Body-bias-driven design strategy for area and performance efficient CMOS circuits

Citation for published version (APA):

DOI:
10.1109/TVLSI.2010.2091974

Document status and date:
Published: 01/01/2012

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Body-Bias-Driven Design Strategy for Area- and Performance-Efficient CMOS Circuits

Maurice Meijer and José Pineda de Gyvez, Fellow, IEEE

Abstract—Worst-case design uses extreme process corner conditions which rarely occur. This limits maximum speed specifications and costs additional power due to area over-dimensioning during synthesis. We present a new design synthesis strategy for digital CMOS circuits that makes use of forward body biasing. Our approach renders consistently a better performance-per-area ratio by constraining circuit over-dimensioning without sacrificing circuit performance. An in-depth analysis of the body-bias-driven design theory is provided. It is complemented by an algorithm that enables fast reconstruction of the area-clock period tradeoff curve of the design. We validated these new concepts through industrial processor designs in 90-nm low-power CMOS. For standard-$V_{th}$ implementations, we observed performance-per-area improvements up to 40%, area and leakage reductions up to 30%, and dynamic power savings of up to 10% without performance penalties as a benefit from our proposed body-bias-driven design strategy. The benefits are larger for high-$V_{th}$ implementations. In this case, we observed performance-per-area improvements up to 90%, area and leakage reductions up to 40%, and dynamic power savings of up to 25% without performance penalties.

Index Terms—Circuit optimization, circuit tuning, CMOS digital integrated circuits, logic design.

I. INTRODUCTION

CONVENTIONAL and well-established digital design practices are based on a worst-case design (WCD) style to guarantee chip operation for meeting timing specifications among the process corners [1]. The circuit is designed in the slow-process corner to meet frequency specifications, while the maximum leakage target is verified in the fast-process corner. However, such extreme process corners rarely occur in most of the fabricated chips. Moreover, WCD makes high performance specifications harder to meet due to over-dimensioning of the design. Over-dimensioning leads to a larger silicon footprint, higher power consumption, and larger leakage. Fig. 1 shows the area–delay tradeoff involved during logic synthesis. Observe that circuit area depends on the process margin for high-performance circuits. If a lower process margin can be tolerated without a parametric yield penalty, circuit performance can be further increased without spending excessive area. Statistical circuit design has long been seen as a viable way to avoid the use of worst-case parameters [2], [3]. However, these approaches have not totally found their way into industrial practices. This is due to, among other reasons, the moving average of process parameters, the flexibility of fabricating the same chip in multiple foundries, and the lack of appropriate EDA tools for statistical logic synthesis. In this paper, we show that a body-bias-driven (BBD) logic synthesis overcomes these drawbacks.

Alternatively, post-silicon tuning has been proposed for improving product-binning yields and for trading off power performance [4]–[7], but does not eliminate the problem of area over-dimensioning. Body biasing is typically used for leakage reduction or performance tuning [4]–[8]. Forward body biasing (FBB) is preferred over supply voltage scaling (VS) to achieve increased performance [4], [7], [8]. This is because the power penalty of FBB is lower for dynamic-power dominant designs. Reverse body biasing (RBB) can effectively achieve leakage reductions [6]. Other post silicon tuning works have been reported. For instance, a joint design-time and post-silicon tuning optimization strategy for minimizing leakage under delay constraints was proposed in [9]. This approach relies on detailed process variability inputs and is capable of reducing process-dependent delay spread. However, it consider neither a timing speed-up nor a circuit area reduction as outcome. Others propose body bias clustering at design-time for minimizing leakage under delay constraints [10], [11], or enhancing circuit performance [12]. These approaches do not consider a (joint) design-time optimization for improving performance or reducing area of the circuit.

Threshold voltage ($V_{th}$) assignment and gate sizing during the design synthesis phase is a known problem [13], [14]. $V_{th}$ assignment has been used for reducing leakage or for reducing dynamic power consumption. Leakage power of digital IP blocks is mostly a concern when the circuit is in standby mode. High-performance circuits typically use low-$V_{th}$ (LVT) devices to speed-up critical delay paths at a higher intrinsic device leakage penalty [13]. This higher leakage is unacceptable for portable
applications since it increases standby power, and thereby reduces battery time. The use of body biasing offers several advantages: 1) it offers a continuum of \( V_{th} \) values; 2) it is \( V_{th} \) technology flavor independent; 3) it can be used on top of multi-\( V_{th} \) assignments; and 4) it can be used applied dynamically or adaptively. FBB can achieve LVT performance during operation, while it can be turned off to achieve low leakage in standby [7]. LVT circuits with RBB cannot achieve such low leakage [15]. Like multi-\( V_{th} \) and gate sizing assignments, BBD designs render smaller footprint area than WCD. Unlike multi-\( V_{th} \) and gate sizing assignments, the \( V_{th} \) choice is not technology-constrained since it is possible to characterize a standard cell library with FBB targeting a given \( V_{th} \) value within a certain range of \( V_{th} \)’s. In [16], we presented a body-bias-driven gate-level optimization method that leverages FBB to improve the performance-per-area (PPA) ratio of digital CMOS circuits. In this paper, we provide an in-depth analysis of the body-bias-driven design theory. This theory allows us to predict the design’s optimum PPA with a minimum number of synthesis trials. We validated these new concepts through industrial processor designs in 90-nm LP-CMOS. In this paper, we discuss as well how our approach is fully integrated in a state-of-the-art commercial design flow.

