# Variability-Aware Design of Energy-Delay Optimal Linear Pipelines Operating in the Near-Threshold Regime and Above

Qing Xie, Yanzhi Wang, and Massoud Pedram
University of Southern California
Department of Electrical Engineering
Los Angeles, California, United States, 90089
{*xqing, yanzhiwa, pedram*}*@usc.edu*

## ABSTRACT

Soft-edge flip-flop based pipelines can improve the performance and energy efficiency of circuits operating in the super-threshold (supply voltage) regime by allowing opportunistic time borrowing. The application of this technique to near-threshold regime of operation, however, faces a significant challenge due to large circuit parameter variations that result from manufacturing process imperfections and substrate temperature changes. This paper thus addresses the issue of variability-aware design of the energy-delay optimal linear pipelines that are aimed at operating in both the near-threshold and super-threshold regimes. Precisely, this goal is achieved by deriving the optimal delay line configuration in the soft-edge flip-flops in the near-threshold and the super-threshold operations regimes. The key is to ensure that the same transistor sizes result in effective operation of the delay lines (and hence appropriate settings of the transparency window size) in both operation regimes under the process induced variations. Experimental results demonstrate the efficacy of the proposed solution.

### Categories and Subject Descriptors

B.6.1 **[Hardware]:** Logic Design – *sequential circuits*

### General Terms

Design, Performance

### Keywords

Energy-delay optimal pipeline design, soft edge flip-flop, time borrowing, near-threshold, ultra-low voltage, parameter variability

## 1. INTRODUCTION

With the increase in demand for battery-powered devices and wireless equipments, the need for energy-efficient design gains growing attention. The ultra-low voltage operation, in particular, *near-threshold* (NT) operation, is very effective in minimizing energy consumption by reducing supply voltage to near the threshold voltage $V_{th}$ with a sacrifice of the timing performance of the circuits [1][2]. It is especially beneficial for those applications with relaxed timing requirements, such as medical monitoring and building health monitoring. Previous work on NT operation proved the existence of the *minimum energy* (*operation*) *point* (MEP), which is the optimal supply voltage level to minimize energy consumption, and derived the MEP voltage analytically for some integrated circuits [3][4]. Although offering promising energy reduction, one of the major concerns of NT operation is that digital circuits operating in the NT regime are very sensitive to the process variation. For example, $(3\sigma/\mu)$ delay variation of a combinational logic block at 0.5 V increases by a factor of 2.5 compared to that at 1 V using 90nm technology [5].

Pipelined data path in a modern processor is a major contributor to the total energy consumption of the processor. Many techniques such as pipeline gating [6], clock gating [7], and voltage scaling [8] are proposed to reduce the energy consumption in traditional *super-threshold* (ST) regimes. In the NT regime, the pipelined circuits should be designed not only energy-efficient, but also variability-tolerant. The authors in [11] developed a statistical methodology to enhance the yield of the pipelined circuits, considering the inter-die and intra-die variation. Recently, a two-phase latch-based design strategy is proposed for the synchronous pipelined circuits in the NT operation regime to derive the optimal ratio between the total width of sequential logic and that of combinational logic [10]. However, the latch-based design has many limitations including hold time violation issues, design difficulties using standard EDA tools, and requirement of an extra clock network, which is power and area inefficient.

An SEFF is the same as a D-flip-flop except that a *delay line* (DL) is added to postpone the clock edges of the master latch to create a *transparency window*, during which both master and slave latches are turned on. Using the transparency window, the SEFFs can pass the slack time across the pipeline stages. However, the DLs mentioned above consume extra amount of energy. Authors in [9][12] showed that the larger *transparency window size* (also known as *softness*) is, the more energy overhead it brings. Applying soft-edge flip-flops (SEFFs) is a useful technique to combat the process variation and improve the operating frequency. The authors of [12] proposed to utilize soft-edge flip-flops in sequential circuits to improve the yield in the presence of process variation. In [9], the authors proposed an SEFF-based pipeline design methodology jointly considering the voltage scaling and time borrowing, and demonstrated noticeable reduction in the *energy-delay product* (EDP) in the ST regime.

In this work, we present an SEFF-based pipeline design methodology for both ST and NT operation regimes targeting at minimizing the EDP. The stage delay increases exponentially with the supply voltage of pipelined circuits in the NT regime. However, we show that the SEFF-based pipeline designed and optimized in the ST regime [9][12] cannot provide enough softness to achieve the minimum EDP in the NT regime due to the delay variation in delay lines embedded within the SEFFs. To design a single delay line that can minimize the EDP in both the ST and NT regimes, we present a novel delay line structure. Precisely, we add a PMOS header on top of the traditional delay line (beneath the supply voltage rail). The header PMOS results in a slight supply voltage drop on the delay line, which is negligible in the ST regime but has a significant impact on the transparency window size in the NT regime. In this way, the proposed

This research is sponsored in part by grants from the Defense Advanced Research Projects Agency and the National Science Foundation.


*GLSVLSI*'13, May 2–3, 2013, Paris, France.

pipelined circuits can provide appropriate softness in both ST and NT regimes and subsequently achieve the optimal EDP.

