

# Analysis and Optimization of Programmable Delay Elements for 2-Phase Bundled-Data Circuits

Guilherme Heck\*, Leandro S. Heck\*, Ajay Singhvi<sup>†</sup>, Matheus T. Moreira\*<sup>‡</sup>,  
Peter A. Beerel<sup>‡§</sup> and Ney L. V. Calazans\*

\*Pontifícia Universidade Católica do Rio Grande do Sul - Porto Alegre, Brazil

<sup>†</sup>Birla Institute of Technology and Science Pilani, Pilani Campus - Pilani, India.

<sup>‡</sup>University of Southern California - Los Angeles, United States

{guilherme.heck, leandro.heck, matheus.moreira}@acad.pucrs.br, ajaysinghvi93@gmail.com  
pabeerel@usc.edu, ney.calazans@pucrs.br

**Abstract**—We present the design and analysis of three commonly used types of programmable delay elements suitable for use in 2-phase bundled-data asynchronous circuits. Our objective is to minimize power consumption and delay margins needed for correct operation under voltage scaling. We propose both circuit design and transistor sizing strategies to optimize these elements and discuss the relative trade-offs observed in a 65 nm bulk CMOS technology.

**Keywords**—Delay elements, 2-phase bundled-data, asynchronous circuits, voltage scaling

## I. INTRODUCTION

Delay elements (DEs) are often employed for the design of clock generation and distribution. For the latter, different DEs exist to fix skew problems post-silicon [1]. DEs are also needed in the design of bundled-data (BD) asynchronous circuits [2]–[4]. However, due to the widespread acceptance of synchronous design techniques, less attention has been devoted to the needs of delay lines in that field. But as IC manufacturing technologies scale down into ultra deep submicron nodes, problems traditionally ignored in synchronous design become increasingly challenging and costly to overcome. This enables asynchronous techniques to regain attention.

Among the families of design templates for asynchronous circuits, BD ones stand out as energy, speed, and area-efficient solutions. Such templates employ explicit request and acknowledge control signals for sending data between blocks [5] and internally rely on delay lines, sequences of DEs whose overall delay must be matched to the critical path(s) of the block's datapath logic. Delay lines must be conservatively designed to be longer than the logic critical path(s) but this delay difference must be minimized, as it represents an overhead that reduces performance. Unfortunately, increasing process variations in deep submicron technologies forces the addition of margins to delay lines designed in traditional ways. For this reason, *programmable DEs* that can be tuned post-silicon to match the actual datapath delays in silicon provide a far more attractive solution. Moreover, for BD design templates based on 2-phase control schemes, delay lines need to have balanced rise and fall propagation delays. This occurs because both delays must be equally matched to the critical path, since any difference between them represents additional overhead.

Another challenge to modern IC designers are the tight power budgets that extend from mobile applications concerned with battery lifetime, to high-end server processors concerned with heat dissipation [6]. A trend to cope with such a challenge is the dynamic scaling of the supply voltage as discussed in [6] and [7]. While this approach can be challenging for synchronous designs, asynchronous circuits are known to efficiently support voltage scaling (VS) techniques as discussed in [7]. Although earlier works discussed the application of asynchronous techniques for VS applications [6] [7], they are focused on application-specific examples and, as far as we could verify, none of them addresses the effects of VS in DEs. Yet, this is a very important concern because the DEs must remain matched with the datapath as voltage scales without unnecessarily increasing margin. In other words, VS affects the datapath delay and its associated DEs must be able to conservatively, yet closely, track such variations. The objective must be to ensure correct operation while avoiding performance losses.

We present an analysis of the effects of VS on three commonly used programmable DEs for 2-phase BD design: *i*) one based on multiplexers and inverters [8]; *ii*) one based on directly-controlled current-starved inverters [9]; and *iii*) one based on current mirror-controlled current-starved inverters [10]. Accordingly, we assess how margins are affected and propose a method to optimize the design of these DEs to reduce performance losses as voltage scales. The method includes a transistor dimensioning technique and an approach to balance rise and fall propagation delays. The main contribution of this work is to provide guidelines for the design of DEs when VS is important, moving forward the state of the art in low power asynchronous design.

## II. ANALYSIS

### A. Delay Elements

In the last decades, DEs with both analog and digital input control have been proposed. An example of a DE with an analog control signal is the work proposed by Vezyrtzis et al. in [4], which also targets low-power applications. Some drawbacks of analog-controlled DEs is the distribution of a global analog signal and the challenges of dealing with VS applications. Hence, our focus is digitally-controlled DEs, which avoid such problems.

