INTEGRATION, the VLSI journal ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

journal homepage: www.elsevier.com/locate/vlsi

Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes ☆

Qing Xie *, Yanzhi Wang, Massoud Pedram

University of Southern California, Department of Electrical Engineering, Los Angeles, CA 90089, United States

ARTICLE INFO

Keywords:
Soft-edge flip-flop
Pipelined circuits design
Near-threshold computing
Process variation

ABSTRACT

Soft-edge flip-flop (SEFF) based pipelines can improve the performance and energy efficiency of circuits operating in the super-threshold (supply voltage) regime by enabling the opportunistic time borrowing. The application of this technique to the near-threshold regime of operation, however, faces a significant challenge due to large circuit parameter variations that result from manufacturing process imperfections. In particular, delay lines in SEFFs have to be over-designed to provide larger transparency windows to overcome the variation in path delays, which causes them to consume more power. To address this issue, this paper presents a novel way of designing delay lines in SEFFs to have a large enough transparency window size and low power consumption. Two types of linear pipeline design problems using the SEFFs are formulated and solved: (1) designing energy-delay optimal pipelines for the general usage that requires SEFFs to operate in both the near-threshold and super-threshold regimes, and (2) designing minimum energy consumed pipelines for particular use case with a minimum operating frequency constraint. Design methods are presented to derive requisite pipeline design parameters (i.e., depth and sizing of delay lines in SEFFs) and operating conditions (i.e., supply voltage and operating frequency of the design) in presence of process-induced variations. HSPICE simulation results using ISCAS benchmarks demonstrate the efficacy of the presented design methods.

1. Introduction

With the increase in demand for battery-powered devices and mobile equipment, the need for energy-efficient designs gains growing attentions. The ultra-low voltage operation, in particular, *near-threshold* (NT) operation, is quite effective in minimizing the energy consumption of a design by reducing its supply voltage to a level close to the threshold voltage of the transistors, V_{th}. Indeed, previous work on NT operation proved the existence of and analytically derived the *minimum energy* (*operation*) *point* (MEP), which is the optimal supply voltage level that minimizes the energy consumption [2,3]. However, NT operation comes at the cost of sacrificing the circuits' timing performance [4,5]. Hence, it is especially beneficial for applications that have relaxed timing requirements, e.g., medical monitoring devices and many types of environmental sensors. Note that digital circuits operating in the NT regime become quite sensitive to process-induced variations. For example, in 90 nm CMOS technology, the relative delay variation, defined as $3\sigma/\mu$, of a combinational logic block operating at 0.5 V increases by 2.5X compared to that of a block operating at 1 V [6].

An SEFF is a D flip-flop with an additional *delay line* (DL). The DL is added to postpone clock edges of the master latch to create a *transparency window* during which both master and slave latches are transparent. The transparency window allows the SEFF to pass a positive slack from one pipeline stage to the next. The *transparency window size* is also referred as *softness*. The SEFF-based pipelined circuits improve the operating frequency of the pipeline by enabling time borrowing from the non-critical stages to the critical stage. In addition, process variations are alleviated for "deeper" combinational logics since the local random variation cancel out when the logic depth increases [6,7]. Thus, soft edges between pipeline stages help to reduce the sensitivity of pipelined circuits on process variations.

Thanks to its properties of frequency enhancement and variation tolerance, we bring the SEFF-based pipeline to the NT regime. Precisely, we focus on designing and optimizing SEFF-based linear pipelines operating in multiple supply voltage regimes from the NT to *super-threshold* (ST) in this work. We show that SEFF-based pipelines that are optimized for operations in the ST regime [7,8] are not suitable in the NT regime. In particular, since circuit delays in the NT regime have an exponential relationship with the transistor threshold voltage, the variation in transistor threshold voltage causes a huge amount of variation to the SEFF softness. Therefore, SEFFs designed in [7,8] cannot provide enough softness in the NT regime to deliver a

☆The preliminary version of this work [1] has been presented in *Great Lakes Symposium on VLSI*, Paris, France, May, 2013.

* Corresponding author. Tel.: +1 213 740 4460.
E-mail addresses: xqing@usc.edu (Q. Xie), yanzhiwa@usc.edu (Y. Wang), pedram@usc.edu (M. Pedram).

0167-9260/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.vlsi.2013.11.003

certain timing yield. In order to design SEFF-based pipelined circuits that can operate in all supply voltage regimes, the key is to ensure that the same transistor sizes result in effective operation of the DL in all voltage regimes (and hence, appropriate setting of the FF softness) under process-induced variations.

We present a novel way of designing DLs in SEFFs to provide enough softness and reduce their power consumptions. Precisely, we add a PMOS header between the conventional DL and the supply voltage rail. The PMOS header results in a slight supply voltage drop on the DL, which is negligible in the ST regime but has a significant impact on the softness in the NT regime. Along with the increased softness, another benefit is the reduced leakage power consumption of the DL due to the *transistor stack effect*. We formulate the SEFF-based linear pipeline design problems and present methods to derive the optimal design parameters, such as the configuration and sizing of DLs. The presented methods also determine the optimal operating conditions, including the supply voltage and operating frequency, according to the problem setup and pipeline designs.

We solve two SEFF-based linear pipeline design problems. The first problem setup targets pipelined circuits operating in mixed NT and ST regime. We adopt the *energy-delay product* (EDP) as the cost function and minimize the EDP over all supply voltage levels. In particular, we formulate the EDP minimization problem for SEFF-based pipelined circuits as a mathematical programming problem, where the clock period, PMOS header width, and configurations of delay lines are optimization variables. The second problem targets the situation in which target pipelined circuits are required to meet some pre-specified minimum operating frequency. Since the frequency is constrained by the problem setup, we adopt the total energy consumption in one clock cycle as the cost function. We find the MEP and corresponding pipeline design parameters that satisfy the frequency constraints. The timing constraints of pipelined circuits are imposed by using the 3σ delay and accounting for the delay variation from every circuit component. We generate fitted Pareto-optimal lines of the DL between energy/power consumptions and FF softness so that we can select the DL configuration and sizing solution that are closest to optimal ones. Experimental results based on ISCAS benchmarks show significant reduction of up to 18.4% in EDP for the first problem and 9.1% in energy consumption for the second problem, respectively.

The remainder of the paper is organized as follows. The notion of SEFF-based pipelined circuits is reviewed in Section 2. In Section 3, we discuss design challenges in the NT regime and provide a novel way of designing DLs so that we can better apply the SEFF-based pipelining technique in the NT regime. We formulate two versions of the SEFF-based pipeline design problem and present solution methods for these in Section 4. Finally, experimental results are presented in Section 5.

2. Background

SEFF-based pipelines are capable to combat process variations and improve the timing performance of pipelined circuits. This section starts off by reviewing the related work, and continues by explaining how SEFFs enable opportunistic time borrowing across pipeline stages and presenting setup and hold timing conditions in a SEFF-based pipeline. Finally the energy-delay product is proposed as the metric to use in order to quantify the performance of pipelined circuits.

2.1. Related work

Pipelining is a well-known technique to improve the timing performance and energy efficiency of the processor [9]. Considerable efforts have been invested to design energy-efficient pipelined circuits. Jacobson in [10] presented clock-gating technique to reduce the power consumption of a microprocessor pipeline. Kim in [11] combined pipelining and parallel processing to reduce power consumption by 40%. Srinivasan in [12] investigated the optimal pipeline depth to balance the power consumption and timing performance. However, none of these work focused on the ultra-low voltage operation regime, which requires special techniques to handle process variations. Recently, authors in [13] explored aggressive latch-based super-pipelining technique and demonstrated the energy efficiency improvements for a 65 nm FFT core operating in ultra-low voltage regime. Although the latch-based pipeline design shows good capability of handling process-induced variations, it has many limitations including hold time violation issues, design difficulties using standard EDA tools, and the requirement of an extra clock network, which makes it power and area inefficient.

Applying soft-edge flip-flops (SEFFs) to digital circuits is a useful technique to improve the circuit performance. Joshi [7] presented to utilize SEFFs in sequential circuits to increase the timing yield in the presence of process variation. Authors in [14] adopted SEFFs to combat delay variations caused by NBTI. Authors in [8 and 15] presented an SEFF-based pipeline design method that utilizes voltage scaling and time borrowing for pipelined circuits, and demonstrated a sizeable reduction in the *energy-delay product* (EDP) and *power-delay product*. Dillen in [16] designed and implemented area-efficient SEFF-based x86-64 AMD microprocessor module and demonstrated enhancements of energy efficiency. In this work, we focus on designing SEFF-based pipelined circuits in the NT regime.

