Reliability Enhancement of Low-Power Sequential Circuits Using Reconfigurable Pulsed Latches

Wael M. Elsharkasy, Member, IEEE, Amin Khajeh, Senior Member, IEEE, Ahmed M. Eltawil, Senior Member, IEEE, and Fadi J. Kurdahi, Fellow, IEEE

Abstract—Pulsed latches are gaining increased visibility in low-power ASIC designs. They provide an alternative sequential element with high performance and low area and power consumption, taking advantage of both latch and flip-flop features. While the circuit reliability and robustness against different process, voltage, and temperature variations are considered as critical issues with current technologies, no significant reliability study was proposed for pulsed latch circuits. In this paper, we present a study on the effect of different PVT variations on the behavior of pulsed latches, considering the effect on both the pulser and the latch. In addition, two novel design approaches are presented to enhance the reliability of pulsed latch circuits, while keeping their main advantages of high performance, low power, and small area. Experiments performed using Synopsys 28nm PDK demonstrate the ability of the proposed approaches to keep the same reliability level at different supply voltages and temperatures in the presence of process variations, with a very small area overhead of around 3%. The two proposed designs have negligible power overhead when running at nominal supply voltage, and they have higher yield per unit power when compared with the traditional design at different voltages and temperatures.

Index Terms—Pulsed latches, flip-flops, pulsed flip-flops, variability, process variation, voltage scaling, low power.

I. INTRODUCTION

FLIP-FLOPS are considered the most popular sequential elements used in conventional ASIC designs. This is mainly because of the simplicity of their timing model, which makes the design and timing verification processes much easier. Master-Slave Flip-Flops (MSFFs) are considered the most common and traditional implementations of flip-flops, due to its stable operation and its simple timing characteristics. However, the fact that the MSFF micro-architecture is usually built using two consecutive latches, it takes an appreciable portion of the clock period, power consumption, and area. A typical MSFF has a significant nominal timing overhead (sum of the clock-to-Q delay and the setup time) of 6 FO4 (fanout-of-4) and can reach 10 FO4 when considering clock skew and jitter [1]. In addition, the clock network, including the flops, often consumes one third to one half of the total dynamic power of the chip [2], [3]. In addition to the mentioned overheads associated with MSFF, some additional margins, which can reach up to 15% (depending on the sign off methodology), are usually added to the nominal timing margins to ensure correct operation under different process, voltage, and temperature (PVT) variations [4]. This, in turn, increases the already existing high timing and power overheads. For the above reasons, MSFF can be considered as a good choice for low-to-medium performance designs as they provide a good balance between delay, power, and easy design and verification processes for chips working at a relatively low frequency [5].

On the other hand, high performance custom designs tend to use latches due to their lower timing overhead that can reach 2 FO4 in some designs [1]. Although latch based designs are typically robust to clock skew and jitter (due to the latch transparency period), latches have a complicated timing model, which, in turn, complicates the design and the verification processes and increases the risk of hold time violations, especially with PVT variations.

To fill in the missing gap between MSFFs and latches, pulsed latches (sometimes called pulsed flip-flops) have been used in some high-performance designs [6]–[8]. Pulsed latches (PLs) are latches driven by short pulses generated from the normal clock signal using a pulse generator circuit called a pulser as shown in Fig. 1. The pulser can be either embedded in the latch, or can be separated as a standalone circuit as shown in Fig. 1. If the latter approach is used, a single pulser can be shared by more than one latch. Thus, it has the advantage of area and power consumption savings
over the former approach, and it is the focus of our discussion in this paper. In addition, the pulser usage can eliminate the need for some of the clock buffers used in the clock tree, thus providing an additional amount of power and area savings.

Having only one latch between the input and the output, PLs have lower timing overhead than MSFFs. At the same time, since the driving pulse is very short, the transparency period for the latch becomes very narrow, allowing the PLs to have a timing behavior close to that of MSFFs, to the extent that they are sometimes classified among flip-flop families [9], [10]. Also, due to the presence of the narrow transparent window of the latch, pulsed latches have an inherent tolerance to clock skew and jitter [2]. Since they have fewer transistors that are triggered by the clock signal, they have the advantage of reducing a significant amount of clocking power [8], and they consume much less leakage power compared to MSFFs due to the smaller area and fewer transistors.

One complication in PL design is the choice of the pulser output pulse width. Too short of a pulse width may not be enough for the latches to store the input data correctly, while too long of a pulse width will result in a longer latch transparency window; which, in turn, increase the timing overhead or can violate hold time requirements [11]. This issue becomes more complicated when considering different sources of variations. PVT variations have significant impacts on different circuit components. Since sequential elements, in general, are by nature time sensitive elements, they are among the circuit categories that are highly affected by any PVT variations [12]. Since pulsed latches, in particular, are very time sensitive, good study of the effect of different sources of variations has to be considered. Since some of these variations, such as voltage and temperature, are temporal variations that change over the operating period of chips, careful analysis and design has to be performed to ensure that reliable circuit operation can always be achieved without any significant loss in performance, power and area.