We formulate the EDP minimization problem for the SEFF-based pipelined circuits as a mathematical programming problem, where the clock period, sizes of the PMOS header, and configurations of the delay lines are the optimization variables. We construct a Pareto-optimal energy/power consumption vs. flip-flop softness tradeoff curve with respect to PMOS header size and the delay line configuration. We take into account the delay variation of both combinational and sequential logics, and derive the joint distribution of clock-to-q delay, setup time, stage delay, and flip-flop softness. The timing constraints of pipelined circuits are enforced by using the $3\sigma/\mu$ delay. Experimental results show that the reduction in the EDP is 14% and 16.5% in the ST and NT operation regimes, respectively.

# 2. SEFF-BASED PIPELINE

## 2.1 Soft-Edge Flip-Flops

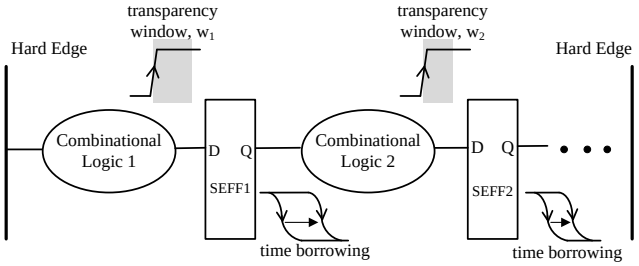


**Figure 1. A linear pipeline with soft-edge flip-flops.**

Figure 1 depicts a general synchronous SEFF-based linear pipelined circuit. Considering the data consistency between the SEFF-based pipeline and the input and output environments, we impose hard boundary conditions using traditional hard-edge flip-flops at the beginning and end of the pipelined circuits. Between the two hard edges, the pipelined circuit has multiple combinational logic stages whose delay is affected by process, voltage, and temperature (PVT) variations. We build the stage register using the SEFFs. The key idea is to postpone the clock signal for the master latch to create a transparency window, which enables time borrowing across the pipeline stages.

The SEFF-based pipeline has the advantages of higher operating frequency and variability tolerance due to the following two reasons. First, the operating frequency of a traditional hard-edge pipeline is always limited by the delay of the critical stage. While in SEFF-based pipelines, the slack time from the non-critical stages can be effectively passed to the critical stage so that the overall performance can be improved [2]. Second, we utilize the fact that the local random process variations are alleviated for deeper combinational logics since the variations cancel out when the logic depth increases [5][12]. Thus, the soft edges between the stages reduce the sensitivity of the circuits on process variations.

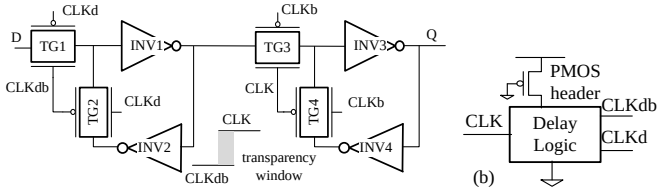


**Figure 2. Design of the positive-edge triggered soft-edge master slave flip-flops (a) and the proposed delay line (b).**

The SEFFs come with the price of extra amount of power consumption. As shown in Figure 2, the DLs are typically a series of properly sized inverters. We tune the softness, which is the delay of DL, through sizing or adding/removing the inverters in DL. In general, the power consumption of the SEFF satisfies a positive relationship versus the softness in both ST and NT regimes, as shown in Figure 3. We fit the power consumption of an SEFF versus softness using a linear function.

SEFF also affects the timing conditions the pipeline stages. Intuitively, all three *critical times*: the setup time, hold time and clock-to-q delay, are postponed by the transparency window size, denoted by $w$. Since the data can be captured at the end of the transparency window, the setup time decreases by $w$. The hold time increases by $w$ since the data needs to be stabilized during the transparency window. In the worst case, the clock-to-q delay also increases by $w$ since the input data may come at the end of the transparency window. Thus, we have the modified timing conditions for the $i$-th stage of the SEFF-based pipeline as follows

$$\begin{aligned} t_{cq,i-1} + w_{i-1} + D_{max,i} + t_{se,i} - w_i \le T_{clk}, \\ t_{cq,i-1} + D_{min,i} \ge t_{h,i} + w_i \end{aligned} \quad (1)$$

where $D_{max,i}$ and $D_{min,i}$ are the worst-case and best-case delay of the $i$-th combinational logic, respectively. $T_{clk}$ is the clock period of the pipeline.

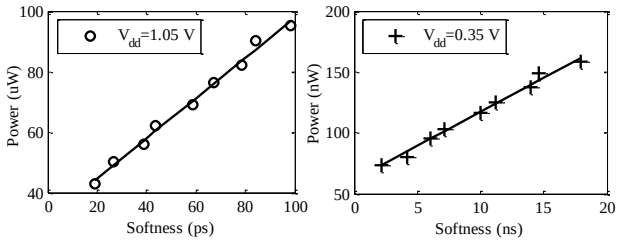


**Figure 3. Power consumption of an SEFF versus softness.**

## 2.2 Energy-Delay Product

*Energy per throughput* is typically used as the metric of the pipelined circuits. The throughput is defined as the number of observed output data divided by the observation duration, which is the product of number of clock cycles and the clock period,

$$thruput = \frac{\#\ of\ output\ data}{\#\ of\ clock\ cycles \times T_{clk}} \propto \frac{1}{T_{clk}}. \quad (2)$$

For a pipeline in steady-state, throughput defined in (2) is proportional to the inverse of the clock period. To account for energy consumption, we define the *energy per throughput* as,

$$\frac{Energy}{thruput} \propto E_c \cdot T_{clk}, \quad (3)$$

where $E_c$ is the total energy consumption per clock cycle of the entire pipeline. We use the *energy-delay product*, $E_c \cdot T_{clk}$, as the cost function in this work.