2.2. Soft-edge flip-flop-based pipelines

Fig. 1 illustrates general synchronous SEFF-based linear pipelined circuits. Considering the data consistency between the SEFF-based pipelined circuits and the input and output environments, we impose hard boundary conditions using conventional hard-edge flip-flops at the first and last stage of pipelined circuits. Between the two hard edges, pipelined circuits have multiple combinational logic stages whose delays are affected by process-induced variations. We build stage registers using SEFFs. The key idea is to postpone the clock signal for the master latch, as shown

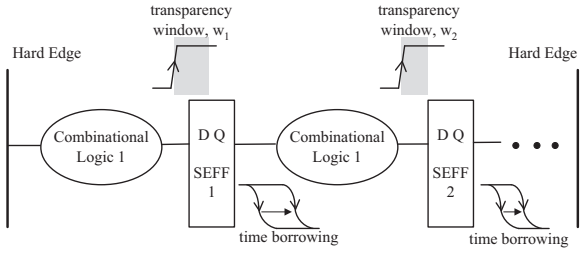


Fig. 1. A linear pipeline with soft-edge flip-flops.

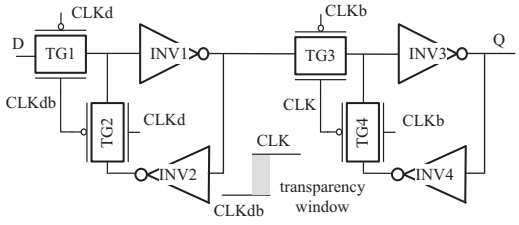


Fig. 2. Design of the positive-edge triggered soft-edge master slave flip-flops.

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

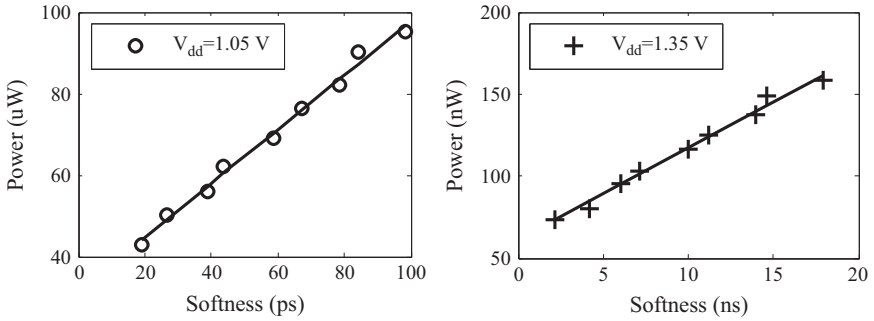


Fig. 3. SEFF power consumption versus softness in the ST regime (left) and the NT regime (right).

in Fig. 2, to create a transparency window, which enables time borrowing across pipeline stages.

The SEFFs come at the price of an additional amount of power consumption. Because DLs are typically a series of properly sized inverters, we tune the softness, which is the delay of the DL, through sizing or adding/removing inverters in the DL. We try a number of different SEFFs, i.e., different DL sizing and depths, operate them at different supply voltage levels, and plot their power consumption versus the softness in Fig. 3. The power consumption is calculated by measuring the average current in HSPICE and multiplying it by the voltage level. As shown in Fig. 3, the power consumption of the SEFF increases with softness in all regimes. The trend in Fig. 3 also agree with observations in [7]·[8].

SEFF also affects timing constraints of pipeline stages. Timing constraints in pipelined circuits mainly consist of a setup time constraint and a hold time constraint. We refer *critical timing parameters* as the setup time, hold time, and clock-to-q delay, and denote the transparency window size by w. We take the i-th stage as an example and show the waveforms in Fig. 4, where D and Q denote the input and output data for the flip-flop. Since the data can be captured by the end of the transparency window, the setup time constraint is relaxed by the transparency window size w_i. However, in the worst case that the data comes right before the transparency window is closed in the previous stage, the clock-to-q delay $t_{cq,i-1}$ is postponed by w_{i-1}. As shown in Fig. 4(a), Q_{i-1} is captured at the end of the previous stage transparency window and D_i must arrive by the setup time before the end of current stage transparency window. For the hold time constraint, the data cannot change until the transparency window is closed. Thus, in the worst case, we need to hold the data stable for a total time adding up hold time $t_{h,i}$ and transparency window size w_i. As shown in Fig. 4(b), Q_{i-1} is captured at the beginning of previous stage and D_i shall not change before the hold time. Therefore, in the worst case, for a stage in the middle of pipeline, timing constraints are summarized in (1).

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

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 pipelined circuits.

2.3. Energy and delay in pipelined circuits

Although SEFFs consume additional energy compared to conventional hard-edge flip-flops, especially for large transparency windows, they can be utilized to improve the energy efficiency when properly designed. The reason is that the leakage energy in one clock cycle of the whole circuit decreases with the clock period. This is extremely helpful for pipelined circuits operating in the NT operation regime. Fig. 5 shows that the leakage energy consumption plays a more important role in the NT regime.

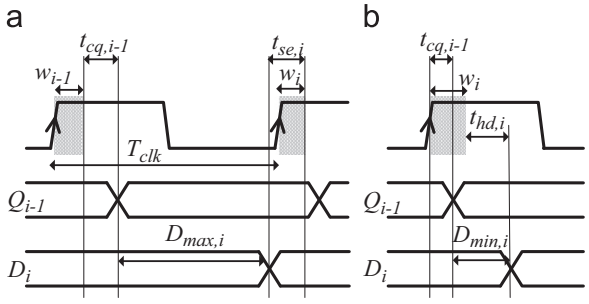


Fig. 4. Waveform explanation of (1): (a) setup-time constraint and (b) hold-time constraint.

Fig. 5 depicts the leakage energy consumption during a period equal to the worst-case delay of the combinational circuit at the given voltage level for some selected ISCAS'85 benchmarks at different supply voltage levels, under the 32/28 nm technology [18]. The leakage power decreases linearly whereas the circuit delay increases exponentially when the supply voltage is reduced [4]·[5], and therefore, the leakage energy consumption as a whole increases. In contrast, the dynamic energy consumption decreases quadratically when the supply voltage drops. In the ST operation regime ($V_{dd} > 0.6$ V), the leakage energy as a percentage of total energy consumption is very low ($< 20\%$), whereas in the NT operation regime ($V_{dd} > 0.6$ V), the leakage energy tends to be the dominant part of the total energy (up to 60%).

To evaluate the performance of pipelined circuits, we account for both timing performance and energy consumption. *Throughput*, which is typically used as the metric of the timing performance in pipelined circuits, is defined as the number of produced output values divided by the duration of time in which the data values were produced. For synchronous clocked circuits, we have

$$\text{Thruput} = \frac{\#\text{ of output data}}{\#\text{ of clock cycles} \times T_{clk}} \tag{2}$$

For pipelined circuits in steady-state (after the pipeline is full), the throughput defined in (2) is simply proportional to the inverse of the clock period. To account for the energy consumption as well, *energy per throughput* is normally used, and we have

$$\frac{\text{Energy}}{\text{Thruput}} \propto E_c \times T_{clk} \tag{3}$$

where E_c denotes the *total energy consumption in one clock cycle* for pipelined circuits. In this work, we use either $E_c \cdot T_{clk}$ or E_c (when T_{clk} is constrained) as the cost function, depending on the problem setup.

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

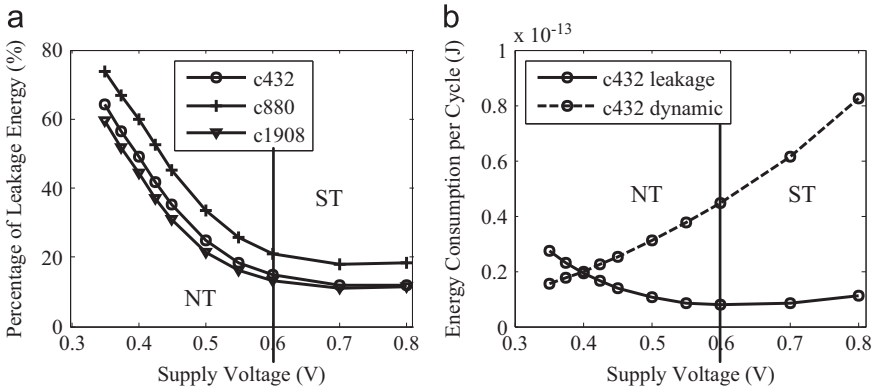


Fig. 5. Leakage energy consumption as a percentage of total energy consumption (a) and the energy consumption breakdown (b) for some ISCAS'85 benchmarks operating at different supply voltages.