The study presented in this paper shows that some variation effects on PLs can be compensated with careful analysis during the design process, while some others can not be compensated without significant degradation in reliability or performance. As an example, pulsed latches designed to work at certain voltage corner may not work at another voltage corner without degradation in either performance or reliability. Since sequential elements such as pulsed latch are prolifically used within the die, any degradation in their performance or reliability can significantly affect the performance of the entire chip or can even cause a large yield loss. In addition, since designing a chip that can perfectly operate at just one voltage corner is not an acceptable solution nowadays, adding a reconfiguration ability to PL can help to reach the required target design goals. With this added feature, PLs can be configured to run with the minimum timing overhead to guarantee correct operation at different voltage levels in the presence of different sources of variation.

In this paper, we are presenting variability analysis of one of the popular topologies of pulsed latches, Transmission Gate Pulsed Latch (TGPL), studying the effects of process, voltage, and temperature variations, as well as proposing design modifications that can help in decreasing the probability of circuit failure (i.e. enhancing pulsed latch reliability) at different supply voltage values. With the proposed approaches, pulsed latches present a formidable alternative to MSFFs, providing higher performance, lower area and power consumption, and higher reliability and robustness to different kinds of variations.

The main contribution areas of this paper are:

- A study of the effect of PVT variations on the operation of pulsed latches in advanced technology nodes, considering the effects on both the pulser and the latches.
- Two novel pulse generator designs for the pulsed latch that can be utilized to increase the reliability of pulsed latch circuits, while keeping its main advantages of high performance and low power consumption.
- Comprehensive comparisons of reliability, power, and area between different data registers implemented using the traditional transmission gate pulsed latches and the proposed reconfigurable pulsed latches.

The remainder of the paper is organized as follows. Section II discusses some previous work in this area. Section III discusses PVT variations and their effects on pulsed latches behavior. Section IV discusses the two proposed design approaches to improve the reliability of pulsed latch under supply voltage scaling. Section V discusses the results obtained for the enhancement of circuit reliability, power, and area on a case study of a typical register. Finally, some conclusions are drawn in section VI.

II. PREVIOUS WORK

Pulsed latches have been always proposed to decrease power consumption and increase performance. In [13], PLs with relatively wide pulse widths were used to allow cycle borrowing and tolerate any clock skew. In order to compensate for any hold time issues, deracer circuits were used to block any incoming fast data before the end of the pulse. In [8], PLs were used as the main sequential elements to increase the performance of the Intel XScale microprocessor without consuming high clock power. Although the minimum pulse widths to ensure reliable operation across different PVT corners were used, delay buffers were still needed to decrease the risk of hold time violations. Baumann et al. [7] proposed three options for selectively replacing some of the MSFFs by PLs to improve the performance of the ARM926 microprocessor. However, some area and power overhead were presented due to buffer insertion. Baumann et al. [14] proposed replacing MSFFs by PLs in an ARM microprocessor. This was used to gain some performance improvement that was utilized as timing margins to compensate for the within-die variations.

In [6], a pulse generator with a pulse width control option was presented to enhance the reliability against Bias Temperature Stability (BTI). In [15], a traditional pulser and a latch formed by a tri-state inverter and a static keeper were proposed. The pulse width was chosen large enough to ensure correct operation under different PVT corners, assuming that delay cells can be used to fix any hold time issues. In [16], the impact of process variations on several pulsed flip-flops were presented, in addition to two techniques...
to reduce that impact. However, the effect of voltage and temperature were not studied and the proposed techniques were not quantified under these two effects. Dhong et al. [17] presented a novel pulser design whose output pulse width is determined by the voltage level at an input of a NAND gate instead of the delay chain of the traditional pulser of the TGPL. The paper showed that the proposed pulser is less affected by the clock rise time when compared with the traditional pulser at different supply voltages. However, the paper defined the failure criteria by the ability of the pulser to output a valid pulse, without quantifying the satisfaction of the pulse width for the needed latch transparency window in order to achieve successful writing.

In [9] and [10], comparisons in energy, delay and area between different flip-flop and pulsed latch classes and topologies were presented. The study showed that TGPL can be considered as the most efficient topology across wide range of applications when considering energy, delay and area trade-offs. In [18] and [19], the effects of PVT variations were studied on different flip-flops and pulsed latches topologies. The study showed also that TGPL has the highest performance and resiliency against process variation. In addition, it also showed that TGPL is still highly affected (as much as the other considered topologies) by voltage variations. However, it showed that the TGPL (pulsed topologies in general) has lower robustness against hold time violation under PVT variations. Although this can be solved by adding buffers, this results in some additional area and power overhead. In addition, the study didn’t quantify the effect of these variations on the design yield. Also, the studied voltage variations was limited to 10% only, while the temperature variations studied was only 20°C around 85°C.

As shown from the previous studies, TGPL, which will be our focus in this paper, is one of the most attractive architecture for PL circuits. However, there are still some challenges in the TGPL design (PL in general) to ensure reliable operation under PVT variations. In addition, a more comprehensive study of the probability of failure based on both the pulser and the latch is still missing. Although the study and proposed architectures presented in this paper will focus on TGPL, the same approaches can be applied to any other PL topology.