Although the SEFFs consume extra energy, especially for large transparency windows, they can be utilized to improve the energy efficiency if properly designed. The reason is that the leakage energy per clock cycle decreases as the clock period is reduced. This is extremely helpful for pipelined circuits operating in the NT operation regime due to the fact that the leakage energy consumption plays a more important role in the NT operation regime [1][2], as shown in Figure 4.

Figure 4 depicts the leakage energy consumption per cycle (here a cycle is defined as the worst-case delay of that combinational circuit at that voltage level) for some selected ISCAS'85 benchmarks at different supply voltage levels, under the 32/28nm technology [13]. The circuit delay increases exponentially and the leakage power decreases linearly when the supply voltage drops [1][2] thus the leakage energy consumption increases. In contrast, the dynamic energy consumption drops continuously. In the ST operation regime ($V_{dd} > 0.6$ V), the

leakage energy as a percentage in total energy is low compared with the dynamic energy, while in the NT operation regime ($V_{dd} < 0.6$ V), it plays an important role.

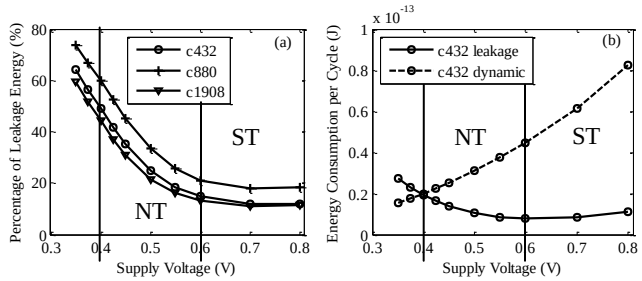


**Figure 4. Leakage energy consumption as a percentage of total energy consumption (a) and the comparison of leakage and dynamic energy consumption (b) for some ISCAS'85 benchmarks operating at different supply voltages.**

# 3. NEAR-THRESHOLD REGIME

The authors in [9] addressed the transparency window assignment problem to minimize the EDP for SEFF-based pipelined circuits in the traditional ST regime, considering the voltage scaling and delay variability. In this work, we focus on the pipelined circuits that are designed to operate in both NT and ST regimes. Thus the issues that we need to address include: 1) how to jointly determine the transparency window sizes and the clock period considering the delay variations and the timing yield; and 2) how to obtain the optimal DL design that minimizes the EDP in both the NT and ST regimes, accounting for the fact that the circuit delay follows different relations versus the supply voltage (i.e., $\alpha$-power law in the ST regime and exponential relation in the NT regime).

## 3.1 Timing Variability

The process variation, such as Random Dopant Fluctuation (RDF), results in variation of some important parameters such as the threshold voltage $v_{th}$. Both of the inter-die (global) and intra-die (local) threshold voltage variations cause the delay variation in logic circuits. The inter-die variation shifts the delays of all logic circuits in the same direction, i.e., either all increase or all decrease. In contrast, the delay variations caused by the intra-die variation are completely shifted to random directions. These two types of variations are handled in different ways. The inter-die variation can be effectively mitigated by body biasing [14]. Thus, in this work, we account for the intra-die variation only and consider the delays follow the Gaussian distribution $N(\mu, \sigma)$.

### *3.1.1 Timing variability in combinational circuits*

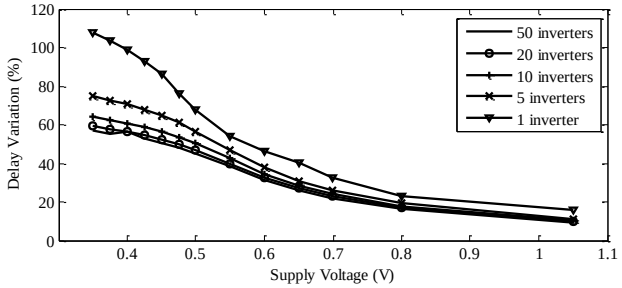


**Figure 5. $3\sigma$ delay variations of an inverter chain versus the supply voltage obtained using Monte Carlo simulation.**

It is known that the on-current of a CMOS gate, which determines the circuit delay, is very sensitive to the variation of threshold voltage in the NT operation regime. Precisely, the on-current of a circuit in the NT regime is exponentially proportional to the threshold voltage [1][2]. Thus, the variation in delay becomes much more significant in the NT regime [5]. We perform 5000 Monte Carlo simulations using 32/28nm technology and assume 10% intra-die $v_{th}$ variation. Figure 5 shows the $3\sigma/\mu$ delay variation of several FO4 inverter chains, where $\mu$ and $\sigma$ are the mean and standard variation of the delay distribution. The delay variation increases by 5X at $v_{dd} = 0.35$ V, compared to that at $v_{dd} = 1.05$ V. The results also show that the delay variation reduces as the length of inverter chain increases. This is due to the effect that the random $v_{th}$ variations have more chance to cancel out with each other for a long inverter chain.