<sup>§</sup> Peter A. Beerel is also a Chief Scientist, Technology Development at Intel, Calabasas, CA 91302.

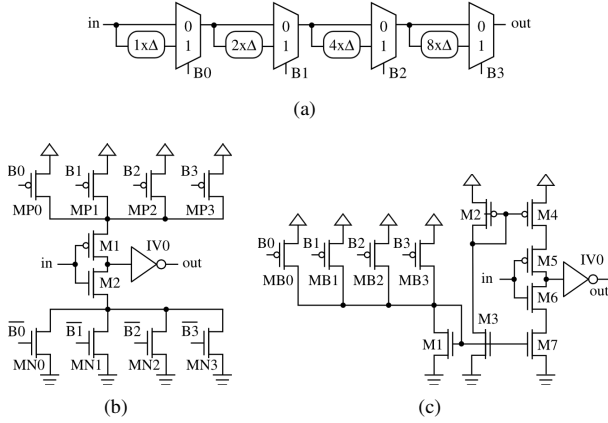


Fig. 1. Schematics of standard DEs, (a) M-DE [8] (b) DCCS-DE [9] and (c) CMCS-DE [10].

The above-mentioned DEs have previously been explored by others and use a variety of circuit structures, including: transmission gates (TG) [3]; thyristors [11]; cascaded inverters [12]; and current-starved inverters [9] [13]. The TG-based DE is a simple approach but can suffer from poor signal integrity. Adding Schmitt triggers to the design improves the signal integrity, but at the cost of area and power efficiency [3]. The thyristor-based DE uses positive feedback control and differential logic [11] that makes it challenging to balance rise/fall transitions across a wide range of voltages.

This paper focuses on three designs based on cascaded and/or current-starved inverters: *i*) one based on cascaded inverters controlled by multiplexers as presented in [8]; and two versions, *ii*) and *iii*), of digitally controlled current-starved inverters (CSIs), as presented in [9] and [10]. The DE *i*), referred to as the MUX DE (M-DE), was chosen due to the fact that it is one of the most popular DEs, given its relatively simple implementation that can be built using standard cells. Fig. 1(a) shows the schematic of a 4 bit M-DE. It consists of four sets of inverters, where each set has twice the delay  $\Delta$  of its predecessor, and four MUXes that select the path used to delay the input signal. The chosen codeword configures the MUXes, selecting the number of cascaded inverters in the delay path and hence the total delay.

DEs *ii*) and *iii*) are intuitively more power- and area-efficient than the M-DE. However, their complexity makes balancing their rise/fall delays challenging, especially under VS. DE *ii*) is a directly-controlled current-starved delay element (DCCS-DE) illustrated in Fig. 1(b) [9]. It uses a parallel set of transistors as current sources to digitally control the propagation delay. Lastly, DE *iii*) is a current mirror current-starved delay element (CMCS-DE). It is depicted in Fig. 1(c), and comprises a CSI controlled by a configurable current mirror, similar to the one presented in [10].

DEs *ii*) and *iii*) are programmable via current sources that control the amount of charge that flows through the CSIs. The current is determined via a binary codeword that selects the transistors which source the current. These transistors are carefully sized to provide a range of delays. The difference between the two CSI designs is that one uses a direct pull-up and pull-down current source, while the other uses a

current mirror to control the current. Previous work [13] proposed optimizations for CSI DEs. However, the proposed optimizations do not target 2-phase operation. Accordingly, we disregard these in our analysis.

## B. Voltage Scaling (VS)

Scaling the voltage down to the near-threshold region has a significant impact on device parameters and circuit performance. To determine how a DE propagation delay scales with voltage, we analyze the delay of CMOS transistors ( $t_{pd}$ ) using the following equation:

$$t_{pd} = C_L \frac{V_{ds}}{I_{ds}} = C_L \frac{2LV_{ds}}{kW(V_{gs} - V_t)^\alpha} \quad (1)$$

where  $C_L$  is the load capacitance,  $V_{ds}$  is the drain-source voltage and  $I_{ds}$  is the drain current given by the  $\alpha$ -power law model suitable for short-channel technologies [14].  $k$  is a technology-dependent parameter,  $L$  and  $W$  are the drawn gate length and width.  $\alpha$  is the velocity saturation used as a simple approximation to capture a region where the velocity increases linearly with field, nor is completely saturated [15].  $V_t$  is the threshold voltage and  $V_{gs}$  is the gate-source voltage. An important point is the increasing impact of transistor sizing on the threshold voltage as the feature size decreases, which further changes the propagation delay based on  $L$  and  $W$  [15].