III. EFFECT OF PVT VARIATIONS ON PULSED LATCHES

The operation of PLs is based on enabling the latch for a short time using a pulse generated by the pulser circuit. Hence, to study the effect of variation on PL operation, variation effects on both the latch write time and the pulser pulse width should be studied. The effect of process variations is carried out for each of the latch and the pulser independently and the same study is repeated for different voltage and temperature values of interest.

A. Process Variations

Due to the extreme miniaturization of device parameters in current and upcoming technology processes, even a small variation in the manufacturing process may cause parameter variations that can lead to a failed circuit operation [20]. Thus, one of the significant challenges in the design phase is the ability to evaluate the effect of different sources of variations on the functionality of complex circuits and to provide circuit solutions to guarantee correct functionality under different sources of variations.

Process dependent sources of variability such as effective length variation, oxide thickness variation, Line Edge Roughness (LER), and Random Dopant Fluctuation (RDF) (for planar MOSFETs) result in variations in the value of the threshold voltages of transistors, which in turn impact the timing and power of digital circuits [21]. The threshold voltage variations due to RDF (which is usually the principal source of threshold voltage variations in planar MOSFETs) are considered as zero-mean Gaussian independent random variables with standard deviation denoted as \( \sigma_{V_{th}} \) which is given by [21]:

\[
\sigma_{V_{th}} = \sigma_{V_{th0}} \sqrt{\frac{L_{min}W_{min}}{LW}}
\]

where \( \sigma_{V_{th0}} \) is the \( \sigma_{V_{th}} \) for minimum sized transistor and it is given by:

\[
\sigma_{V_{th0}} = \frac{qT_{ox}}{\epsilon_{ox}} \sqrt{\frac{N_{d}W_{d}}{3L_{min}W_{min}}}
\]

where \( N_{d} \) is the effective channel doping, \( W_{d} \) is the depletion region width, \( T_{ox} \) is the oxide thickness, \( L_{min} \) and \( W_{min} \) are the minimum channel length and width, respectively.

While the scaling down of CMOS technology reduces the nominal supply voltage, the threshold voltages are not scaled by the same factor, leading to a significant reduction of the transistor’s available voltage headroom (the difference between the supply voltage and the threshold voltage). Hence, even any small variation in the transistor threshold voltage can lead to a significant degradation of the circuit behavior or can even cause complete circuit failure.

Studying the effect of process variations on PLs includes studying the effect of variations on both the latch write time and the pulser pulse width. As shown in Fig. 2, this is represented by the probability distribution functions (PDFs) of both the write time (Latch WR Time) and the pulse width (Pulser PW). To ensure correct write operation, the pulse width should be larger than the required transparent window for the latch (i.e. time needed
to capture the input data and pass it through the internal nodes to the storing cross coupled inverters). The area under the intersection between the two PDFs represents the failure of write operation, since this is the region where there is a high probability that the pulse width will be smaller than the time needed by the latch. Alternatively, knowing the information about the distribution of the latch write time and for easiness of timing analysis, a maximum value for latch write time can be calculated for certain sigma value of the designer choice. In this case, the probability of write failure can be calculated as the probability of having the width of the pulser output smaller than this desired maximum value. In both cases, depending on the target yield, the designer can determine the minimum acceptable value for circuit failure, and hence, the transistors’ dimension can be adjusted to reach the target yield.

B. Voltage Scaling

Voltage scaling is a popular run-time technique used for reducing the power consumption of circuits. It significantly decreases both dynamic power (with its two components of switching power and internal power) and leakage power. On the other hand, the ability to reduce the operating supply voltage is limited by a minimum value determined usually by some timing constrains (critical path delay as an example), in addition to some margins for the PVT variations, and usually adding a margin for aging effects [22]. As the supply voltage is scaled down, the available voltage headroom decreases further and the transistors become more sensitive to any variations.

The effect of voltage scaling is naturally associated with the increase of timing delays for different circuit components. While this can be handled at design time for several circuit components, the case may not be as easy for PLs. Since PL operation depends on two different components (pulser and latches) of different micro-architectures, the timing of each of them is affected differently. As shown in Fig. 3, voltage scaling affects the probability distribution of the pulser and the latch differently. Hence, failure probability calculated at one voltage will not be the same when applying voltage scaling. Indeed, PL reliability degrades significantly when scaling down the supply voltage. As shown in Fig. 4, the probability of write failure for a PL can increase by up to two order of magnitude when the supply voltage is scaled down by around 30%. Even if the PL circuit is designed to operate reliably at an intermediate supply voltage (0.9V as an example), the reliability will still significantly degrade at lower voltages, especially at low operating temperatures.