### *3.1.2 Timing variability in flip-flops*

Similar to the combinational circuits, the flip-flops also have delay variations in their components, which results in the variations of the critical times [15][16]. For example, according to [15], the setup time is the summation of the delays of gates TG1, INV1, and INV2. Figure 6 shows the distributions of the setup time and clock-to-q delay under same $v_{th}$ variation as above.

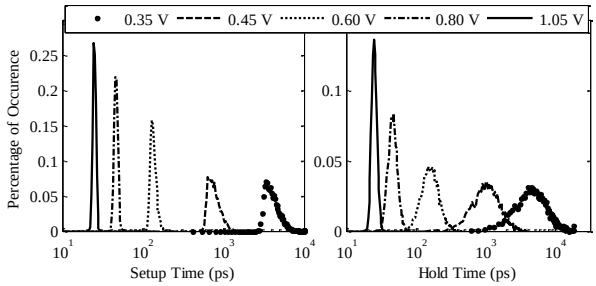


**Figure 6. Distributions of the setup time (left) and clock-to-q (right) time under threshold voltage variation.**

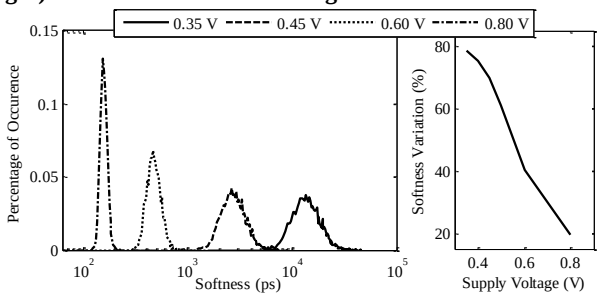


**Figure 7. Distribution (left) and $3\sigma/\mu$ variation (right) of the transparency window under the threshold voltage variation. This DL provides 84 ps transparency window at 1.05V.**

In addition, the SEFFs incur another type of the delay variation on the DLs, that is, the transparency windows are also affected by the delay variation. Since the DL is an independent part in the SEFF and only postpones the clock signal of the mater latch, the variation in DLs does not affect the critical times. Figure 7 shows the distribution and $3\sigma/\mu$ variation of the transparency window size under the threshold voltage variation.

## 3.2 Yield of the Pipeline

The timing yield of the pipelined circuits is defined as the probability that all the stages in the designed pipeline meet a certain delay target. Under the process variation, the delay is a random variable and can be modeled using a normal distribution $N(\mu, \sigma)$ with mean value of $\mu$ and standard variation of $\sigma$. In (1), we define two new random variables for setup time constraint and hold time constraint as follow,

$$D_{setup,i} = t_{cq,i-1} + w_{i-1} + D_{max,i} + t_{se,i} - w_i$$
$$D_{hold,i} = t_{h,i} + w_i - t_{cq,i-1} - D_{min,i} \quad (4)$$

To enforce the timing yield, we rewrite (1) as follows and set the timing conditions using the $\pm 3\sigma$ delays of $D_{setup,i}$ and $D_{hold,i}$,

$$D_{setup,i}(3\sigma) \le T_{clk}$$
$$D_{hold,i}(-3\sigma) \le 0 \quad (5)$$

In this work, we apply frequency scaling to find the minimum $T_{clk}$ so that condition (5) is met.

Note that all random variables in $D_{setup,i}$ are independent of each other due to the intra-die variation. First, the distribution of the combinational delay $D_{max,i}$ is independent of SEFF. Second, the critical times in SEFF are also independent of the transparency window, because the DL is a separate part in SEFF and is not involved in determining the critical times. Finally, $t_{se,i}$ and $t_{cq,i-1}$ comes from different SEFFs, and thereby they are independent. Similarly, all random variables in $D_{hold,i}$ are also independent of each other. Since the sum of several independent normal random variables is still a normal random variable, we calculate the mean value and standard variation of $D_{setup,i}$ and $D_{hold,i}$ as follows:

$$\mu_{total} = \sum_{i=1}^{N} \mu_i$$
$$\sigma_{total} = \left(\sum_{i=1}^{N} \sigma_i^2\right)^{1/2} \quad (6)$$

### 3.3 PMOS Header in Delay Line

As we mentioned above, the low supply voltage in NT regime significantly slows down the circuits, including both the combinational logics in the pipeline stages and the DLs that determine the transparency window sizes. In addition, the delay variation of the circuit depends on the logic depth. However, the critical paths in the combination circuits are typically much longer than the DLs. Thus, a fundamental issue is that the DLs, which are traditionally designed and optimized in the ST regime to provide the optimal transparency windows, may not work optimally in the NT regime. In this work, we propose a novel structure of the DL and a design methodology to optimize the EDP of the pipeline in both NT and ST regimes, in order to ensure that the SEFF-based pipeline works optimally in these two regimes.