3. Near-threshold regime

The authors in [8] addressed the transparency window assignment problem to minimize the EDP for SEFF-based pipelined circuits in the conventional ST regime. In this work, we focus on pipelined circuits operating in both NT and ST regimes. We need to address the issues of (1) how to determine the best-suited pipeline design parameters, such as DL configurations; and (2) how to derive the optimal operating conditions, such that the cost function is minimized and the timing constraint is satisfied. We account for process-induced delay variations of various components in pipelined circuits, as well as different relationships between the circuit delay and the supply voltage (i.e., α-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 variations of the threshold voltage, V_{th}. Both 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 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 [19]. Thus, in this work, we account for the intra-die variation only and consider that the delay follows the Gaussian distribution $N(\mu,\sigma)$.

3.1.1. Timing variability in combinational circuits

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 [4]·[5]. Thus, the variation in delay becomes much more significant in the NT regime [6]. We perform 5000 Monte Carlo simulations using 32/28 nm technology and assume 10% intra-die V_{th} variation. Fig. 6 shows the $3\sigma/\mu$ relative delay variation of several FO4 inverter chains, where μ and σ denote the mean and standard deviation of the circuit delay, respectively. The delay variation increases by 5X at $V_{dd}=0.35$ V, compared to that at $V_{dd}=1.05$ V. Results also show that the delay variation reduces as the length of inverter chain increases. This is due to the effect that random V_{th} variations will cancel out with each other in a long inverter chain.

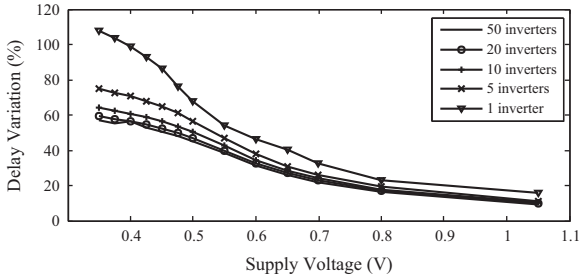


Fig. 6. $3\sigma/\mu$ relative delay variations of an inverter chain versus the supply voltage obtained using Monte Carlo simulation.

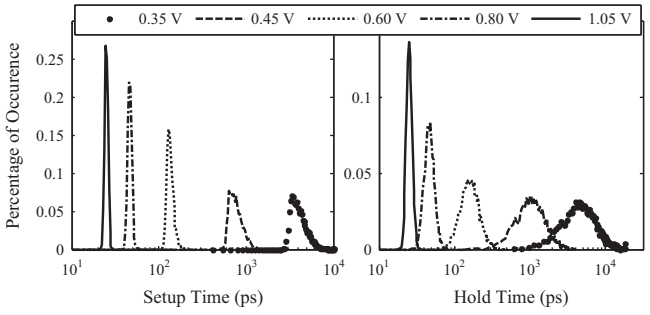


Fig. 7. Probability distributions of the setup time (left) and clock-to-q (right) time under threshold voltage variation.

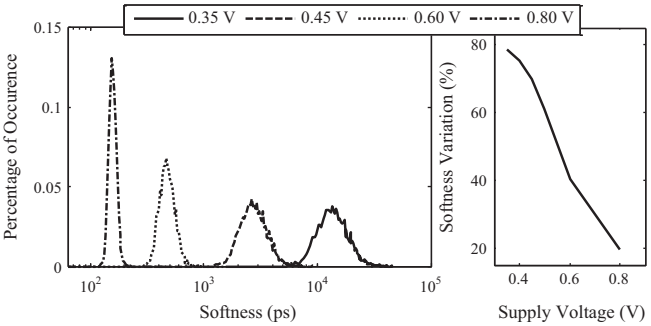


Fig. 8. Probability distribution (left) and $3\sigma/\mu$ relative variation (right) of transparency window under the threshold voltage variation. This DL provides 84ps transparency window at 1.05 V.

3.1.2. Timing variability in flip-flops

Similar to combinational circuits, flip-flops also have delay variations in their components, which result in variations of some critical timing parameters [20]·[21]. For example, according to [20],

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

the setup time is the summation of the delays of gates TG1, INV1, and INV2 shown in Fig. 2, and thereby is affected by delay variations of those gates. Fig. 7 shows probability distributions of the setup time and clock-to-q delay under the same level of V_{th} variation. Besides the setup time and clock-to-q time, transparency window sizes of SEFFs are also subject to process variations. Fig. 8 shows the probability distribution and $3\sigma/\mu$ variation of the transparency window size in presence of the threshold voltage variation.

3.2. Timing yield of the pipelined circuits

The timing yield of pipelined circuits is defined as the probability that all stages in the designed pipeline meet a certain delay target [17]. Under process variations, each delay value is a random variable and can be modeled using a normal distribution $N(\mu,\sigma)$ with the mean value of μ and standard deviation of σ. In (1), we define two new random variables for the setup time constraint and hold time constraint for the i-th pipeline stage as follows

$$D_{stage,se,i} = t_{cq,i-1} + w_{i-1} + D_{max,i} + t_{se,i} - w_i \tag{4}$$

$$D_{stage,hd,i} = t_{h,i} + w_i - t_{cq,i-1} - D_{min,i} \tag{5}$$

Note that (4) and (5) describe timing constraints for a pipeline stage in the middle of a series of SEFF-based pipeline stages. w_{i-1} or w_i are zero if hard-edge flip-flop registers are used in the previous stage or the current stage, respectively. To impose the timing yield constraint, we rewrite (1) as follows and set the timing constraints using the $\pm 3\sigma$ delays of $D_{setup,i}$ and $D_{hold,i}$ as follows.

$$\begin{aligned} &D_{stage,se,i}(3\sigma) \le T_{clk} \\ &D_{stage,hd,i}(-3\sigma) \le 0 \end{aligned} \tag{6}$$

In this work, we use (6) as the timing yield constraint in the formulated optimization problems.

Note that all random variables in right-hand side of (4) are independent of each other due to intra-die variations. First, the probability distribution of the combinational delay $D_{max,i}$ is independent of the SEFF. Second, the critical timing parameters in the 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 $t_{cq,i-1}$ come from different SEFFs belonging to different pipeline stages, and thereby, they are independent of each other. Similarly, all random variables in right-hand side of (5) are also independent of each other. We know that the sum of a set of independent Gaussian (normal) random variables, each with the mean value μ_i and standard deviation σ_i, is still a Gaussian random variable with mean and standard deviation given by

$$\begin{aligned} \mu_{total} &= \sum_{i=1}^{N} \mu_i \\ \sigma_{total} &= \left(\sum_{i=1}^{N} \sigma_i^2 \right)^{1/2} \end{aligned} \tag{7}$$

We calculate the mean value and standard deviation of $D_{stage,se,i}$ and $D_{stage,hd,i}$ based on (7).

3.3. PMOS header in the delay line

3.3.1. Effect on transparent window sizes

As mentioned above, transparency window sizes of DLs are also subject to delay variations when operating at the low supply voltage in the NT regime. Fig. 8 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 short inverter chain. To satisfy the timing yield constraint in (6), we have to over-design the DL that can provide enough softness to tolerate delay variations. For example, to provide 10 ns softness at the supply voltage level of 0.35 V, we need to over-design the DL to have a mean softness of 50 ns if the $3\sigma/\mu$ delay variation is 80%. However, this issue is less significant in the ST regime due to the smaller delay variation. Therefore, a fundamental issue is that DLs, which are conventionally designed and optimized in the ST regime may not work properly in the NT regime.

We add a PMOS header on top of the conventional DL, as shown in Fig. 9(a). The PMOS header affects the DL in two ways. First, the on-current decreases so that the DL takes more time for a transition. Second, the PMOS header causes a voltage drop at the source terminal of the pull-up network in the DL, which is equivalent to reducing its supply voltage. This effect is relatively small in the ST regime since the supply voltage is much larger than the voltage drop. However, in the NT regime, this voltage drop is no longer negligible given that the supply voltage is low. In contrast, it plays an important role in extending the softness since the delay is exponentially dependent on the supply voltage in the NT regime. Therefore, the presented DL structure greatly increases the transparency window size in the NT regime while only slightly affects that in the ST regime.