According to (1), delay increases as voltage scales down. We propose here the definition of the *voltage scaled delay ratio* (VSDR) parameter, to represent the relationship between the delay at near-threshold ( $t_{pd\_near}$ ) and the delay at nominal voltage ( $t_{pd\_nom}$ ). The computation of VSDR using the  $t_{pd}$  equation is:

$$VSDR = \frac{t_{pd\_near}}{t_{pd\_nom}} = K \frac{(V_{gs\_nom} - V_{t\_nom})^{\alpha_{nom}}}{(V_{gs\_near} - V_{t\_near})^{\alpha_{near}}} \quad (2)$$

where  $K$  is a constant that depends only on the operating voltage. According to [16] variations in  $C_L$ ,  $W$  and  $L$  with respect to voltage are insignificant. Hence, these were assumed as negligible. On the other hand,  $V_t$  depends not only on transistor sizing but also on the operating voltage, albeit the dependence on the latter is small.  $\alpha$  also varies slightly with voltage, decreasing as we decrease the operating voltage [15]. In fact, (2) implicitly states that VSDR depends on transistor sizing parameters  $W$  and  $L$ .

It is traditionally assumed that  $V_t$  increases as  $L$  increases until it asymptotically reaches a constant. However, in some processes, including the employed bulk 65 nm CMOS technology, the reverse short-channel effect causes the opposite effect -  $V_t$  decreases as  $L$  increases [15] [17]. The magnitude of this decrease is larger at near-threshold voltages, where  $V_{gs}$  is close to  $V_t$ . This means that VSDR decreases as  $L$  increases. In addition, narrow-channel effects causes  $V_t$  to *slightly* increase with width until it reaches a constant value [15] [17]. Using the same analysis as above, this means that VSDR *slightly* increases as  $W$  increases, until becoming constant. Moreover, because the impact of changing  $W$  is smaller than that of changing  $L$ , VSDR asymptotically decreases as  $W$  and  $L$  are simultaneously increased keeping the  $W/L$  ratio constant.

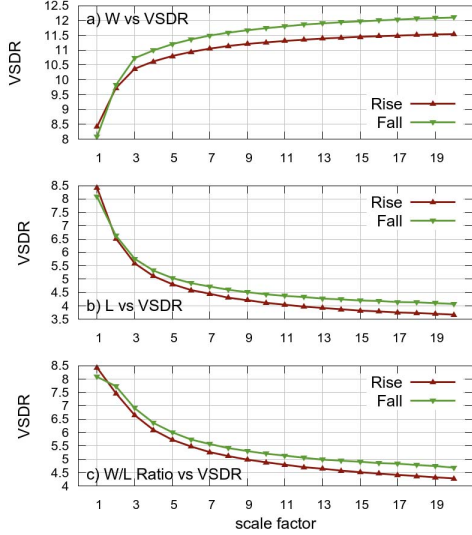


Fig. 2. Transistors' sizing of (a)  $W$ , (b)  $L$  and (c)  $\frac{W}{L}$  ratio versus VSDR.

Lastly, without loss of generality VSDR can be employed for circuits composed by many transistors, also representing the ratio between the delay at near-threshold and nominal voltages.

### C. VSDR Transistor Sizing

Traditionally, optimum transistor sizing for VS focuses on achieving the best power-delay product [18] [19]. However, our goal is to explore how to size transistors such that both delay element and datapath slow down by the same amount as voltage scales, i.e. both display similar VSDRs. This means the delay element would remain matched to the datapath logic at all voltage levels, requiring the smallest amount of margin. We assume that for proper operation the transistor sizing of the datapath critical path cannot be altered. Therefore, an important feature for DEs is the ability to create a wide range of VSDRs, controllable by proper transistor dimensioning. To do so, we first analyzed the effect of transistor sizing on the VSDR, simulating an inverter in a 65 nm bulk CMOS technology. Fig. 2 shows the analysis results for three scenarios: *i*) increasing  $W$ , keeping  $L$  at minimum ( $L_{min}$ ); *ii*) increasing  $L$ , keeping  $W$  at minimum ( $W_{min}$ ) and *iii*) increasing the  $W/L$  ratio. The x-axis represents the factor by which the varied parameter was scaled and the y-axis represents the resulting VSDR. Simulation results corroborate the analysis presented in Section II-B. Scenario *i*) results in Fig. 2(a) show VSDR increasing as  $W$  increases. Scenario *ii*) results (Fig. 2(b)) show VSDR decreasing as  $L$  increases and results for *iii*) in (Fig. 2(c)) show that VSDR is dominated by  $L$ , with a behavior similar to that observed for *ii*). The advantage of scaling the  $W/L$  ratio to achieve a particular VSDR is that the delay range and ratio remain unaffected by changes in sizing. Thus, the VSDR can be tuned using a combination of these scaling strategies, which enables reducing delay margins.