Figure 7 shows that the $3\sigma/\mu$ variation of the transparency window size increases significantly in the NT regime due to its simple structure, i.e., a few inverters. To minimize the EDP, we have to design a relatively long DL, which can provide enough softness to tolerate the delay variation. For example, to provide 10ns softness at 0.35V, we need to over-design the DL to have the mean softness of 50ns if the $3\sigma$ delay variation is 80%. Thus, the SEFF-based pipelines that are designed and optimized in the ST regime typically cannot provide enough softness in the NT regime since the delay variation increases significantly.

We propose to add a PMOS header on top of the traditional DL, as shown in Figure 2 (b). The PMOS header delays the DL in two ways. First, the on-current decreases so that the DL takes more time to make the transition. Second, the existence of the PMOS header results in a voltage drop at the source terminal of DL, which is equivalent to reducing the supply voltage. This effect is small in the ST regime since the typical supply voltage is much larger than the voltage drop. However, in the NT regime, this voltage drop is no longer negligible compared to the low supply voltage. In contrast, it plays an important role in extending the softness since the delay is exponentially related to the supply voltage in the NT regime. Therefore, the proposed DL structure provide us a large transparency window size in the NT regime while only slightly affects that in the ST regime.

Figure 8 depicts the normalized transparency window size of two modified DLs with different PMOS header widths. We normalize the window size to the window size of the original header-less DLs. Figure 8 shows that with the proposed PMOS header we can increase the transparency window size by more than 3X in the NT regime while only slightly affect that in the ST regime. Another advantage of the proposed DL structure is the DL power savings due to the stacking effect. We determine the width of the PMOS header using the following design methodology.

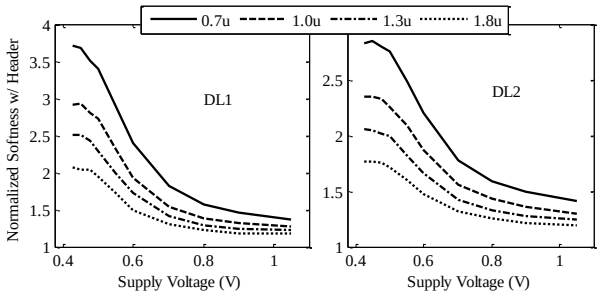


**Figure 8. Normalized softness versus supply voltage for DL1 (left, 20ps at 1.05V) and DL2 (right, 84ps at 1.05V).**

## 4. OPTIMIZATION OF ENERGY-DELAY PRODUCT IN SEFF-BASED PIPELINE

### 4.1 Formulation of the Optimization Problem

We focus on minimizing the EDP in the SEFF-based pipelined circuits. The energy consumption per clock cycle, denoted by $E_c$, contains two components: the dynamic energy consumption $E^{dyn}$ and the leakage energy consumption $E^{leak}$. The component $E^{dyn}$ is determined by the total capacitance being charged and discharged during each transition, the switching factor and the supply voltage, and thus it is independent of the clock period. In contrast, $E^{leak}$ is linearly proportional to the clock period. For each of them, we separate the energy consumed by the combinational logics and flip-flops. Therefore, $E_c$ is given by,

$$\begin{aligned} E_c &= E^{dyn} + E^{leak} = E_{cb}^{dyn} + E_{FF}^{dyn} + E_{cb}^{leak} + E_{FF}^{leak} \\ &= \sum_{i=1}^{N} \left(E_{cb,i}^{dyn}(V_{dd}) + P_{cb,i}^{leak}(V_{dd}) \cdot T_{clk}\right) \\ &+ \sum_{j=1}^{N-1} \Big(E_{dff,j}^{dyn}(V_{dd}) + E_{DL,j}^{dyn}(V_{dd}, w_j) + P_{dff,j}^{leak}(V_{dd}) \\ &\qquad \cdot T_{clk} + P_{DL,j}^{leak}(V_{dd}, w_j) \cdot T_{clk}\Big) \end{aligned} \quad (7)$$

The subscript *cb* in (7) stands for combinational logics. We separate the energy consumed by SEFFs into two parts: *dff* stands for a traditional hard-edge D-flip-flop (DFF), and *DL* stands for delay line. We formulate the EDP optimization problem in SEFF-based pipelined circuits as follows.

- Given: Characterized distribution of $w_j$, $t_{se}$, $t_{cq}$, $t_h$, $D_{max,i}$, $D_{min,i}$; energy consumption $E_{cb,i}^{dyn}$, $E_{dff,j}^{dyn}$, $E_{DL,j}^{dyn}$; power consumption $P_{cb,i}^{leak}$, $P_{dff,j}^{leak}$, $P_{DL,j}^{leak}$ for $i \in [1, \dots, N], j \in [1, \dots, N-1]$, at each specific supply voltage $V_{dd}$.
- Find: $w_j^0$, $s_j$, and $T_{clk}$, for $j \in [1, \dots, N-1]$.
- Minimize: $EDP = E_c \cdot T_{clk}$.
- Subject to: Constraints (4), (5), and (6).