Fig. 9(b) and (c) shows normalized transparency window sizes of two PMOS-headed DLs with different PMOS header widths. We normalize the new window size to that of the original header-less DLs. Fig. 9 shows that with the presented 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. Therefore, DLs with PMOS headers are suitable to address the timing yield

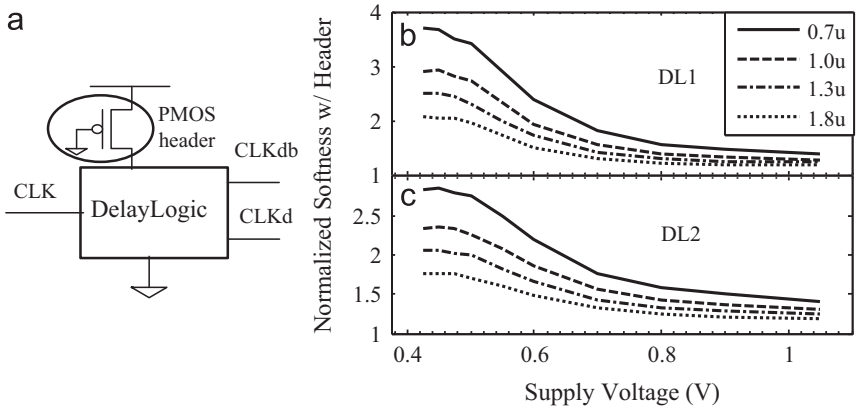


Fig. 9. Presented PMOS header (a) and normalized softness versus supply voltage for DL1 (b) (original softness of 20ps at 1.05 V) and DL2 (c) (original softness of 84ps at 1.05 V).

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

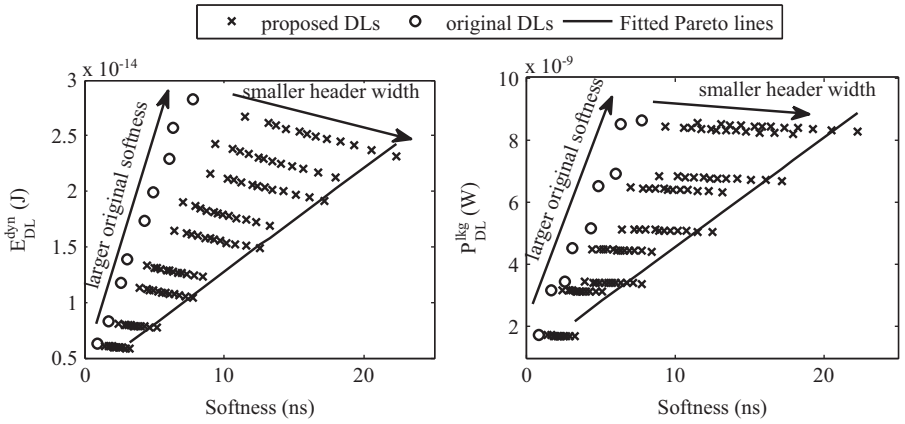


Fig. 10. Dynamic energy and leakage power consumption versus softness versus original header-less DLs and DLs with PMOS headers at $V_{dd} = 0.4V$.

issue in both NT and ST regimes with the same configuration and sizing. Another advantage of the presented DL structure is that it results in energy savings in the DL as well. We determine the best-suited width of the PMOS header using the following design methods.

3.3.2. Analysis of energy and area overhead

The PMOS header affects the DL energy consumption and costs extra area. We plot the DL dynamic energy and leakage power consumption versus different header widths and DL configurations for a 0.4 V supply voltage level in Fig. 10. Circles denote the dynamic energy consumption in one clock cycle and leakage power consumption of DLs in original SEFFs, while crosses denote those of DLs in PMOS headed SEFFs, respectively. Note that each line of crosses in Fig. 10 corresponds to the same original DL configuration but with different PMOS header widths. Results in Fig. 10 are obtained by using HSPICE simulations. Results at other supply voltages, though not plotted due to the size limit for the paper, show similar relations of dynamic energy and leakage power versus PMOS header widths, like in Fig. 10.

The presented PMOS header reduces both the dynamic energy consumption and leakage power consumption of DLs. In Fig. 10, after adding the PMOS header to original DLs, the dynamic energy consumption of DL decreases because smaller supply voltage is delivered to the DL logic, while the leakage power consumption decreases due to the stack effect caused by the PMOS header. In addition, the dynamic energy consumption and leakage power consumption of DLs decrease as the PMOS header width decreases. However, the PMOS header cannot be too small otherwise mismatches between rising and falling transitions of the delayed clock signal may affect the clock signal integrity. We limit the lower bound of the PMOS header width to be 0.7 μm so that the mismatch is within 10%, according to our simulation results. The upper bound is imposed because, according to our analysis in Section 4.1.2, one prefers smaller PMOS header, which brings more softness and results in lower dynamic energy consumption and leakage power dissipation.

Adding a PMOS header causes area overhead. We tried different PMOS header widths (0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.6, 1.8, 2.0, 2.5) μm to generate fitted Pareto lines of Fig. 10. Note that these lines go through solutions that correspond to the smallest PMOS header widths. Therefore, the area overhead of the PMOS headers is quite small even compared to that of hard-edge flop-flop. Note that our proposed design has smaller area for a SEFF compared to that of reference [8] when these two designs of SEFFs provide the same degree of softness.

4. SEFF-based pipelined circuits optimization

We consider two different optimization problem setups in this work. First, we consider the general case where the SEFF-based pipelined circuits operate in both NT and ST regimes, and optimize the SEFF-based pipeline design over the entire operating regimes. To account for both delay and energy, we adopt the energy-delay product as the cost function to be minimized. In the second problem setup, we focus on a particular use case where a minimal operating frequency constraint is imposed to pipelined circuits. We optimize the pipeline design such that the total energy consumption in one clock cycle is minimized while the frequency constraint is met.

4.1. EDP optimization for mixed NT/ST operations

We present a design method for the SEFF-based pipelined circuits targeting the general case in this section. We consider multiple supply voltages for pipelined circuits that include both NT and ST operating regimes. At each supply voltage level, we find the proper clock frequency such that the EDP is minimized and timing constraints are satisfied. Note that pipelined circuits will use same DLs at different supply voltage levels, and thereby they shall be designed judiciously.

4.1.1. Formulation of the optimization problem

The energy consumption in one 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. Thus, it is independent of the clock period. In contrast, E^{leak} is linearly proportional to the clock period. Each of them is comprised of the energy consumed by the combinational logics and flip-flops. Therefore, we have

$$\begin{aligned} E_c &= E^{dyn} + E^{leak} = E^{dyn}_{cb} + E^{dyn}_{ff} + E^{leak}_{cb} + E^{leak}_{ff} \\ &= \sum_{i=1}^{N} \left(E^{dyn}_{cb,i}(V_{dd}) + P^{leak}_{cb,i}(V_{dd}) \times T_{clk} \right) \\ &\quad + \sum_{j=1}^{N-1} \left(E^{dyn}_{dff,j}(V_{dd}) + E^{dyn}_{DL,j}(V_{dd}, w_j) \right. \\ &\quad \left. + P^{leak}_{dff,j}(V_{dd}) \times T_{clk} + P^{leak}_{DL,j}(V_{dd}, w_j) \times T_{clk} \right) \end{aligned} \tag{8}$$

The subscript *cb* in (7) stands for combinational logics, and *ff* stands for SEFFs. $1 \le i \le N$ and $1 \le j \le N-1$ are stage indices of combinational blocks and sequential elements, respectively. We

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

separate the energy consumed by SEFFs into two parts: energy consumption of DLs (*DL*) and that of conventional hard-edge D flip-flops (*dffs*). We formulate the EDP optimization problem as follows.

- Given: characterized probability distribution of w_j, t_{se}, t_{cq}, t_h, $D_{max,i}$, $D_{min,i}$; energy consumption $E^{dyn}_{cb,i}$, $E^{dyn}_{dff,j}$, $E^{dyn}_{DL,j}$; power consumption $P^{leak}_{cb,i}$, $P^{leak}_{dff,j}$, $P^{leak}_{DL,j}$ for $i \in [1, \ldots, N]$, $j \in [1, \ldots, N-1]$, at each specific supply voltage V_{dd}.
- Find: w^0_j, s_j, and T_{clk}, for $j \in [1, \ldots, N-1]$.
- Minimize: $EDP = E_c \times T_{clk}$.
- Subject to: timing constraints (4–7).

