

This paper presented the idea and design of the "output split inverter" delay gate (OSI) exhibiting lower impact of fabrication mismatch on timing parameters than the commonly used current starved inverter (CSI). The superior performance of the proposed circuit was achieved by inserting the current limiting transistors in between the switching ones, unlike in the case of the CSI circuit where these transistors are on the side of the power rails. Despite the similar operation of both circuits, the simulation results have shown significant differences in their dynamic behavior observed during the signal transitions which are of a high importance when process parameter variability is concerned. The analyses and the simulation results were confirmed by measurements of 512 16-stage CSI and OSI delay lines implemented on a test chip fabricated in a standard 90 nm CMOS technology. The experimental results have shown that the proposed OSI delay lines generate 10%–50% less variable delay intervals than the CSI ones with no penalty in terms of additional area, power or complexity increase.

REFERENCES

- J. Christiansen, "An integrated CMOS 0.15 ns digital timing generator for TDC's and clock distribution systems," *IEEE Trans. Nucl. Sci.*, vol. 42, no. 4, pp. 753–757, Aug. 1995.
- [2] P. Dudek, S. Szczepanski, and J. V. Hatfield, "A high-resolution CMOS time-to-digital converter utilizing a Vernier delay line," *IEEE J. Solid-State Circuits*, vol. 35, no. 2, pp. 240–247, Feb. 2000.
- [3] M. Bolt, E. Cantatore, M. Socha, C. Aussems, and J. Solo, "Matching properties of MOS transistors and delay line chains with self-aligned source/drain contacts," in *Proc. ICMTS*'96, Mar. 1996, pp. 21–25.
- [4] E. H. M. Heijne *et al.*, "LHC1: A semiconductor pixel detector readout chip with internal, tunable delay providing a binary pattern of selected events," *Nucl. Instr. Meth. A*, vol. 383, pp. 55–63, 1996.
- [5] E. Cantatore *et al.*, "Statistical analysis and optimization of delay line chains for pixel readout electronics," *Nucl. Instr. Meth. A*, vol. 395, pp. 318–323, 1997.
- [6] T. Toifl, R. Vari, P. Moreira, and A. Marchioro, "4-channel Rad-hard delay generation ASIC with 1 ns timing resolution for LHC," *IEEE Trans. Nucl. Sci.*, vol. 46, no. 3, pp. 139–143, Jun. 1999.

- [7] P. Mroszczyk and P. Dudek, "Trigger-wave collision detecting asynchronous cellular logic array for fast image skeletonization," in *Proc. ISCAS 2012*, May 2012, pp. 2653–2656.
- [8] G. Indiveri, E. Chicca, and R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 211–221, Jan. 2006.
- [9] R. Wang, G. Cohen, T. J. Hamilton, J. Tapson, and A. van Schaik, "An improved aVLSI axon with programmable delay using spike timing dependant delay plasticity," in *Proc. ISCAS 2013*, May 2013.
- [10] G. S. Jovanovic and M. K. Stojcev, "Current starved delay element with symmetric load," *Int. J. Electron.*, vol. 93, no. 3, pp. 167–175, Mar. 2006.
- [11] P. Mroszczyk and P. Dudek, "Tunable CMOS delay gate with reduced impact of fabrication mismatch on timing parameters," in *Proc. NEWCAS 2013*, Jun. 2013.
- [12] J. Shyu, G. C. Temes, and F. Krummenacher, "Random error effects in matched MOS capacitors and current sources," *IEEE J. Solid-State Circuits*, vol. 19, no. 6, pp. 948–956, Dec. 1984.
- [13] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.
- [14] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterization and modeling of mismatch in MOS transistors for precision analog design," *IEEE J. Solid-State Circuits*, vol. 21, no. 6, pp. 1057–1066, Dec. 1986.
- [15] P. R. Kinget, "Device mismatch and tradeoffs in the design of analog circuits," *IEEE J. Solid-State Circuits*, vol. 40, no. 6, pp. 1212–1224, Jun. 2005.
- [16] S. J. Lovett, M. Welten, A. Mathewson, and B. Mason, "Optimizing MOS transistor mismatch," *IEEE J. Solid-State Circuits*, vol. 33, no. 1, pp. 147–150, Jan. 1998.
- [17] A. Rodriguez-Vasquez, G. Linan, S. Espejo, and R. Dominguez-Castro, "Mismatch-induced trade-offs and scalability of analog processing visual microprocessor chips," *Analog Int. Circuits Signal Proc.*, vol. 37, no. 2, pp. 73–83, Nov. 2003.
- [18] M. J. M. Pelgrom, H. P. Tuinhout, and M. Vertregt, "Transistor matching in analog CMOS applications," in *IEDM'98 Tech. Dig.*, Dec. 1998.
- [19] C. J. Abel, C. Michael, M. Ismail, and C. S. Teng, "Characterization of transistor mismatch for statistical CAD of submicron CMOS analog circuits," in *Proc. ISCAS*'93, May 1993, vol. 2, pp. 1401–1404.
- [20] P. G. Drennan and C. C. McAndrew, "Understanding MOSFET Mismatch for Analog Design," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 450–456, Mar. 2003.

[21] C. Michael and M. Ismail, "Statistical modeling of device mismatch for analog MOS integrated circuits," *IEEE J. Solid-State Circuits*, vol. 27, no. 2, Feb. 1992.



Przemysław Mroszczyk received his five-year masters (mgr inż.) degree in electronic engineering from the University of Science and Technology (former AGH), Kraków, Poland, in 2010. He is currently working towards the Ph.D. degree at the University of Manchester (School of Electrical and Electronic Engineering), Manchester, UK. His research work is in the field of mixed-mode and asynchronous CMOS VLSI processor arrays for fast and power efficient computation with applications in image processing, Bayesian inference and machine learning.



Piotr Dudek (S'98–M'01–SM'06) received his mgr inż. degree in electronic engineering from the Technical University of Gdańsk, Poland, in 1997, and the M.Sc. and Ph.D. degrees from the University of Manchester Institute of Science and Technology (UMIST), Manchester, UK, in 1996 and 2000 respectively.

He worked as a Research Associate, and since 2002 as a Lecturer at UMIST/The University of Manchester. In 2009 he was a Visiting Associate Professor in the Department of Electronic and Com-

puter Engineering at the Hong Kong University of Science and Technology. Since 2011 he has been a Reader in the School of Electrical and Electronic Engineering, The University of Manchester, leading the Microelectronics Design Lab. His research interests are in the area of integrated circuit design, cellular processor arrays, vision sensors, novel computer architectures, and brain-inspired systems.

Dr. Dudek is a member of IEEE Circuits and Systems Society Technical Committees on Cellular Neural Networks and Array Computing, Neural Systems and Applications, and a Chair-Elect of the CAS Sensory Systems Technical Committee. He has been a member of scientific committees of several international conferences.