Note that in (7), $w_j$ stands for the transparency window size of the $j$-th SEFF. It depends on the window size $w_j^0$ of the original header-less $j$-th DL, and PMOS header width $s_j$. $w_j^0$'s and $s_j$'s are the actual design variables for the DL. Thus, we find the appropriate $w_j^0$ and $s_j$ values as our final results.

### 4.2 Pareto Optimal Curve of Energy/Power-Softness Trade-off

To solve the EDP optimization problem, we first derive the relation between $w_j$ and $E_{DL,j}^{dyn}(V_{dd}, w_j)$, $P_{DL,j}^{leak}(V_{dd}, w_j)$,

respectively. The former term depends on $w_j$ since the PMOS header introduces extra capacitance and reduces the voltage swing of the DLs, while the latter one does due to the stacking effect. To characterize these relations, we sweep different original DL window sizes $w_j^0$'s and PMOS header sizes $s_j$'s, as shown in Figure 9. Since the energy overhead is positive related to the transparency window size, it would be more desirable if we can achieve larger window size with smaller energy overhead. In Figure 9, the circle markers represent the original header-less DL windows and the cross markers represent the proposed DLs using different PMOS headers. We find that by adopting PMOS headers, we reduce the energy/power overhead and simultaneously extend the window size. We achieve better energy-softness trade-off by applying relatively small PMOS header. However, the PMOS header cannot be too small otherwise it increases the skew between the delayed CLKd and CLKdb signals. We limit the minimum width of PMOS to be 0.7um so that the skew between CLKd and CLKdb signals is acceptable, according to our simulation results.

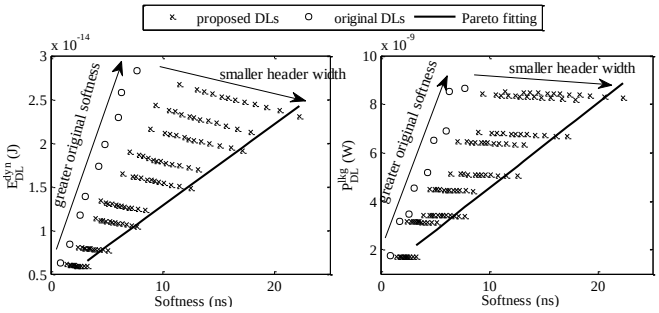


**Figure 9. Dynamic energy and leakage power overhead versus softness for different original header-less DLs and header width at $V_{dd} = 0.4\,V$.**

Figure 9 shows many energy/power-softness trade-off options at a specific $V_{dd}$, however, we are only interested in those with large transparency window size and small energy/power consumption. Therefore, we create Pareto-optimal curves for both (dynamic energy consumption vs. softness) and (leakage power vs. softness), as shown in Figure 9. We fit the two Pareto-optimal curves using linear relations so that, we have,

$$Pareto(E_{cb,i}^{dyn}(w_j)) = a_E \cdot w_j + b_E$$
$$Pareto(P_{DL,j}^{leak}(w_j)) = a_P \cdot w_j + b_P \quad (8)$$

where $a_E, b_E, a_P, b_P$ are fitting parameters. The Pareto-optimal curves provide the optimal energy/power-softness trade-off points that we can achieve using the proposed DLs. We use (8) to substitute $E_{DL,j}^{dyn}$ and $P_{DL,j}^{leak}$ in (7).

## 4.3 Solution Method

The cost function in (7) includes the product of $T_{clk}$ and $w_j$, and thereby is not convex. To solve the optimization problem, we perform a ternary search on $T_{clk}$. During each search, we solve a linear programming (LP) problem for a fixed $T_{clk}$ to obtain the optimal $w_j$ and the corresponding cost function value, based on which we narrow down the search range of $T_{clk}$. The optimal clock period is determined where we find the minimal value of $E_c \cdot T_{clk}$. Note that although the Pareto-optimal curves in (8) provide optimal trade-off points, the DL configurations are discrete so that not all points along the Pareto curves are feasible. In general, we pick the feasible point (each feasible point gives a combination of $w_j^0$ and $s_j$) that has the closest but smaller softness value compared with the optimal $w_j$, since a larger value of $w_j$ could potentially cause hold time violation, which is more difficult to handle. However, smaller values of $w_j$ could also potentially lead to setup time violations. Thus we check the setup time inequality (5) and slightly extend the clock period to solve the setup time violation, if there is. The design flowchart is given in Figure 10.

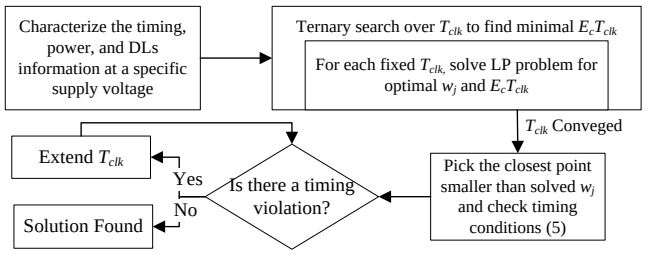


**Figure 10. Proposed design flowchart of SEFF-based pipeline.**

Figure 10 show the proposed design flowchart of the SEFF-based pipeline at a specific supply voltage. To design a pipelined circuits working in both the ST and NT operation regime, we perform the design flow for two desired supply voltage in these regimes, e.g. 0.4V and 0.8V, and configure the DLs and PMOS header as,

$$w_j = \eta \cdot w_{j,NT} + (1-\eta) \cdot w_{j,ST}$$
$$s_j = \eta \cdot s_{j,NT} + (1-\eta) \cdot s_{j,ST} \quad (9)$$

where the $\eta$ is the ratio of the time that the pipeline is being operated in the NT operation regime.