Note that in (7), w_j stands for the transparency window size of the j-th SEFF. It depends on the configuration of the original header-less j-th DL with original transparency window size w^0_j, and the PMOS header width s_j. w^0_j's and s_j's are actual design parameters in DLs that we are interested in. Thus, we seek to derive appropriate w^0_j and s_j values. Notice that the critical timing parameters of the SEFF, t_{se}, t_{cq}, t_h, are only determined by the design of the D flip-flop part and thus they are independent of w^0_j's and s_j's, which are affected by the design of the DL part.

4.1.2. Pareto-optimal energy/power vs. softness trade-offs

To solve the EDP optimization problem, we first derive the relationships between $E^{dyn}_{DL,j}(V_{dd}, w_j)$, $P^{leak}_{DL,j}(V_{dd}, w_j)$ and w_j. Fig. 10 plots measured $E^{dyn}_{DL,j}(V_{dd}, w_j)$ and $P^{leak}_{DL,j}(V_{dd}, w_j)$ for different PMOS header widths at a specific V_{dd}. As we only interested in those points with a large softness and small energy/power consumption, we obtain Pareto-optimal energy/power trade-off points for the dynamic energy and leakage power vs. softness where small PMOS header widths are used. Next, we perform linear curve fittings along those Pareto-optimal trade-off points. Solid lines in Fig. 10 show fitted Pareto lines of optimal dynamic energy consumption vs. softness and leakage power consumption vs. softness at supply voltage 0.4 V, respectively. For each V_{dd} level, we repeat this and obtain fitted Pareto lines by using linear functions denoted by

$$\begin{aligned} \text{Pareto}(E^{dyn}_{DL,j}(w_j), V_{dd}) &= a_E(V_{dd}) \times w_j + b_E(V_{dd}) \\ \text{Pareto}(P^{leak}_{DL,j}(w_j), V_{dd}) &= a_P(V_{dd}) \times w_j + b_P(V_{dd}) \end{aligned} \tag{9}$$

where a_E, b_E, a_P, b_P are supply voltage-dependent fitting parameters. Fitted Pareto lines provide optimal energy/power-softness trade-off points that we can achieve using DLs with PMOS headers. We use (9) to substitute $E^{dyn}_{DL,j}$ and $P^{leak}_{DL,j}$ in (8).

4.1.3. Solution method

The cost function in (8) is not convex. To solve the optimization problem, we perform a ternary search on T_{clk}. For each T_{clk} value, we solve a linear programming (LP) problem for the fixed T_{clk} value to obtain the optimal w_j and the corresponding cost function value, based on which we can narrow down the search range of T_{clk} The optimal clock period is determined when we find the minimal value of $E_c \cdot T_{clk}$. Note that although fitted Pareto lines in (9) provide the optimal trade-off points, DL configurations are discrete so that not all points along fitted Pareto lines are achievable. In general, we round the softness value down to the closed feasible point (w^0_j, s_j). This is because 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 potentially lead to setup time violations. Thus we check the setup time constraint (6) and slightly extend the clock period to resolve the potential setup time violation. The design flow is provided in the Algorithm 1.

Algorithm 1. Find T_{clk} to minimize the EDP (*MINEDP*)

Inputs: t_{se}, t_{cq}, t_h, $D_{max,i}$, $D_{min,i}$, $E^{dyn}_{cb,i}$, $E^{dyn}_{dff,j}$, $E^{dyn}_{DL,j}$; $P^{leak}_{cb,i}$, $P^{leak}_{dff,j}$, $P^{leak}_{DL,j}$ for $i \in [1, \ldots, N]$, $j \in [1, \ldots, N-1]$, at a specific supply voltage V_{dd}, convergence parameter ϵ_T.
Set a reasonable initial search range of T_{clk}, i.e., $[T^{bot}_{clk}, T^{up}_{clk}]$
Do ternary search in $[T^{bot}_{clk}, T^{up}_{clk}]$:
  Pick T'_{clk} and $T''_{clk} \in [T^{bot}_{clk}, T^{up}_{clk}]$;
  $[w'_j$'s, $E'_c T'_{clk}] = solveLP(T'_{clk}, t_{se}, t_{cq}, \ldots, P^{leak}_{DL,j})$;
  $[w''_j$'s, $E''_c T''_{clk}] =$
    $solveLP(T''_{clk}, t_{se}, t_{cq}, \ldots, P^{leak}_{DL,j})$;
  Update $[T^{bot}_{clk}, T^{up}_{clk}]$ based on the values of $E'_c T'_{clk}$ and $E''_c T''_{clk}$;
Until $T^{up}_{clk} - T^{bot}_{clk} \le \epsilon_T$
$T^{min}_{clk} = T''_{clk}$; w^{min}_j's $= w''_j$'s; $E^{min}_{clk} T^{min}_{clk} = E''_c T''_{clk}$;
Derive $w^{0,min}_j$'s and s^{min}_j's from w^{min}_j's based on the fitted Pareto lines at V_{dd};
If there is a timing yield violation using T^{min}_{clk}:
  Extend T^{min}_{clk} to eliminate the timing violation;
Return: T^{min}_{clk}, $w^{0,min}_j$'s, s^{min}_j's, and $E^{min}_{clk} T^{min}_{clk}$.

Algorithm 1 shows the presented design flow of the SEFF-based pipelined circuits at a specific supply voltage. To design pipelined circuits working in both the ST and NT operation regimes, we perform the design flow for two desired supply voltages in these regimes, e.g. 0.4 V and 0.8 V, and configure the parameters of the DL and the PMOS header as

$$\begin{aligned} w^0_j &= \eta \times w^0_{j,NT} + (1-\eta) \times w^0_{j,ST} \\ s_j &= \eta \times s_{j,NT} + (1-\eta) \times s_{j,ST} \end{aligned} \tag{10}$$

where η is the ratio of the time that pipelined circuits operate in the NT operation regime.

4.2. Energy minimization under a frequency constraint

In this section we focus on a particular use case where pipelined circuits are constrained by a pre-specified minimum clock frequency. We adopt the energy consumption in one clock cycle as the cost function. We determine the optimal operation frequency/voltage and optimal transparency window sizes, in order to find the minimal energy point of pipelined circuits.

4.2.1. Formulation of the optimization problem

Different from the previous problem, we have another constraint on the minimal operating frequency, which makes the supply voltage another optimization variable. T_{clk} is still kept as an optimization variable because operating circuits at maximally allowed T_{clk} does not necessarily result in minimal energy consumption due to the existence of minimal energy point. Therefore, we formulate the energy minimization problem as follows.

- Given: The relationships between the supply voltage level V_{dd} and the following variables: w_j, t_{se}, t_{cq}, t_h, $D_{max,i}$, $D_{min,i}$; energy consumption $E^{dyn}_{cb,i}$, $E^{dyn}_{dff,j}$, $E^{dyn}_{DL,j}$; power consumption $P^{leak}_{cb,i}$, $P^{leak}_{dff,j}$, $P^{leak}_{DL,j}$ for $i \in [1, \ldots, N]$, $j \in [1, \ldots, N-1]$.
- Find: w^0_j, s_j, T_{clk} and V_{dd}, for $j \in [1, \ldots, N-1]$.
- Minimize: E_c given in (8).
- Subject to: $T_{clk} \le T_{clk,req}$ and timing constraints (4)–(7).

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

4.2.2. Trans-regional curve fitting

We characterize the timing and energy parameters of the pipelined circuits at discrete V_{dd} levels. In order to find the optimal supply voltage level, we need to derive the continuous relationships between V_{dd} and optimization variables. It is known that the physical principles of a transistor's driving current are different in the NT and ST regimes [1,4,5]. Therefore, we perform a trans-regional curve fitting that applies different fitting equations in these two regimes. The boundary between these two regimes is set to be $V_{dd}=0.6$ V.

In the NT regime, we adopt the exponential relationship between the driving current and supply voltage, according to [22]. We denote variables related to timing in the problem formulation above by *tv* and relate them to the supply voltage using the following equation

$$tv_{NT} = c \times V_{dd} \exp\left(aV_{dd}^2 - bV_{dd}\right) \tag{11}$$

where *a*, *b*, *c* are fitting parameters. In the ST regime, we apply the well-known alpha-power law [23] such that

$$tv_{ST} = f \times V_{dd}(V_{dd} - d)^{-e} \tag{12}$$

where *d*, *e*, *f* are fitting parameters in the ST regime.