Another important aspect of the design is to balance the rise and fall times for 2-phase bundled data operation. This again involves sizing transistors such that they have similar driving strength and thus similar propagation times. However, VS does

not affect  $W$  and  $L$  of nMOS and pMOS devices by the same amount, resulting in different VSDRs, as noticeable in Fig. 2. Although these differences are not significant for simple gates like the case study inverter presented here, for complex gates, like the DEs presented in Section II-A, variations in rise and fall VSDRs can lead to significant mismatches in rise/fall propagation delays, as voltage scales. For instance, the worst case rise/fall VSDR variation presented in Fig. 2(b) for big  $L$  devices is roughly 4.1 for the fall delay and 3.7 for the rise delay. If we assume a DE with a balanced delay of 500 ps at nominal voltage, this means that in the near-threshold region, its fall propagation delay will be 2.05 ns and its rise propagation delay will be 1.85 ns, which is a difference of 200 ps that would require adding a margin of 10%. In other words, a delay element designed to give balanced operation at nominal voltage would not be able to maintain the balance in rise and fall times as the operating voltage scales down. The imbalance in rise and fall times would grow significantly as voltage approaches the near-threshold region. Thus, it becomes necessary to introduce architectural changes in the design of delay elements, to ensure that the mismatch in the driving strength of nMOS and pMOS devices does not affect the ability to provide balanced operation at all voltage levels.

### III. DESIGN

Because the delay of pMOS and nMOS transistors vary differently as voltage scales, a single CSI is not suited for maintaining rise and fall propagation delays balanced. In fact, even when using an output inverter for regenerating signal levels, as suggested in [9], simulations showed that delays are still not sufficiently balanced. A classic technique applicable to cope with this is to replicate the same circuit in a series connection. For example, a sequence of two CSIs can generate an output signal with the same variation for rising and falling edges as voltage scales. In this way, we expect that rising and falling edges present very similar VSDRs. To verify this behavior, we performed experiments with one CSI and with a series connection of two of them. We observed that the mismatch in the VSDR for rise/fall transitions ranged from 10% to 15% for a single CSI, while the mismatch was less than 1% for the series version. Therefore, to achieve balanced DEs we suggest the use of this technique.

The first DE we consider here is the DCCS-DE. Fig. 3(a) shows the schematic of its series version. We use a 2-component series of CSIs, one composed by M1-M2 transistors and the other by M3-M4 transistors. However, the first CSI is typically fed by a well-defined signal, which is generated by a conventional standard cell. By merely placing the CSIs in series, the second one would be fed by a signal with a worse slew rate. Simulations showed that this can compromise the balance between rise and fall delays. Another two problematic scenarios occurs when the CSIs feed gates with different drive strengths and when they feed different loads. Therefore, we added identically sized inverters IV1 and IV2 to drive the CSIs and IV0 and IV3 to provide equivalent loads. This topology helps ensure that we obtain a balanced DCCS-DE.

The range of delays generated by a DCCS-DE depends on tuning three parameters: *i*) the  $W$  and  $L$  of the current source transistors (MPx and MNx); *ii*) the  $W$  and  $L$  of the CSIs (M1-4) and *iii*) the output capacitive load of the CSIs ( $C$ ). The