# 5. EXPERIMENTAL RESULTS

We apply the proposed design flow to some example pipelined circuits, while the stages are synthesized using ISCAS'85 benchmarks. We adopt the Synopsys 32/28nm technology and explore the supply voltage ranging from 0.35V to 1.05V. Based on these combinational benchmarks, we create four example pipelines, as shown in Table 2. We compare the proposed design method with the method proposed in [9], which designs and optimizes the SEFF-based pipeline only in the ST regime. We use hard-edge DFFs-based pipelines as the baseline, in which the clock period is simply determined by the slowest stage, and normalize the optimized EDP to it.

**Table 1. Distribution of maximum and minimum delay (ps) of selection benchmarks at nonimal voltage 1.05V.**

| Benchmarks | $\mu(D_{\max})$ | $\sigma(D_{\max})$ | $\mu(D_{\min})$ | $\sigma(D_{\min})$ |
|---|---|---|---|---|
| c432 | 803 | 19.3 | 119 | 3.2 |
| c499 | 614 | 9.5 | 259 | 2.4 |
| c880 | 759 | 20.5 | 144 | 3.6 |
| c1355 | 1047 | 31.9 | 387 | 10.8 |
| c1908 | 994 | 31.6 | 281 | 8.6 |

**Table 2. Four example pipelined circuits.**

| Pipeline | Configuration |
|---|---|
| TB1 | c1908, SEFF, c880 |
| TB2 | c432, SEFF, c1908, SEFF, c499 |
| TB3 | c1908, SEFF, c432, SEFF, c1355, SEFF, c880 |
| TB4 | c432, SEFF, c880, SEFF, c1908, SEFF, c499 |

Table 3 shows the results of the EDPs obtained by using different design methods. The percentage reductions of the EDP range from 6.0% to 18.4%. Compared to the hard-edge DFFs-based pipelines, the SEFF-based pipelines achieve better EDPs using the frequency scaling. Precisely, the SEFF-based pipeline improves the operating frequency via time borrowing when there are slacks in some of the pipeline stages. Furthermore, since the leakage energy consumption is the product of the leakage power and the clock period, the leakage energy consumption is also reduced as the clock period decreases.

The leakage energy reduction mentioned above becomes more noticeable in the NT regime and below where the leakage energy consumption is dominant. As seen in Figure 4, the leakage energy is only a small portion of the total energy consumption in the ST regime, while it dominates the total energy consumption in the NT regime. Figure 11 compares the delay reduction for TB1 achieved by the SEFF-based pipelines. Notice that the normalized delay in Figure 11 is lower than the corresponding normalized EDP in Table 3 at high supply voltages, which indicates that the total energy consumption of SEFF-based pipeline is slightly higher than hard-edge DFFs-based pipeline due to the energy overhead of the SEFFs. However, at low supply voltage levels, the normalized EDP of SEFF-based pipeline is lower because both energy and delay are reduced.

Our design method consistently outperforms the design method proposed in [9]. In the ST regime, our method outperforms the method in [9] by achieving better energy-delay trade-off curves, as shown in Figure 9. In other words, we provide the same softness with lower energy consumption by using the proposed DL design. In the NT regime, where the DL in [9] cannot provide enough softness, our method achieves better EDPs by providing more softness with acceptable energy consumption. Thus, we achieve lower EDP in the NT regime.

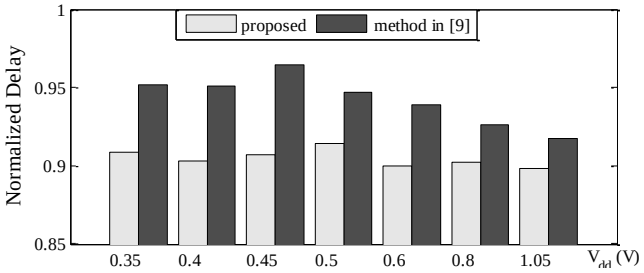


**Figure 11. Normalized stage delay reduction for TB1 for the proposed method versus the baseline method of [9].**

# 6. CONCLUSION

Previous work on soft-edge flip-flop (SEFF)-based pipeline mainly focused on the super-threshold operation regime. In this work, we proposed a design methodology for SEFF-based pipeline in both the super- and near-threshold (NT) operation regimes. We considered the high process variation in NT regime and maintained the timing yield of the pipeline by setting the timing constraints using the $3\sigma$ delay. We modified the structure of the delay lines by adding a PMOS header to achieve better energy-softness trade-off, which is demonstrated to be very effective in NT regime due to the high process variation. We applied the proposed method to some example pipelines constructed using ISCAS'85 benchmarks and demonstrated significant energy-delay savings in both the ST and NT regimes.