We separate the energy consumption into two parts: dynamic and static. The dynamic energy consumption of combinational blocks or SEFFs, denoted by e_{dyn}, depends on the supply voltage V_{dd} in the same manner in NT and ST regime. The relationship is given by

$$e_{dyn} = \alpha \times (V_{dd})^{\beta} \tag{13}$$

where α and β are fitting parameters. Please note that (13) applies for both combinational blocks and SEFFs. On the other hand, because the leakage current is an exponential function of the supply voltage V_{dd}, the leakage power consumption p_{leak} is approximated using

$$p_{leak} = \delta V_{dd} \times \exp(V_{dd} - \gamma) \tag{14}$$

where δ and γ are fitting parameters. The leakage energy consumption in one clock cycle is derived as the product of p_{leak} and clock period T_{clk}.

Fig. 11 shows the trans-regional curve fitting results for the benchmark circuit c1355 by using the presented fitting equations (11)–(14). Due to the size limit for the paper, we do not present all fitting results. Eqs. (11)–(14) achieve very high fitting quality with a mean error of 4.9%.

4.2.3. Solution method

In this section, we find the minimal energy operating point of pipelined circuits and corresponding pipeline design parameters. The energy consumption in one clock cycle is again a non-convex function of V_{dd}, T_{clk} and w_j's. However, previous work has shown that the energy consumption is a quasi-convex function of the supply voltage with a single global minima [2,5]. Thus, we perform

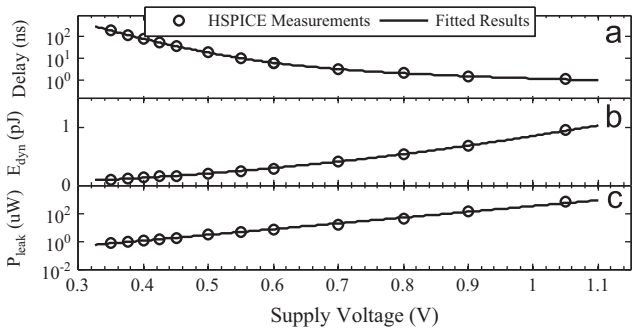


Fig. 11. Curve fitting examples using (11)–(14) for ISCAS'85 benchmark c1355.

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

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

a ternary search over the supply voltage V_{dd} in the outer loop. In the inner loop, we find the optimal operating frequency at a specific V_{dd} similar to the MINEDP algorithm. In order to achieve the minimal E_c, we replace the cost function of EDP in MINEDP algorithm with the energy E_c. We name this modified algorithm *MINE*.

To ensure that the minimal energy operating point returned by the MINE algorithm satisfies the frequency constraint, we compare T_{clk}^{min} with $T_{clk,req}$. If $T_{clk}^{min} > T_{clk,req}$, we set $T_{clk,req}$ as the clock period and derive the corresponding energy consumption E_{clk}^{req} and pipeline design parameters w_j^{req}'s. Otherwise, we return T_{clk}^{min} and corresponding design parameters as they are. The algorithm terminates once it converges to the supply voltage and operating frequency that results in the minimal amount of energy consumption in one clock cycle. The pipeline design parameters are determined accordingly. Details of the algorithm are summarized in Algorithm 2, where *derive WE* $(V_{dd}, T_{clk}, \ldots)$ is a function that returns w_j's and E_c when the pipelined circuits are operated with clock period T_{clk} and supply voltage V_{dd}.

Algorithm 2. Minimal energy point searching (*MEPS*).

Inputs: t_{se}, t_{cq}, t_h, $D_{max,i}$, $D_{min,i}$, $E_{cb,i}^{dyn}$, $E_{dff,j}^{dyn}$, $E_{DL,j}^{dyn}$; $P_{cb,i}^{leak}$, $P_{dff,j}^{leak}$, $P_{DL,j}^{leak}$ for $i \in [1,\ldots,N]$, $j \in [1,\ldots,N-1]$, as functions of V_{dd}, convergence parameters ϵ_V and ϵ_T.
Set a reasonable initial search range of V_{dd}, i.e., $[V_{dd}^{bot}, V_{dd}^{up}]$;
Do ternary search in $[V_{dd}^{bot}, V_{dd}^{up}]$:
  Pick $V_{dd,1}$ and $V_{dd,2} \in [V_{dd}^{bot}, V_{dd}^{up}]$;
  $[T_{clk,1}^{min}, w_{j,1}^{min}\text{'s}, E_{c,1}^{min}] = MINE(V_{dd,1}, \epsilon_T, t_{se}, t_{cq}, \ldots, P_{DL,j}^{leak})$;
  If $(T_{clk,1}^{min} > T_{clk,req})$:
    $T_{clk,1}^{min} = T_{clk,req}$;
    $[w_{j,1}^{min}\text{'s}, E_{c,1}^{min}] = deriveWE(V_{dd,1}, T_{clk,req}, t_{se}, t_{cq}, \ldots, P_{DL,j}^{leak})$;
  $[T_{clk,2}^{min}, w_{j,2}^{min}\text{'s}, E_{c,2}^{min}] = MINE(V_{dd,2}, \epsilon_T, t_{se}, t_{cq}, \ldots, P_{DL,j}^{leak})$;
  If $(T_{clk,2}^{min} > T_{clk,req})$:
    $T_{clk,2}^{min} = T_{clk,req}$;
    $[w_{j,2}^{min}\text{'s}, E_{c,2}^{min}] = deriveWE(V_{dd,2}, T_{clk,req}, t_{se}, t_{cq}, \ldots, P_{DL,j}^{leak})$;
  Update $[V_{dd}^{bot}, V_{dd}^{up}]$ based on the values of $E_{clk,1}^{min}$ and $E_{clk,2}^{min}$;
Until $V_{dd}^{up} - V_{dd}^{bot} \le \epsilon_V$
$V_{dd}^{min}=V_{dd,1}$; $T_{clk}^{min}=T_{clk,1}^{min}$; $w_j^{min}\text{'s}= w_{j,1}^{min}\text{'s}$; $E_c^{min} = E_{c,1}^{min}$;
Derive $w_j^{0,min}$'s and s_j^{min}'s from w_j^{min}'s based on fitted Pareto lines at V_{dd}^{min};
Return: $V_{dd}^{min}, T_{clk}^{min}, w_j^{0,min}\text{'s}, s_j^{min}\text{'s}, E_c^{min}$.

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

5. Experimental results

5.1. Experiment setup

We apply the presented design methods to some example pipelined circuits, where pipeline stages are synthesized using ISCAS'85 benchmarks. We adopt the Synopsys 32/28 nm technology and explore the supply voltage ranging from 0.35 V to 1.05 V in HPSICE. Table 1 shows the distribution of maximum and minimum delay of ISCAS'85 benchmarks. Based on these combinational logic benchmarks, we create four example pipelined circuits, as shown in Table 2. We compare the presented design method with the baseline method presented in [8], which adopts the conventional DL structure and optimizes the SEFF-based pipeline only in the ST regime. We use hard-edge DFFs-based pipelines as another baseline, in which the clock period is simply determined by the slowest pipeline stage. We normalize the optimized EDP value using the presented method to the hard-edge DFFs-based pipelines baseline.

5.2. Minimizing the energy-delay-product

We run the MINEDP algorithm to solve the first design problem for all example pipelined circuits. Table 3 shows returned clock periods and softness assignments for TB1 circuit. We determine the PMOS header width to be 0.9 μm. Compared to the baseline method, we achieve higher clock speed through better time borrowing. The baseline method in [8] cannot further increase the softness to reduce the EDP, due to the positive relationship between the softness and SEFF energy consumption. If the baseline method keeps increasing the softness, the benefits brought by delay reduction will be cancelled by the increase of energy consumption in SEFFs. Table 4 compares the normalized EDP for all test benches. The percentage reductions of the EDP range from 6.0% to 18.4%. Compared to hard-edge DFFs-based pipelines, SEFF-based pipelines achieve higher EDPs because the clock period is reduced. Precisely, the SEFF-based pipeline improves the operating frequency via time borrowing when there are slacks in some of the pipeline stages.

The leakage energy consumption also decreases because clock period decreases in the NT regime. As we can see in Fig. 5, the leakage energy takes only a small portion of the total energy consumption in the ST regime, while it plays a more important role in the total energy consumption in the NT regime. Fig. 12 compares the normalized delay reduction for TB1 achieved by SEFF-based pipelines. Notice that the normalized delay achieved by the presented method in Fig. 12 is lower than the corresponding normalized EDP values in Table 4 at high supply voltages. It indicates that the total energy consumption of the SEFF-based pipeline is slightly higher than the corresponding hard-edge DFFs-based pipeline due to the energy overhead introduced by SEFFs. On the other hand, at low supply voltage levels, the normalized EDP (around 80–90% for TB1 in Table 4) of the SEFF-based pipeline is lower than the normalized delay ($>95\%$ in Fig. 5) because the total energy consumption is reduced in the NT regime as well.