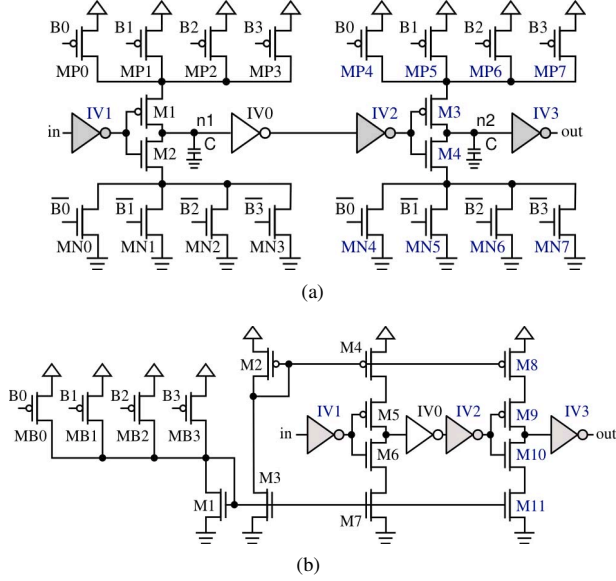


Fig. 3. Modified (a) DCCS-DE and (b) CMCS-DE.

setting of  $i$ ) is important since these transistors are responsible for feeding the CSIs. We set their widths to  $W_{min}$  and the  $L$ s to 4:8:16:32 times  $L_{min}$ , respectively. This allows 15 binary selections of delay where 0 represents the codeword when all transistors are on, and 14 when only the  $32L_{min}$  is on. We used two separate current sources, because sharing one current source compromised the range of delays achievable with the DCCS-DE. In particular, to maintain the same delay range and share a single current source between the two CSIs we would need transistors with bigger  $L$ s, which complicates the design and generates additional area overheads. The transistors of  $ii$ ) are responsible for driving the output of the CSI. Therefore, their size is directly related to the delay range. These transistors need to be sufficiently wide to avoid constraining the maximum current set by the current source transistors in  $i$ ). In our case, minimum width transistors were sufficient for this purpose.

We observed that during the switching of the CSIs the drain capacitance of transistors in  $i$ ), which are at virtual power rails, are connected with the capacitance  $C$  in Figure 3(a) and charge sharing occurs. This speeds up the transitions on  $n1$  and  $n2$ , which undesirably decreases the delay set by the current source transistors and therefore reduces the achievable delay range. We mitigate this effect by artificially increasing  $C$  by adding a shunt capacitor. Note that sharing a single current source for both CSIs would increase the severity of this problem, as more shared capacitance would be added (the drain-source capacitance of transistors in  $ii$ ). We accordingly increased  $C$  by adding a 5 fF shunt capacitor to provide a delay range of 300 ps to 1.4 ns, which we fixed for all the DEs evaluated herein, to provide a fair analysis.

We next explore the CMCS-DE design. Similar to the DCCS-DE, we employed a series of two CSIs with similar input and output inverters, as Fig. 3(b) shows. Transistors MB0-MB3 control the amount of current through M1, which through its diode connection sets the gate voltage of M3, M7 and M11. Transistors M7 and M11 control the current

through the nMOS transistors of the CSIs (M6 and M10) which determines the CSIs falling delay. Transistors M2 and M3 translate the voltage bias from the nMOS network to the pMOS transistors M4 and M8, which control the CSIs rise delay. Thus, transistors M4, M7, M8 and M11 make up the current sources of the CSIs and transistors M1-3 constitute the current mirrors that limit the CSIs current.

Taking these factors into account, transistors were sized to give a delay range matching that obtained from the DCCS-DE design discussed earlier. Transistors of the current sources must be sufficiently large to avoid increasing the minimum delay of the CSIs, i.e. their resistance must be comparable to the resistance of the CSI transistors, otherwise the delay range would shrink. Similar to the DCCS-DE, the CSIs transistors of CMCS-DE were also minimum-sized and experiments confirmed that sizing for current sources were sufficient to avoid the aforementioned issue. The sizing of current mirror transistors is more complex. M1 needs to be sufficiently large to guarantee an adequate distribution of voltage on its gate across the range of currents set by MB0-MB3. In our design, it needs 12x the minimum width to produce the necessary M1 gate voltage. The ratio of the  $W/L$  ratio of M3 to that of M1 determines the fraction of current through M1 that flows through M3. By keeping M3 minimum-sized we ensure that this ratio is a fraction of 1, and hence the current mirrored to M3 is always a fraction of that in M1. The size of M2 affects how this current translate to the voltage on M4 and M8 and to the current through these transistors. This affects the matching of the CSIs rise and fall delays. However, because we are placing two CSIs in series it is less important for the rise and fall delays of an individual CSI to be balanced and thus the size of M2 can be minimum. Additionally, by keeping M2 and M3 small, we are minimizing the design idle power, as these transistors are significant sources of idle power.