One possible solution is to design the PL circuit to operate with the needed level of reliability at the lowest possible operating voltage. Since chips usually operate at different supply voltages with different operating modes, when pulsed latches are operating at a voltage higher than that minimum value, they will be operating with extra timing margin (the pulse width will be larger than the needed width to achieve the required level of reliability). Hence, this will negatively affect one of the main advantage of PLs which is their low timing overhead, in addition to increasing the risk of hold time violations. The proposed circuit approaches in this paper will help in forcing PLs to reliably operate with just the needed timing margins at different supply voltages. Hence, this will assure gaining the maximum benefits of pulsed latches at different operating modes without any unnecessary waste in the design performance.

C. Temperature Effect

Studying the effect of temperature variation on the design is very important. Not only does the variation in temperature affect leakage power and performance, but it also affects the probability of having an error during circuit operation, as well as impacting the life span of different chip parts [23]. Factors such as the increase of leakage power with technology process scaling, the nonequivalent down scaling of the supply voltage when compared to geometry scaling, and the increase in the dynamic power associated with the increase in performance required in current designs, all lead to the increase of the operating temperature. Careful study of the effect of temperature is required especially for time-sensitive sequential elements.

The study of temperature effects on pulsed latches is much more critical, since each of the pulser and the latches can have a different response to temperature variation. The study done in this paper shows that both circuits become more sensitive to process variation with the decrease in temperature.
However, the pulser is more significantly affected by temperature variations. In addition, the entire PL would have high failure rates when operating at a lower temperature.

When running at nominal supply voltage, the transistors become faster as the temperature decreases. Since the pulser is more timing sensitive than the latch, the timing margin between the latch write time and the pulser output pulse width will decrease with the decrease in temperature as shown in Fig. 5. Hence, the probability of write failure is expected to increase with the decrease of temperature.

When scaling down the supply voltage beyond certain limit, the relation between the circuit delays and temperature is revered due to temperature inversion effect [24]. Again, due to the higher sensitivity of the pulser over the latch, the time margin increases as temperature decreases. However, the process variation effect increases significantly with the decrease in temperature. As shown in Fig. 6, the standard deviation for the latch distribution is doubled with the temperature decrease from 125°C to −40°C, while the standard deviation for the pulser distribution increases by more than 60%. This significant increase in the variation of the latch and pulser timing with the decrease in temperature leads to the increase of the probability of write failure.

Therefore, for different supply voltages, even when taking temperature inversion into account, pulsed latches become less reliable at lower temperature. Hence, to ensure correct operation, pulsed latches should be designed at the lowest operating temperature expected for the system. This can be a little different from one process technology to another, but similar results can be expected, at least for planar CMOS devices at closer process nodes.

IV. PROPOSED DESIGN APPROACHES

The basic structure of a conventional pulser of the TGPL is shown in Fig. 7, where the delay unit, usually consisting of an even number of inverters, is responsible for determining the width of the needed output pulse. To guarantee correct operation, the pulser is designed to generate a pulse width that is larger than the latch write time.

As described in the previous section, it is not easy to design a non-configurable pulsed latch circuit that can operate with just the needed timing margins at different supply voltages in the presence of process and temperature variations, while keeping the needed level of reliability. To be able to reach the needed reliability level, the pulser circuit should be reconfigured at run time to generate an output pulse whose width can be controlled based on the operating condition. Shown in Fig. 8(a) is a pulsed latch circuit designed to operate correctly at nominal supply voltage with high level of reliability. However, when scaling down the supply voltage as shown in Fig. 8(b), the circuit become less reliable with higher probability of failure. The required level of reliability can be achieved at the lower supply voltage by increasing the width of the generated pulse. As shown in Fig. 8(c), this is equivalent to shifting the pulser probability distribution to the right, compensating for the increased variation effects at lower voltages and therefore, decreasing the probability of circuit failure.

In this section, two design approaches are proposed. Both approaches depend on controlling the delay path (the delay
unit and its following inverter) of the pulser circuit by using an external control signal (CTRL) to generate a controllable pulse width. The first approach considers splitting the supply rail of the pulser circuit, and applying an additional controllable level of voltage scaling on the delay path when needed. The second approach relies on using multiple delay units in the pulser circuit and choosing a certain delay unit at run-time according to the operating condition. Detailed discussions of the two approaches are presented in the next two subsections.

A. First Approach

This approach is based on using a virtual supply rail for the delay path of the pulser, driven from the main supply rail used for the rest of the pulser circuit and the latches. This can be accomplished using header PMOS switches for the delay path of the pulser circuit, similar to the local power gating topology [25] as shown in Fig. 9, where turning off some of these switches will result in lowering the supply voltage of the delay path. Since this delay path is the main part of the circuit that control the width of the generated pulse, controlling the supply voltage of this path will result in controlling the output pulse width.

Separate control signals can be used for different switches, where at least one of these switches must be always turned on (i.e., the gate of this PMOS switch should be tied to the ground) giving the maximum output pulse width, while the other switches can be turned on or off to achieve the required narrowing of the pulse width. The number of these parallel switches and their sizes will depend on the number and values of the virtual supply voltage levels, which corresponds to the needed pulse widths to achieve the target reliability level at different operating conditions. Since the delay chain current represents only 20-30% of the total pulser current, the sizes of these PMOS switches should be reasonable, adding a small area overhead to the pulser circuit.