The presented method consistently outperforms the baseline method presented in [8]. In the ST regime, our method outperforms the method in [8] by achieving Pareto-optimal trade-off points, as shown in Fig. 10. In other words, we provide the same softness with lower energy consumption by using the presented DL design. In the NT regime where the conventional DL in [8] cannot provide enough softness, the modified DL structure allows us to provide greater softness with acceptable energy consumption. Thus, we reduce both the clock period and total energy

Table 2
Configurations of four example pipelined circuits.

Pipeline	Configuration
TB1	c1908, $SEFF_1$, c880
TB2	c432, $SEFF_1$, c1908, $SEFF_2$, c499
TB3	c1908, $SEFF_1$, c432, $SEFF_2$, c1355, $SEFF_3$, c880
TB4	c432, $SEFF_1$, c880, $SEFF_2$, c1908, $SEFF_3$, c499

Table 3
Comparison of achieved clock period and assigned softness for TB1. The PMOS header width is set to be 0.9 um.

V_{dd} (V)	Hard-edge		Method in [8]		MINEDP	
	Clock period (ns)	Soft-ness (ns)	Clock period (ns)	Soft-ness (ns)	Clock period (ns)
1.05	1.15	0.099	1.05	0.121	1.03
0.8	2.36	0.185	2.19	0.245	2.13
0.6	8.49	0.576	7.97	0.939	7.64
0.5	28.5	1.73	27.0	3.35	26.1
0.45	59.6	2.41	57.5	7.02	54.0
0.4	138.1	7.8	131.3	15.1	124.7
0.35	321.9	17.9	306.4	34.7	292.4

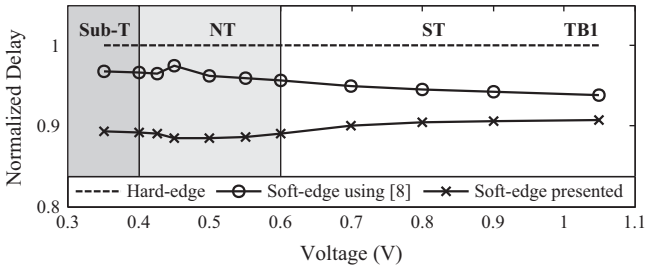


Fig. 12. Normalized stage delay reduction for TB1 achieved by the presented method versus the baseline method in [8].

Table 4
The energy-delay products achieved by the presented design method for four example pipelines.

V_{dd} (V)	Normalized energy-delay product (%)											
	TB1			TB2			TB3			TB4		
	Hard-edge	Method in [8]	Pro-posed	Hard-edge	Method in [8]	Pro-posed	Hard-edge	Method in [8]	Pro-posed	Hard-edge	Method in [8]	Pro-posed
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

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

consumption (i.e., less leakage) in the NT regime, which results in lower EDPs.

5.3. Finding the minimal energy operating point

We apply the presented MEPS algorithm to find minimal energy operating points for test bench pipelined circuits. We show the softness assignment and PMOS width returned by MEPS algorithm in Table 5. Note that some SEFFs have zero softness because time borrowing is not beneficial across those stages. We can simple use a normal DFF for that stage register.

Experimental results of optimal supply voltages and corresponding minimum energy consumptions are shown in Table 6. The baseline method in [8] is not tested here because it is not designed for find the minimal energy point. One can observe that, by replacing the hard-edge DFF registers with SEFF registers, we can reduce the supply voltage of the entire pipelined circuits while still meeting the frequency constraint. This is due to the opportunistic time borrowing enabled by SEFFs. Operating pipelined circuits at lower supply voltages reduces the energy consumption in one clock cycle. Compared to hard-edge DFF-based pipelined circuits, we achieve energy reductions up to 9.1%.

The presented design method achieves greater relative energy savings for those cases with a higher minimal operating frequency constraint (i.e., we reduce more energy at the 500 MHz frequency constraint compared with the 10 MHz frequency constraint). This is due to the fact that, for relaxed frequency constraints, we operate pipelined circuits in the NT regime. In the NT regime, the circuit delay is exponentially dependent, and thus highly sensitive, on the supply voltage. Therefore, we can only reduce the supply voltage by a small amount, which results in a relatively small energy reduction. Considering the additional energy overhead introduced by SEFFs, it is difficult to achieve further energy savings in the NT regime. In contrast, we significantly reduce supply voltage of pipelined circuits operating in the ST regime, and thereby achieve higher energy savings.

Fig. 13 shows the energy consumption in one clock cycle and the corresponding operating frequency versus the supply voltage for TB3. One can observe that scaling down the supply voltage helps to reduce the energy consumption in the ST regime. In contrast, there is a global minima of the energy consumption at V_{dd}=0.42 V in the NT regime. The corresponding operating frequency at that point is 10.6 MHz. Further downscaling the supply voltage below this minimal energy point does not introduce any more energy benefit. For the frequency constraint higher than 10.6 MHz, we find that the most energy efficient way is to operate pipelined circuits at the supply voltage level where the corresponding operating frequency exactly meets the frequency requirement (constraint). For relaxed situations in which the required frequency levels are lower than 10.6 MHz, operating pipelined circuits at this minimal energy point is the most energy efficient.

Table 5
Softness and PMOS width assignment returned by MEPS algorithm.

	Softness (ns) at different f_{req} (MHz)					PMOS width (um)
	10	100	200	300	500	
TB1						
$SEFF_1$	12.66	1.25	0.64	0.39	0.23	0.9
TB2						
$SEFF_1$	0	0	0	0	0	N/A
$SEFF_2$	18.30	1.78	0.86	0.51	0.28	0.7
TB3						
$SEFF_1$	7.29	0.76	0.38	0.24	0.14	0.8
$SEFF_2$	0	0	0	0	0	N/A
$SEFF_3$	13.21	1.29	0.62	0.37	0.21	0.8
TB4						
$SEFF_1$	0	0	0	0	0	N/A
$SEFF_2$	0	0	0	0	0	N/A
$SEFF_3$	18.70	1.81	0.87	0.51	0.28	0.7

6. Conclusion

Previous work on soft-edge flip-flop (SEFF)-based pipelines mainly focused on the super-threshold operation regime. In this work, we presented design methods for SEFF-based pipelines in both the super- and near-threshold (NT) operation regimes. We addressed two design problems in this work: how to design and operate SEFF-based pipelines (1) for the general usage that requires operating in multiple voltage regimes so that the energy-delay product is minimized; and (2) for a particular use

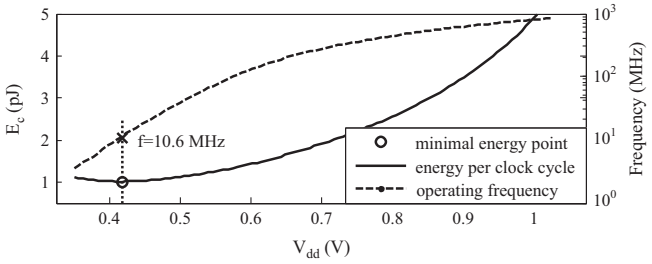


Fig. 13. Energy consumption in one clock cycle and the corresponding operating frequency versus supply voltage for TB3.