Another option that we considered was sharing the current sources for the two CSIs. However, such a design is prone to generating glitches on its output and thus excluded from consideration. In particular, when there is a transition at the input of the first CSI, the virtual power rail at the drain of the current sources can see a voltage bump due to charge sharing. The size of this charge sharing bump will be particularly significant for larger delays when the current through the current sources is configured to be small. This bump can be seen at the output load of the second CSI as it remains connected to the virtual power rail for the short period of time it takes for IV0 and IV2 to switch. For instance, assume that the input  $in$  is 1, which means that the output of the first CSI is 1 and the second CSI's output is 0. Now, assume that  $in$  switches to 0. For the delay of IV0 and IV2, the outputs of the CSIs will both be connected to the virtual ground and the positive bump on it can propagate to the output. Note that an equivalent glitch can also be generated when the input switches from 0 to 1. However, this glitch would be a negative bump as it occurs in the pMOS network. Note that a similar glitch phenomenon would also occur if we tried to share the current sources in the DCCS-DE.

A similar technique of alternating between pMOS and nMOS structures can be employed for the design of the M-DE. Accordingly, for achieving a balanced VSDR, each fixed delay element must be composed by an even number of inverters.

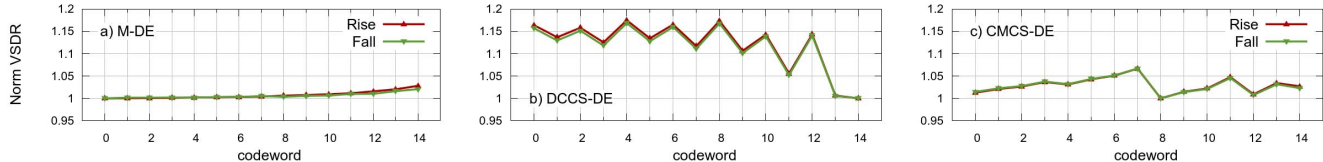


Fig. 4. Delays margins.

In our case, we implemented them using 6, 12, 24 and 48 inverters between MUXes, where all inverters had the same pMOS and nMOS transistor sizes, set to meet a desired VSDR as described above. The set of inverters enables a delay range of 300 ps to 1.4 ns. Moreover, the MUXes should also have a balanced VSDR, which is typically not achievable with MUXes available in standard-cell libraries, as their design is not optimized for this purpose. We therefore constructed our MUXes using classic sum-of-products composed of three NAND gates and one inverter for the control signal. Given that NANDs were constructed with the same transistors sizes, this allows obtaining a balanced VSDR for falling and rising edges in the DE.

#### IV. EXPERIMENTS AND DISCUSSION

A 65 nm bulk CMOS technology was used to design the delay elements, with 1.2V used as the operating voltage and 0.6V taken as the near-threshold voltage. All simulations were carried out using the Cadence Spectre Analog Simulator. To compare the different delay elements effectively, they were designed to give delays in the range of 300 ps to 1.4 ns. The simulation environment was identical for all DEs to keep consistency across designs and ensure a fair comparison. A trapezoidal input passed through an input buffer allows producing a realistic input signal. This feeds each of the DEs. In addition, a fixed fan-out was maintained at the output of each DE. All simulations assumed an operating temperature of 25°C and typical fabrication process parameters.

The first step in our experiments was to get the desired range of delays, by sizing the transistors as described in Section III. We then performed set of simulations on each DE, measuring the following characteristics: rise and fall propagation delay for each codeword; energy per transition (EPT) for rise and fall delays, for each codeword; and leakage power for two static states, with a steady 1 and 0 at the input. All experiments were performed at both nominal and near-threshold voltages. Each DE was evaluated based on different parameters: *i)* rise/fall delay ratio across codewords; *ii)* rise/fall VSDR ratio across codewords; *iii)* average VSDR across codewords; *iv)* average EPT; and *v)* average leakage power;

Parameters *i)-iii)* give a notion of the overhead in terms of margins, which translates to losses in performance. For *i)*, as explained before, the closer this ratio is to one the better, as lower margins will be required for 2-phase BD designs. In fact, what we observed in the results obtained from simulation was that by employing the balancing technique discussed in Section III this ratio was always very close to one, and presents variations lower than 1%. This is in contrast with the results obtained for the original designs which had variations of over 10% and demonstrates that the technique can indeed be used

TABLE I. TRADE-OFFS BETWEEN ANALYZED DES.