During normal operation, when required to operate at the nominal supply voltage, all header switches are on, driving the whole pulser circuit by nearly the same supply voltage value.

When scaling down the supply voltage, the needed margin for variations in the latch write time increases. By turning off part of the pulser header switches, an additional down scaling of the virtual supply of the delay path (VDI) is provided; i.e., the delay unit and its following inverter is running at a slightly lower supply than the rest of the pulser circuit. This additional voltage scaling of VDI will result in a small increase in the pulser output pulse width. Since the circuit is already operating with small voltage headroom at this lower supply voltage, a very small decrease in VDI will be sufficient to produce an adequate increase in the pulse width without having a significant difference between the supply voltages of the delay path and the rest of the pulser circuit. In addition, the remaining pulser circuit (the NAND gate and the output inverter) will act as a voltage level shifter, driving the latches by the same voltage level as their supply voltage.

B. Second Approach

Since the pulse width depends on the delay unit as shown in Fig. 7, implementing multiple delay units with different delays can help in generating pulses with different widths. One important design consideration is the ability to choose between these different units post silicon or at run-time. The second proposed pulser design is shown in Fig. 10. Each delay unit represents a buffer chain that can be implemented in different ways. It can be as simple as a very small delay unit (i.e., just a wire) and up to multiple even number of inverters of different inverter sizes and/or numbers. The output of the multiplexer is used to drive an odd number of inverters, whose final output is connected to the NAND gate.

By selecting a longer delay chain, the latch transparency window can be increased at run time, which is required when scaling down the supply voltage. The shortest delay unit is designed such that, when operating at a nominal supply voltage, the circuit is verified to run with very low probability of failure in the presence of different process and temperature variations. The rest of the delay units are designed depending on the number and values of the supply voltage scaling levels.

V. Results

To verify the proposed approaches, test circuits of 16-bit register were examined and three implementation choices were compared. The three implementations consists of a single pulser driving sixteen identical latches similar to that shown in Fig. 1. The first choice is the implementation using the traditional non-configurable pulser shown in Fig. 7. The pulser was designed at nominal supply voltage to ensure the required reliability level. The second and the third choices are the two proposed pulser implementations, also driving sixteen identical latches. The effect of voltage scaling of one scaling level was applied on all circuits. An extreme value of voltage scaling which is usually around 30% reduction from nominal supply value was used to show the effectiveness of the proposed approaches. The same approaches can be easily extended to any other scaling values.
The control signal (CTRL) used to select the required pulse width can be generated by one of two methods. In a system with two supply rails, the CTRL signal can be driven by the same control signal used to switch the power gates used to choose the operating supply rail. For a system with a single supply rail, a circuit can be designed to generate an output voltage that is dependent on the value of the VDD. For the proposed register circuits, the circuit in Fig. 13 was designed. The first stage of this circuit is a voltage divider of the supply voltage that generates an output a little less than half the VDD. The output of this voltage divider is used to drive a pseudo-NMOS inverter circuit. The pull down network (PDN) of this inverter is a strong NMOS, while the pull up network (PUN) is built using a weak PMOS in parallel with a weak NMOS. The presence of both NMOS and PMOS in the PUN is to ensure reliable circuit operation at different process corners, especially the fast-NMOS slow-PMOS (FS) corner. The output of this pseudo-NMOS inverter passes through few regular CMOS inverters to generate the final CTRL signal. At nominal supply voltage, when the value of VDD is high, the output of the voltage divider will be high enough to strongly turn on the PDN NMOS of the pseudo-NMOS inverter, generating a lower voltage value at its output. Since the PUN of the circuit has weak NMOS and PMOS devices, the generated output of this pseudo-NMOS inverter will be low enough to be interpreted by the following CMOS inverter as a logic ’0’. At scaled-down voltage, when the value of VDD is low, the output of the voltage divider will be low enough to make the NMOS of the pseudo-NMOS PDN hardly on. Hence, the always-on PMOS and NMOS of the PUN will generate a high enough voltage to be interpreted as a logic ’1’ by the following CMOS inverter. The regular CMOS inverters are used to adjust the voltage levels of the CTRL signal and to generate the needed voltage polarity. Since this circuit can be shared between different resisters in a design block, any area or power overhead associated with it can be negligible. In addition, the same circuit can be repeated to generate additional control signals (for multiple supply voltage scaling levels), where the switching threshold for each circuit can be controlled by adjusting the sizes and the threshold voltages of each transistor.