Table 6
The minimal energy operating points achieved by the presented design method for four example pipelines.

f_{req} (MHz)	Design Methods	Supply voltage/normalized energy consumption in one clock cycle (%)							
		TB1		TB2		TB3		TB4	
		V_{dd}^{min} (mV)	E_c^{min} (%)	V_{dd}^{min} (mV)	E_c^{min} (%)	V_{dd}^{min} (mV)	E_c^{min} (%)	V_{dd}^{min} (mV)	E_c^{min} (%)
10	Hard-edge	419.0	100.00	419.0	100.00	419.0	100.00	419.0	100.00
	Presented	412.2	98.37	412.2	99.21	414.9	99.23	410.8	99.50
100	Hard-edge	583.1	100.00	583.1	100.00	585.8	100.00	583.1	100.00
	Presented	574.9	97.03	573.5	98.37	577.6	98.88	572.2	98.12
200	Hard-edge	662.4	100.00	662.4	100.00	666.5	100.00	662.4	100.00
	Presented	647.4	95.99	645.9	96.71	652.8	96.90	644.6	96.94
300	Hard-edge	728.0	100.00	728.0	100.00	733.5	100.00	728.0	100.00
	Presented	708.8	94.68	707.5	95.49	718.5	96.46	706.1	95.69
500	Hard-edge	845.6	100.00	845.6	100.00	856.5	100.00	845.6	100.00
	Presented	816.9	90.94	818.3	92.42	834.6	93.39	815.5	91.92

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003

case with a specific operating frequency constraint so that the total energy consumption is minimized. We considered the high process variation in the NT regime and guaranteed the timing yield of pipelined circuits by imposing the timing constraints using the 3σ delay. We presented a novel delay lines structure for SEFFs by adding a PMOS header to achieve better dynamic energy-softness and leakage power-softness trade-offs, which is demonstrated to be very effective in the NT regime in combating the high process variation. We applied the presented method to test bench pipelines constructed using ISCAS'85 benchmarks and demonstrated the efficacy of presented methods.

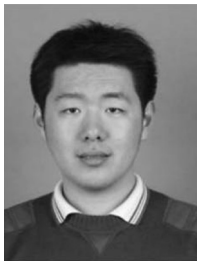
Acknowledgement

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

References

[1] Q. Xie, Y. Wang, M. Pedram, Variability-aware design of energy-delay optimal linear pipeline operating in the near-threshold regime and above, in: Proceedings of Great Lakes Symposium on VLSI, May 2013.
[2] B.H. Calhoun, A. Wang, A. Chandrakasan, Modeling and sizing for minimum energy operation in subthreshold circuits, J. Solid State Circuits 40 (2005) 9.
[3] B. Zhai, D. Blaauw, D Sylvester, K. Flautner, The limit of dynamic voltage scaling and insomniac dynamic voltage scaling, IEEE Trans. on VLSI 13 (2005) 1239–1252.
[4] R.G. Dreslinski, et al., Near-threshold computing: reclaiming Moore's law through energy efficient integrated circuits, Proc. IEEE 98 (2010) 2.
[5] D. Markovic, C.C. Wang, L.P. Alarcon, T.T. Liu, J.M. Rabaey, Ultralow-power design in near-threshold region, Proc. IEEE 98 (2010) 2.
[6] Seo, S., et al., Process variation in near-threshold wide SIMD architectures, in: Proceedings of the Design Automation Conference (DAC 2012), 2012, pp. 980–987.
[7] Joshi, V., Blaauw, D., Sylvester, D. 2007. Soft-edge flip-flops for improved timing yield: design and optimization, in: Proceedings of International Conference on Computer Aided Design.
[8] M. Ghasemazar, M. Pedram, Minimizing the energy cost of throughput in a linear pipeline by opportunistic time borrowing, in: Proceedings of International Conference on Computer Aided Design, (ICCAD 2008), 2008, pp. 155–160.
[9] A. Chandrakasan, S. Sheng, R. Brodersen, Low-power CMOS digital design, IEEE JSSC 27 (1992) 473–484.
[10] H. Jacobson, et al., Stretching the Limits of Clock-gating Efficiency in Server-class Processors, in: High-Performance Computer Architecture, (HPCA 2005), 2005, pp. 238-242.
[11] N.S. Kim, T. Kgil, K. Bowman, V. De, T. Mudge, Total power-optimal pipelining and parallel processing under process variations in nanometer technology, in: Proceedings of International Conference on Computer Aided Design, (ICCAD 2005), 2005, pp. 534–540.
[12] V. Srinivasan, et. al., Optimal pipelines for power and performance, in: Proceedings of the International Symposium on Microarchitecture, (MICRO 2002), 2002, pp. 333–344.
[13] Seok, M., A 0.27 V, 30 MHz, 17.7 nJ/transform 1024-pt complex FFT Core with super-pipelining. In: Proceedings of IEEE International Solid-State Circuits Conferences, (ISSCC 2012), 2012, pp. 342–344.
[14] K. Duraisami, E. Macii, M. Poncino. Using soft-edge flip-flops to compensate NBTI-induced delay degradation, in: Proceedings of ACM Great Lakes Symposium on VLSI (GLSVLSI 2009). 2009, pp. 169–172.
[15] M. Ghasemazar, M. Pedram, Optimizing the power-delay product of a linear pipeline by opportunistic time borrowing, IEEE Trans. on Comput.Aided Des. Integr. Circuits Syst. 30 (2011) 1493–1506.
[16] S. Dillen; D. Priore, A. Horiuchi, S. Naffziger, Design and implementation of soft-edge flip-flops for x86-64 AMD microprocessor modules, in: Proceedings of Custom Integrated Circuits Conference (CICC 2012), 2012, pp. 9–12.
[17] Datta, A., Bhunia, S., Mukhopadhyay, S., Roy, K. 2005. Statistical modeling of pipeline delay and design of pipeline under process variation to enhance yield in sub-100 nm technologies, in: Proceedings of Design and Test in Europe.
[18] Synopsys 32/28 nm Generic Library: ⟨https://sso.synopsys.com/idp/Authn/UserPassword⟩.
[19] Hanson, S., et al. 2007. Performance and variability optimization strategies in a sub-200 mV, 3.5 pJ/inst, 11 nW subthreshold processor, in: Proceedings of International Symposium on VLSI Circuits.
[20] Fisher, S., Dagan, R., Blonder, S., Fish, A. 2011. An improved model for delay/energy estimation in near-threshold flip-flops, in Proceedings of International Symposium on Circuits and Systems.
[21] Lotze, N., Ortmanns, M., Manoli, Y. 2008. Variability of flip-flop timing at subthreshold voltages, in Proceedings of International Symposium on Low Power Electronics and Design.
[22] D.M. Harris, B. Keller, J. Karl, S. Keller, A transregional model for near-threshold circuits with application to minimum energy operation, in: Proceedings of Microelectronics International Conference (ICM), 19–22 Dec. 2010, pp. 64–67.
[23] Takayasu Sakurai, A.Richard Newton., Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas, Solid-State Circuits, IEEE J. 25.2 (1990) 584–594.

Qing Xie (S'12) received B.S. and M.S. degree in Physics from Fudan University, China in 2007 and Northeastern University, Massachusetts in 2009, respectively. He is currently pursuing Ph.D. degree in Electrical Engineering Department at University of Southern California, under the supervision of Prof. Massoud Pedram. His research interests are in the area of low-power systems design, energy storage systems, system-level power management, thermal management, and near-threshold computing. He has received the Best Paper Award from the 30th IEEE International Conference on Computer Design.



Yanzhi Wang (S'12) received the B.S. degree with distinction in electronic engineering from Tsinghua University, Beijing, China, in 2009. He is currently pursuing the Ph.D. degree in electrical engineering at University of Southern California, under the supervision of Prof. Massoud Pedram. His current research interests include system-level power management, next-generation energy sources, hybrid electrical energy storage systems, near-threshold computing, and the smart grid. He has published more than 50 papers in these areas.

Massoud Pedram (F'01), who is the Stephen and Etta Varra Professor in the Ming Hsieh department of Electrical Engineering at the University of Southern California, received a Ph.D. in Electrical Engineering and Computer Sciences from the University of California, Berkeley in 1991. He holds 10 U.S. patents and has published four books, 12 book chapters, and more than 130 archival and 320 conference papers. His research ranges from low power electronics, energy-efficient processing, and cloud computing to photovoltaic cell power generation, energy storage, and power conversion, and from RT-level optimization of VLSI circuits to synthesis and physical design of quantum circuits. For this research, he and his students have received six conference and two IEEE Transactions Best Paper Awards. Dr. Pedram is a recipient of the 1996 Presidential Early Career Award for Scientists and Engineers, a Fellow of the IEEE, an ACM Distinguished Scientist, and currently serves as the Editor-in-Chiefs of the ACM Transactions on Design Automation of Electronic Systems and the IEEE Journal on Emerging and Selected Topics in Circuits and Systems. He has also served on the technical program committee of a number of premiere conferences in his field and was the founding Technical Program Co-chair of the 1996 International Symposium on Low Power Electronics and Design and the Technical Program Chair of the 2002 International Symposium on Physical Design.

Please cite this article as: Q. Xie, et al., Designing soft-edge flip-flop-based linear pipelines operating in multiple supply voltage regimes, INTEGRATION, the VLSI journal (2014), http://dx.doi.org/10.1016/j.vlsi.2013.11.003