The remainder of this paper is organized as follows. In Section II, we introduce BBD design. Section III presents the theoretical background and modeling. In Section IV, we explore the area, performance, and power trends for BBD design. Section V presents the BBD logic synthesis approach. In Section VI, we validate the proposed models. Section VII shows our benchmarked results. Finally, Section VIII presents our conclusions.

II. BBD DIGITAL DESIGN

Here, we will introduce BBD design and present body-bias silicon-tuning capabilities for a 90-nm low-power CMOS process technology.

A. Concept

Under WCD, digital CMOS circuits are implemented to meet timing specifications for slow process conditions. Observe, however, that FBB enhances circuit speed. Bearing this in mind, one does not need to pursue WCD. Instead, it is possible to design the circuit in between the worst and nominal process corners provided that the IC has FBB capabilities to correct performance deviations due to fabrication outcome. This creates opportunities for more cost-effective solutions without sacrificing performance specs and parametric yield. The amount of FBB required can be calibrated at test time or during boot of the chip.

Fig. 2 illustrates the parameters that are under control with BBD design. The right-hand side of Fig. 2 plots the dependency between clock period and FBB. A higher FBB value enables faster circuit operation. The amount of speed-up depends on the process technology, the used transistor threshold voltage option, and the design’s power supply voltage. FBB needs only to be applied to those die samples with a lower speed than the nominal process outcome. The left-hand side of Fig. 2 plots the relationship between circuit area and clock period. For increasing FBB values, the curve shifts linearly proportional to a reducing clock period. Notice that a performance increase by FBB can be traded off against a performance decrease due to a smaller circuit area. In this way, we are able to maximize the PPA ratio of the circuit at design-time, while meeting a target performance.

B. Body-Bias Tuning Capabilities in CMOS 90-nm Process

The effectiveness of BBD design depends on the performance tuning range available with FBB. We briefly summarize our experimental results that were obtained for a set of ring-oscillator test structures in a 90-nm LP CMOS process. These test structures are similar to the ones presented in [4]. Both standard-\( V_{th} \) (SVT) and high-\( V_{th} \) (HVT) versions are available. Measurements have been performed for 61 die samples of the same 300-mm wafer at \( V_{DD} = 1.1 \) V and \( T = 85 \) °C. In our experiments, we applied FBB of, at most, 0.5 V to avoid turning on the device’s junction diodes. FBB is applied simultaneously to pMOS and nMOS transistors through P- and N-well biasing, respectively.

Table I presents the measurement results for a slow die sample. A 24% and 40% performance increase is observed for the SVT and HVT ring-oscillator test structures, respectively, when 0.5-V FBB is applied to both N- and P-wells simultaneously. Contrarily, leakage increases by up to about 25x and 80x, respectively. The leakage increase is more severe for HVT. This is because the forward-biased junction leakage at 0.5-V FBB dominates over the subthreshold leakage.

LVT circuits show a lower performance and leakage increase with FBB as compared with SVT [15]. The intrinsic leakage of LVT is about 10x higher than in case of SVT, which has a large impact on power consumption in standby or low-activity
use-cases. Therefore, we will focus on the use of SVT and HVT in the remainder of this work.

III. THEORETICAL BACKGROUND BBD DESIGN

Here, we present the theoretical background of BBD design for achieving an optimum PPA ratio.

A. Area and Clock Period Modeling

The delay of a digital logic gate can be modeled as [13]

\[
d_{\text{gate}} = \left( d_0 + \frac{d_1}{x} \right) (1 + kV_{\text{BB}}),
\]

(1)

The first term represents the intrinsic \(d_0\) and load-dependent \(d_1/x\) gate delay [17]. Parameter \(x\) is the gate sizing factor \((x \geq 1)\). The gate delay is both \(V_{\text{DD}}\) and \(V_{\text{th}}\)-dependent. The second term models the impact of FBB on gate delay by a linear function. \(V_{\text{BB}}\) represents the FBB voltage: \(V_{\text{BB}} = V_{\text{FBB}} = V_{\text{DD}} - V_{\text{th}}\). Parameter \(k\) is the polynomial coefficient, which can be different for each gate. The maximum error of using such a linear function to model the delay dependency on FBB is lower than 2% for our 90-nm LP-CMOS test-structures.

Based on (1), we model the delay and area of a CMOS digital logic circuit as

\[
D_j = \sum_{i \in j} \left( d_{0i} + \frac{d_{1i}}{x_i} \right) (1 + k_i V_{\text{BB}}) \leq T_{\text{ck}}, \quad \forall j \in \Psi
\]

(2)

\[
A_{\text{total}} = \sum_{i=1}^{m} x_i A_i
\]

(3)

where \(i\) is an index that runs over all gates in the circuit, \(j\) is an index that runs over all paths in the circuit, \(D_j\) is the delay of path \(j\), \(\Psi\) is the collection of all paths in the circuit, and \(A_i\) is the minimum area of gate \(i\). Expression (2) constrains the delay of each circuit path to be less than the targeted clock period \(T_{\text{ck}}\). The total circuit area is the summed gate areas.