The experiments were carried out using the Synopsys 28nm PDK [27], [28]. All of the implementations were examined at the same frequency of 1 GHz and the rise time of the input clock signal was chosen to be 50ps. The nominal supply voltage used was 1.05V and the same level of supply voltage scaling to 0.7V was applied. The analysis for the effect of
The probability of write failure for different register implementations at different supply voltages and different temperatures is shown in Table I. The simulations were done using Hspice and the variability analysis was carried out using Solido Variation Designer [12] and Matlab. All the simulations and analysis were carried out on the post-layout extracted circuits. The power numbers were calculated over a few hundred cycles of common activity levels for the different register implementations. The area calculations were carried out through layouts drawn with Synopsys Custom Designer and verified with Synopsys IC Vaildator using the same 28nm technology. The layouts of proposed PL-SW pulser and PL-MUX pulser are shown in Fig. 14. For simplicity, the shown layouts are for the pulser circuits only. However, the reported areas were calculated on the complete 16-bit register layout, which include the 16 latches in addition to some buffers.

Since sequential elements are replicated in large numbers in a chip die and across the entire wafer, a wide range of up to ±6σ of process variation was considered in the variability analysis [12], where the High-Sigma Monte Carlo (HSMC) tool of the Solido Variation Designer was used to do the variation analysis.

A high design yield that is higher than 99% was chosen. Hence, a target probability of write failure ($P_{WF}$) that is less than $1 \times 10^{-8}$ was set using [29]

$$\text{Yield} = (1 - P_{WF})^n$$

(3)

where $n$ was chosen to be in the order of $10^6$ cells.

The probability of write failure $P_{WF}$ can be calculated as the probability of having a pulser pulse width $PW$ smaller than an estimated maximum value for the latch write time ($T_{lw-\text{max}}$). This $T_{lw-\text{max}}$ is calculated as the latch write time ($T_{lw}$) at 6σ of the write time distribution (i.e., $T_{lw-\text{max}} = \mu_{lw} + 6\sigma_{lw}$, where $\mu_{lw}$ is the mean of the distribution and $\sigma_{lw}$ is the standard deviation). Hence, $P_{WF}$ can be calculated as

$$P_{WF} = P(PW < T_{lw-\text{max}}) \times P(T_{lw} > T_{lw-\text{max}})$$

(4)

A. Reliability Analysis

Since the traditional PL circuits do not have any configuration ability, they were designed to ensure correct functionality at nominal supply condition. Correct functionality means that the pulsed latch circuit can achieve the target level of reliability (i.e. target probability of failure) when running at nominal supply at the entire range of temperature values in the presence of process variation.

When the supply voltage is scaled down, the traditional PL register becomes more susceptible to write failure. As shown in Table I, the $P_{WF}$ for the traditional PL register is within the required value at nominal supply voltage. However, $P_{WF}$ increases when the supply voltage is scaled down to 0.7V. Hence, the circuit becomes much less reliable at this lower voltage. The latch $T_{lw-\text{max}}$ and the pulser PWs at different supply voltages are shown in Fig. 15 for the PL-SW design. The typical values for the PW are chosen to be the mean of their distributions, while the minimum values for the PW are arbitrary chosen to be 3σ lower than the mean. The short configuration of the PL-SW is nearly the same architecture as the traditional PL. When the supply voltage decreases, the latch $T_{lw-\text{max}}$ starts to move away from the typical short PW and closer to the minimum short PW, increasing the probability of write failure as shown in Table I. If the pulser can be configured to generate longer PW at lower supply voltages, the reliable timing relation between the $T_{lw-\text{max}}$ and
the pulser PW can be regained, where the $T_{wr\_max}$ becomes closer to the typical PW, lowering the probability of write failure.

For both proposed designs, PL-SW and PL-MUX, the required reliability levels can be achieved at different supply voltages, within the entire temperature range in the presence of process variations without adding any unnecessary timing overhead. With the added reconfiguration ability, both designs were independently characterized to function properly at the two different operating voltages (nominal and 30% down-scaled voltages). This was used to adjust the sizes of the header switches for the PL-SW design, and adjust the design of the delay units for the PL-MUX design.

When running at the nominal supply voltage, all of the header switches are turned on for the PL-SW design, while the multiplexer is switched to the short delay unit for the PL-MUX design. When scaling down the supply voltage, the two proposed approaches depend on increasing the pulse width. This is accomplished by turning off one of the switches for the PL-SW design or switching to the long delay unit for the PL-MUX design using the CTRL signal generated by the circuit shown in Fig. 13. As shown in Fig. 16 for PL-SW and Fig. 17 for PL-MUX, this results in shifting the probability distribution of the pulser output to the right, compensating for the increased variation effects at lower voltages, and hence, decreasing the probability of circuit failure as shown in Table I. Therefore, the required high level of reliability at different voltages is obtained at the cost of a very small overhead in area and power.

In addition, to ensure a reliable operation of the circuit that generates the CTRL signal, the circuit was tested at different process corners and at the wide temperature range from $-40^\circ C$ up to $125^\circ C$. In addition, the correct operation was verified at 5% lower than the 1.05V and at 5% higher than the 0.7V to ensure tolerance to any variations in the supply voltage regulator and the power delivery network.

B. Power and Area Comparison

Since power consumption is an important metric for such circuits, any power overhead associated with the proposed approaches should be minimized. Fig. 18 shows the average energy per cycle for different register designs normalized to that of the traditional PL register at the 1.05V and 125°C. During normal operation, when running at a nominal supply voltage, both the PL-SW register and the PL-MUX are nearly consuming the same amount of energy as the traditional PL register.