Delay Line	M-DE	DCCS-DE	CMCS-DE
Average margins	<1%	12%	2.9%
Worst-case margins	2.4%	17%	6.7%
Avg. Idle Power @near (nW)	1.362	0.113	2226.141
Avg. Idle Power @nom (nW)	6.165	0.672	80457.280
Avg. EPT @near (fJ)	12.565	2.095	11.561
Avg. EPT @nom (fJ)	49.543	6.798	35.282
Active Area ( $\mu\text{m}^2$ )	2.2656	3.6962	1.272

for balancing DEs. Therefore, results indicate that *i)* does not significantly affects margins overheads. For *ii)*, similar results were obtained. The charts of Fig. 4 show the normalized rise and fall VSDRs for the three DEs. Because rise and fall VSDRs are also similar and present a ratio close to one, effects of *ii)* also do not impose overheads in the margins.

Results for parameters *i)* and *ii)* are expected, since delays are balanced by replicating the circuits and by sizing the transistors properly. Parameter *iii)* provides a metric for analyzing margins that are specific for each DE because it cannot be tuned as in *i)* and *ii)*. In fact, it is clear in Fig. 4 that wide variations in VSDR are observed for the DCCS-DE across different codewords. This translates to margins that need to be added, as codewords are used to compensate process variations and at design time worst case VSDR values need to be considered to guarantee correct operation. Considering *iii)*, the results for the average and worst-case margins are summarized in the first two rows of Table I. Accordingly, this shows that the DCCS-DE imposes up to 17% of margins in the worst-case and 12% in average-case. For the CMCS-DE, these variations are less severe (6.7% in the worst-case and 2.9% in average) and for the M-DE they are practically negligible (2.4% in the worst-case and less than 1% in average). Thus, M-DE would require the least margin to be added and would provide the best performance in terms of speed.

We employed ideal voltage sources but in real circuits these will suffer from effects like IR drop that can cause noise in the power supply. Under this conditions, the delay of the evaluated DEs is expected to vary. However, because the components were designed targeting VSDR, these variations are expected to track the variations in the datapath. Problems could appear when the DEs' supply voltage is higher than the one in the datapath. This will potentially cause the DE to operate faster than the datapath. For dealing with this problem extra margin will need to be added to the DE. We believe that this margin will be the same for all evaluated DEs, given that all were designed to have the same VSDR.

Results for parameters *iv)* and *v)* were measured as the average between rise and fall transitions energy and average

between the leakage of the static states, respectively. These results are also summarized in Table I. As expected, CSI DEs were more energy-efficient than the M-DE because of lesser current flowing through current-starved inverters as compared to the inverters in the M-DE. In other words, for the current-starved DEs, lesser capacitances need to be charged/discharged during propagation. Of the two current-starved designs, the DCCS-DE proved to be more efficient than the CMCS-DE due to the fact that the DCCS-DE had current flowing directly from the pull-up and pull-down networks to the inverter, while the CMCS-DE had the current flowing through the reference arm, mirror arm as well as the inverter.

It was also seen that changing the operating voltage from 1.2V to 0.6V resulted in over 70% energy savings for all the DEs. For leakage power, the DCCS-DE was the most efficient design, with the M-DE having about 10 times more idle power consumption both at nominal and near-threshold regions. The idle power consumption of the CMCS-DE was very large, 4 orders of magnitude larger than the one observed for the DCCS-DE. This is expected due to the current continually flowing through the reference arm of the current mirror to keep up the reference signal even when the circuit is not switching. However, this can be prohibitively expensive in low-power asynchronous circuit design.

Table I also shows area results, measured as the active area of the DEs (*i.e.* the summation of  $W \times L$  of all transistors). The table shows the CMCS-DE is the smallest design which is mainly due to the fact that a single current mirror could be used to bias the entire DE without compromising delay and VSDR balancing. This is in contrast with the DCCS-DE that needed separate pull-up and pull-down current sources for the buffer like design to maintain delay and VSDR balancing. However, one important point for the CMCS-DE is that considerations need to be made about the shielding that might be needed to protect the analog parts of the DE and that is not considered in the results. Other trade-offs that were also observed include the fact that the DCCS-DE produced a somewhat non-monotonic delay behavior with an ascending control word pattern while the M-DE and CMCS-DE produced a more monotonic delay behavior.