# REFERENCES


[1] Dreslinski, R. G., et al. 2010. Near-threshold computing: Reclaiming moore's law through energy efficient integrated circuits. *Proceedings of the IEEE*, 98(2).

[2] Markovic, D., Wang, C. C., Alarcon, L. P., Liu, T. T., and Rabaey, and J. M. 2010. Ultralow-power design in near-threshold region. *Proceedings of the IEEE*, 98(2).

[3] Calhoun, B. H., Wang, A., and Chandrakasan, A. 2005. Modeling and sizing for minimum energy operation in subthreshold circuits. In *Journal of Solid State Circuits*, 40(9).

[4] Zhai, B., Hanson, S., Blaauw, D., and Sylvester, D. 2005. Analysis and mitigation of variability in subthreshold design. In *Proc. of Int'l Symp. on Low Power Electronics and Design*.

[5] Seo, S., et al. 2012. Process variation in near-threshold wide SIMD architectures. In *Proc. of Design Automation Conf*.

[6] Manne, S., Klauser, A., and Grunwald, D. 1998. Pipeline gating: speculation control for energy reduction. In *ACM SIGARCH Computer Architecture News*.

[7] Jacobson, H., et al. 2005. Stretching the limits of clock-gating efficiency in server-class processors. *In Proc. of High-Performance Computer Architecture*.

[8] Partovi, H., Burd, R., Salim, U, Weber, F., DiGregorio, L., and Draper, D. 1996. Flow-through latch and edge-triggered flip-flop hybrid elements. *In Proc. of Solid-State Circuits Conf*.

[9] Ghasemazar, M. and Pedram, M. 2008. Minimizing the energy cost of throughput in a linear pipeline by opportunistic time borrowing. In *Proc. of Int'l Conf. on Computer Aided Design*.

[10] Seok, M., Jeon, D., Chakrabarti, C., Blaauw, D., and Sylvester, D. 2011. Pipeline strategy for improving optimal energy efficiency in ultra-low voltage design. In *Proc. of Design Automation Conf*.

[11] Datta, A., Bhunia, S., Mukhopadhyay, S., Banerjee, N., and Roy, K. 2005. Statistical modeling of pipeline delay and design of pipeline under process variation to enhance yield in sub-100nm technologies. In *Proc. of Design and Test in Europe*.

[12] Joshi, V., Blaauw, D., and Sylvester, D. 2007. Soft-edge flip-flops for improved timing yield: design and optimization. In *Proc. of Int'l Conf. on Computer Aided Design*.

[13] Synopsys 32/28 nm Generic Library: https://sso.synopsys.com/idp/Authn/UserPassword.

[14] Hanson, S., et al. 2007. Performance and variability optimization strategies in a sub-200mV, 3.5 pJ/inst, 11nW subthreshold processor. In *Proc. of Int'l Symp. on VLSI Circuits*.

[15] Fisher, S., Dagan, R., Blonder, S., and Fish, A. 2011. An improved model for delay/energy estimation in near-threshold flip-flops. In *Proc. of Int'l Symp. on Circuits and Systems*.

[16] Lotze, N., Ortmanns, M., and Manoli, Y. 2008. Variability of flip-flop timing at sub-threshold voltages. In *Proc. of Int'l Symp. on Low Power Electronics and Design*.

[17] Pu, Y., et al. 2010. Misleading energy and performance claims in sub/near threshold digital systems. In *Proc. of Int'l Conf. on Computer Aided Design*.


**Table 3. The energy-delay products of the proposed design method for four example pipelines.**

| $V_{dd}$ (V) | Normalized Energy-Delay Product (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TB1 | | | TB2 | | | TB3 | | | TB4 | | |
| | Hard-edge | Method in [9] | Proposed | Hard-edge | Method in [9] | Proposed | Hard-edge | Method in [9] | Proposed | Hard-edge | Method in [9] | Proposed |
| 1.05 | 100 | 92.70 | 89.93 | 100 | 96.45 | 90.98 | 100 | 95.83 | 92.89 | 100 | 95.53 | 89.94 |
| 0.8 | 100 | 93.19 | 90.01 | 100 | 96.53 | 91.03 | 100 | 95.27 | 93.11 | 100 | 95.91 | 90.18 |
| 0.6 | 100 | 93.21 | 88.47 | 100 | 96.70 | 88.79 | 100 | 96.71 | 91.64 | 100 | 96.21 | 87.25 |
| 0.5 | 100 | 93.82 | 89.89 | 100 | 97.22 | 88.05 | 100 | 97.57 | 92.03 | 100 | 96.83 | 86.02 |
| 0.45 | 100 | 95.52 | 88.30 | 100 | 98.48 | 87.18 | 100 | 99.49 | 94.00 | 100 | 98.25 | 84.67 |
| 0.4 | 100 | 93.12 | 86.58 | 100 | 96.39 | 86.55 | 100 | 98.74 | 92.82 | 100 | 95.55 | 83.54 |
| 0.35 | 100 | 92.56 | 86.02 | 100 | 95.79 | 85.29 | 100 | 98.15 | 92.60 | 100 | 94.65 | 81.65 |