When scaling the supply voltage down to 0.7V, both the PL-SW and PL-MUX registers seem to consumes more power. However, the traditional PL register has higher probability of failure at this lower voltage, which make its energy numbers of no sense unless that working at this low level of reliability is acceptable. Comparing the two proposed approaches, PL-MUX register consumes between 9% to 14% more power than the PL-SW register due to the additional switching delay units.

Regarding the area overhead, each approach has added some transistor to the pulser circuit, however, the area overhead of the added circuits is very small. The overhead in area...
of the PL-SW register compared to the traditional PL one is only 2.4%, while the area overhead of the PL-MUX register is 3%.

C. Discussion

This paper presents an analysis of the effect of process and temperature variations at different supply voltages on the reliability of the traditional TGPL. This analysis includes the effect on both the pulser and the latch, in addition to the relation between them. TGPL was chosen as it is one of the most efficient PL topologies. Without any configuration ability, the TGPL should be designed to operate with high reliability at the lowest supply voltage (the pulse width is increased to ensure correct operation with low failure probability). This will result in much wider pulser PW when the circuit is operating at higher voltages. Hence, this adds unnecessary extra timing overhead, in addition to increasing the chances of hold time violations. These hold time violations can be solved by adding some delay buffers. However, this will degrade some of the advantages of PLs which are their lower sequential and clock network power. Therefore, reconfigurable PLs are needed. In addition, any overhead associated with adding the configuration ability should be minimized.

The two proposed approaches have an advantage of keeping the required level of reliability at different supply conditions with the minimum power and area overhead when compared to the static traditional PLs. At the same time, each of the two approaches is running with just the needed margins at different supply voltage values, hence, minimizing the timing overhead. To evaluate the gained benefits from the proposed designs when compared to the added power overhead, the design yield, calculated using equation 3, per unit power is calculated for the three different register implementations. The results of the design yield per unit power for the two proposed implementations when normalized to that of the traditional PL register at the same voltage and temperature are shown in Table II. When running at nominal supply voltage, the proposed designs are running as reliable as the traditional PL register with negligible power overhead at different temperatures. When the supply voltage is scaled down, only the two proposed designs can keep the same level of reliability with a very small power overhead at different temperatures. Hence, the two proposed designs show great advantages over the traditional PL register at the entire range of voltages and temperatures. Although the two proposed approaches was used for the TGPL, both approaches can be easily used with any other PL topology whose pulser uses a delay chain to control the generated pulse width.

In addition, the two proposed designs are not significantly affected by the clock rise time. This can be shown in Fig. 19 for PL-SW when running 1000-sample Monte Carlo simulations for different clock rise time. Also, an output pulse was generated successfully with no failure at both 1.05V and 0.7V when running 1000-samples of Monte Carlo simulations while varying the clock rise time up to 150ps. Similar results were also obtained for the PL-MUX design. This makes the two proposed designs very comparable to the design proposed in [17] for the range of supply voltage studied in this paper.

When compared to similar studies presented in the literature, the approaches proposed in this paper have some advantages. In [16], only the effect of process variations on PL was studied and the two proposed techniques were not quantified under the effect of temperature and voltage variations. The study in [18] and [19] considered the effect of PVT variations, but it didn’t quantify the effect of these variations on the design yield. In addition, the covered temperature variation range was very limited and didn’t study the entire range covered in this paper. In [15], the pulser used for the proposed PL is similar to the traditional pulser studied in this paper, but it was designed to generate wider pulse for correct operation at different voltages. The paper reported the usage of delay cells to fix hold time issues. Therefore, with the very small area overhead of PL-SW and PL-MUX and the elimination of the need to delay buffers, our two proposed approaches are expected to save area when compared to the one proposed in [15]. Since PL-SW doesn’t add any significant power overhead, power saving are also expected (PL-MUX can also save power if large number of delay cells must be used). Similarly, [7], [8] reported the need of delay buffers to fix hold time violations, which could be eliminated to save power and area by using the proposed reconfigurable approaches.

Comparing the results of the two proposed approaches, each has advantages and drawbacks. While the PL-SW design is

<table>
<thead>
<tr>
<th>T</th>
<th>( V_{DD} = 1.05 \text{ V} )</th>
<th>( V_{DD} = 0.7 \text{ V} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PL-SW Register</td>
<td>PL-MUX Register</td>
</tr>
<tr>
<td>125°C</td>
<td>1.03</td>
<td>1.01</td>
</tr>
<tr>
<td>25°C</td>
<td>1.04</td>
<td>1.02</td>
</tr>
<tr>
<td>-40°C</td>
<td>1.04</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Fig. 19. Box plot of pulse width of PL-SW at 0.7V and 25C over 1000-samples Monte Carlo simulation.
smaller in area and requires less power, the design of the power switches is more complicated specially if several control level is needed. In addition, when all the switches are ON (i.e. running at nominal supply voltage), there will still be a small voltage drop on the power switches. Hence, the delay unit and its consecutive inverter will still run at a slightly lower voltage than the rest of the pulser circuit, generating a pulse width slightly larger than expected. One possible solution is to increase the sizes of the switches, however, this will increase the pulser area.