## V. CONCLUSION

We assessed power, energy and area trade-offs together with delay margins, for three programmable DEs. Based on VSDR, a newly proposed metric, we suggested optimization strategies that rely on transistor sizing and circuit replication to make DEs track variations in the datapath delay as voltage scales, to reduce the margins needed to add to designs. Simulation results in a 65 nm bulk CMOS technology validated the method efficiency and indicate a clear set of design trade-offs among the evaluated DEs. M-DE provides the best margins reductions, DCCS-DE provides better energy efficiency, and CMCS-DE enabled high density.

## VI. ACKNOWLEDGEMENTS

Authors acknowledge the support of CNPq under grants 401839/2013-3, 200147/2014-5, 202519/2014-7 and 310864/2011-9 and the support of FAPERGS under grant 11/1445-0.

## REFERENCES

- [1] A. Chakraborty *et al.*, "Dynamic Thermal Clock Skew Compensation using Tunable Delay Buffers," *IEEE Transactions on VLSI Systems*, vol. 16, no. 6, pp. 639–649, June 2008.
- [2] P. A. Beerel *et al.*, *A Designer's Guide to Asynchronous VLSI*. Cambridge University Press, 2010.
- [3] N. Mahapatra *et al.*, "An Empirical and Analytical Comparison of Delay Elements and a New Delay Element Design," in *IEEE Computer Society Workshop on VLSI*, 2000, pp. 81–86.
- [4] C. Vezirtzis, Y. Tsividis, and S. Nowick, "Designing Pipelined Delay Lines with Dynamically-Adaptive Granularity for Low-Energy Applications," in *Computer Design (ICCD), 2012 IEEE 30th International Conference on*, Sept 2012, pp. 329–336.
- [5] A. Ghiribaldi *et al.*, "A Transition-Signaling Bundled Data NoC Switch Architecture for Cost-effective GALS Multicore Systems," in *Design, Automation & Test in Europe (DATE)*, 2013, pp. 332–337.
- [6] J. Hamon and E. Beigne, "Automatic Leakage Control for Wide Range Performance QDI Asynchronous Circuits in FD-SOI Technology," in *International Symposium on Asynchronous Circuits and Systems (ASYNC)*, 2013, pp. 142–149.
- [7] I. Chang *et al.*, "Exploring Asynchronous Design Techniques for Process-tolerant and Energy-efficient Subthreshold Operation," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 2, pp. 401–410, 2010.
- [8] J. Tschanz *et al.*, "Tunable Replica Circuits and Adaptive Voltage-Frequency Techniques for Dynamic Voltage, Temperature, and Aging Variation Tolerance," in *Symposium on VLSI Circuits (VLSI)*, 2009, pp. 112–113.
- [9] M. Maymandi-Nejad and M. Sachdev, "A Digitally Programmable Delay Element: Design and Analysis," *IEEE Transactions on VLSI Systems*, vol. 11, no. 5, pp. 871–878, Oct 2003.
- [10] —, "A Monotonic Digitally Controlled Delay Element," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 11, pp. 2212–2219, Nov 2005.
- [11] G. Kim *et al.*, "A Low-voltage, Low-power CMOS Delay Element," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 966–971, July 1996.
- [12] N. Mahapatra *et al.*, "Comparison and Analysis of Delay Elements," in *45th Midwest Symposium on Circuits and Systems (MWSCAS)*, 2002, pp. 473–476.
- [13] S. Kobenge and H. Yang, "A Power Efficient Digitally Programmable Delay Element for Low Power VLSI Applications," in *Asia Symposium on Quality Electronic Design*, 2009, pp. 1–5.
- [14] A. Hiroki *et al.*, "An Analytical MOSFET Model Including Gate Voltage Dependence of Channel Length Modulation Parameter for 20 nm CMOS," in *International Conference on Electrical and Computer Engineering (ICECE)*, 2008, pp. 139–143.
- [15] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*. Addison Wesley Publishing Company Incorporated, 2011.
- [16] J. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*, ser. Prentice Hall Electronics and VLSI Series. Pearson Education, 2003.
- [17] T. Tsunomura *et al.*, "Effect of Channel Dopant Profile on Difference in Threshold Voltage Variability Between NFETs and PFETs," *IEEE Transactions on Electron Devices*, vol. 58, no. 2, pp. 364–369, Feb 2011.
- [18] C. Chen and M. Sarrafzadeh, "Simultaneous Voltage Scaling and Gate Sizing for Low-Power Design," *IEEE Transactions on Circuits and Systems II*, vol. 49, no. 6, pp. 400–408, Jun 2002.
- [19] A. Chandrakasan *et al.*, "Low-Power CMOS Digital Design," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 4, pp. 473–484, April 1992.