On the other side, the PL-MUX is simpler and easier in design. However, its power overhead is much larger due to the increase of dynamic power with the extra inverters. In addition, the area and power will exponentially increase with each additional voltage scaling level, due to the additional delay units and the larger multiplexer.

Hence, for designs with few voltage scaling levels, the PL-MUX design can be preferable over the PL-SW one, as it is easier in design, generates more precise pulse widths, and its overheads (power and area) are reasonable. On the other hand, for designs with large number of voltage scaling levels, the PL-SW design is preferred, as the area and power overheads of the PL-MUX design will be significant.

VI. CONCLUSION

In this paper, an analysis of the effect of PVT variations on the pulsed latch performance was presented. The analysis considered both the pulser and the latch to evaluate the reliability of the entire pulsed latch circuit. In addition, the benefits of having a reconfigurable pulsed latch circuit was discussed. Two novel modifications to add the reconfiguration ability to TGPL circuits were proposed.

The benefits of using the proposed design approaches in enhancing the robustness of pulsed latch circuits at different supply voltages were demonstrated using 16-bit registers. Both proposed approaches were able to ensure reliable operation of the pulsed latch-based register under different supply voltages in the presence of process and temperature variations, without any unnecessary timing overhead. Both approaches have a very small area overhead of around 3% or less. In addition, the power overhead of both approaches is minimal when compared to the traditional pulsed latch based register at the same reliability level. Both approaches are easily scalable to cover different levels of voltage scaling. In addition, they can be applied to any other pulsed latches topology that depends on a delay path to generate the output pulse.

ACKNOWLEDGMENT

The authors gratefully acknowledge the staff of Solido Design Automation Inc. for their support with the Solido Variation Designer tool.

REFERENCES


Wael M. Elsharkasy (S’04–M’14) received the B.Sc. and M.Sc. degrees in electrical engineering from Alexandria University, Egypt, in 2007 and 2011, respectively. He was an Intern with the Power Optimization Center of Excellence, Broadcom Corporation, from 2014 till 2016. He is currently pursuing the Ph.D. degree in electrical engineering and computer science with the University of California at Irvine, Irvine, CA, USA. His research interests include low power design for SoCs, variation-aware design of digital integrated circuits, and low power reconfigurable architectures.

Ahmed M. Eltawil (S’97–M’03–SM’14) received the Ph.D. degree from the University of California at Los Angeles, Los Angeles, in 2003. Since 2005, he has been with the Department of Electrical Engineering and Computer Science, University of California at Irvine, Irvine. He is the Founder and the Director of the Wireless Systems and Circuits Laboratory. His current research interests are in low power digital circuit and signal processing architectures for wireless communication systems. He received several distinguished awards, including the NSF CAREER Award in 2010 supporting his research in low power wireless systems.

Amin Khajeh (S’01–M’11–SM’14) received the B.Sc. degree in electrical engineering and communication from Shiraz University, Iran, in 2002, and the M.Sc. degree in electrical engineering from The University of Texas at Arlington, TX, USA, in 2005, and the Ph.D. degree in EECS from the University of California at Irvine, Irvine, CA, USA, in 2010. He was an Intern in research and development division of Siemens Company in 2002, and he later joined Siemens as a Research Staff Member from 2002 to 2003. He was with Qualcomm low power DSP Team from 2010 to 2012, where he was involved researching and implementing advance low power techniques for DSP cores for wireless multimedia applications and the Circuit Research Lab, Intel Research Labs, from 2012 to 2014, where he was involved in low power SoC design. He is currently a Principal Scientist with Broadcom Limited, where was involved in low power design and methodology for network switches. His research interests include low power design and methodology for SoCs, design of low power high performance circuits, and high performance high yield memory design, where he has authored over 30 technical papers and holds five patents on these subjects.

Fadi J. Kurdahi (F’05) received the B.E. degree in electrical engineering from the American University of Beirut in 1981 and the Ph.D. degree from the University of Southern California in 1987. Since then, he has been a Faculty with the Department of Electrical Engineering and Computer Science, University of California at Irvine, where he conducts research in the areas of computer aided design, high-level synthesis, and design methodology of large scale systems, and serves as the Director of the Center for Embedded & Cyber-physical Systems, comprised of world-class researchers in the general area of Embedded and Cyber-Physical Systems. He is a fellow of the AAAS. He was the Program Chair or the General Chair on program committees of several workshops, symposia, and conferences in the area of CAD, VLSI, and system design. He received the best paper award of the IEEE TRANSACTIONS ON VLSI in 2002, the best paper award in 2006 at ISQED, and four other distinguished paper awards at DAC, EuroDAC, ASP-DAC, and ISQED. He also received the Distinguished Alumnus Award from his Alma Mater, the American University of Beirut, in 2008. He served on numerous editorial boards.