Fig. 3 shows a typical area-clock period tradeoff curve for a generic digital logic circuit. The curve is constructed from a multitude of synthesis runs such that the same design meets distinct clock period constraints. In Fig. 3, the area and clock period have been normalized to the best performing design \((A_{\text{max}}, T_{\text{min}})\). This design is obtained by constraining gate sizing of digital gates to their maximum size in the library. The faster designs \((T_{\text{ck}} < T_{\text{min}})\) are obtained for unconstrained gate sizing. Observe that high-performance circuits consume more area than slow circuits. This is due to gate upsizing and logic reordering to speed up critical circuit paths. The trend shown in Fig. 3 can be modeled by a rational function, given as follows, with \(\chi\), \(\delta\), and \(\eta\) as independent fitting parameters:

\[
A_{\text{total}} = \frac{x}{\delta + T_{\text{ck}}} + \eta
\]

(4)

The general form of (4) describes a rectangular hyperbola. Parameters \(\delta\) and \(\eta\) model the shift of origin. The vertical asymptote is located at \(T_{\text{ck}} = 1/\delta\), which represents the minimum clock period of the design that is theoretically possible. The horizontal asymptote is located at \(A_{\text{total}} = \eta\), which represents the minimum area of the design in case of no gate upsizing or logic restructuring. Parameter \(\chi\) models the hyperbola scale factor, which accounts for the impact of gate upsizing and logic restructuring. A characteristic point is the clock period value \(T_A\) at which the slope of the hyperbola equals \(-1\). This clock period \(T_A\) will be used to reconstruct the design’s area and clock period tradeoff curve, as will be discussed in Section V-B. \(T_A\) and the corresponding circuit area \(A_A\) can be determined as follows:

\[
T_A = \frac{\chi}{\delta} + \sqrt{\chi} \quad \forall \frac{T_{\text{ck}}}{\min} \geq T_{\text{min}} \land \chi \geq 0
\]

(5)

\[
A_A = \eta + \sqrt{\chi} \quad \forall T_{\text{ck}} \geq T_{\text{min}} \land \chi \geq 0.
\]

(6)

Next, we will discuss the relationship between the fitting parameters of (4) for WCD and BBD design styles. Parameter \(\eta\) is identical for both design styles because of the very relaxed or unconstrained timing. Now, notice that, if both WCD and BBD circuits were optimized in the same way over the entire clock period range, then \(\chi\) would be the same. However, this is not true in general since BBD libraries are faster than conventional libraries, e.g., the gate drive of a forward bias cell is larger than the one of the same cell without FBB.

Let us now take a look at speed and area tradeoffs between WCD and BBD design styles. Suppose first that a given circuit area \((A_{\text{total,lbdd}} = \lambda \cdot A_{\text{total,wcld}})\) is desired, then the clock period of the BBD circuit can be obtained from the WCD clock period

\[
T_{\text{ck,lbdd}} = T_{\text{ck,wcld}} + \frac{\chi_{\text{lbdd}}(\delta_{\text{wcld}} + T_{\text{ck,wcld}})}{\chi_{\text{wcld}} + (\delta_{\text{wcld}} - \eta_{\text{lbdd}})(\delta_{\text{wcld}} + T_{\text{ck,wcld}})}
\]

(7)

where \(\lambda\) represents the fraction \(A_{\text{total,lbdd}}/A_{\text{total,wcld}}\). Parameter \(\lambda\) equals 1 for a constant circuit area between both design styles. Alternatively, suppose now that a given clock period \((T_{\text{ck,lbdd}} = \sigma \cdot T_{\text{ck,wcld}})\) is pursued, then the circuit area of the BBD circuit can be obtained from the WCD circuit area as follows:

\[
A_{\text{total,lbdd}} = \eta_{\text{lbdd}} + \frac{\chi_{\text{lbdd}}(A_{\text{total,wcld}} - \eta_{\text{lbdd}})}{(\delta_{\text{lbdd}} - \sigma \delta_{\text{wcld}})(A_{\text{total,wcld}} - \eta_{\text{lbdd}}) + \sigma \chi_{\text{wcld}}}
\]

(8)

where \(\sigma\) represents the fraction \(T_{\text{ck,lbdd}}/T_{\text{ck,wcld}}\) and equals 1 for a constant clock period. Notice from (7) that the speed advantage of the BBD circuit depends only on the difference between \(\delta\)'s provided that \(\chi_{\text{lbdd}} = \chi_{\text{wcld}}\). The smaller area of BBD...
in (8) is also due to the difference of \(\delta\)'s. These results are expected since digital gates with FBB have a greater output drive than without FBB. Consequently, smaller area gates are employed in BBD designs. Equations (7) and (8) enable designers to estimate the effectiveness of BBD over WCD in trading off circuit speed against area. Design and process technology alternatives can be compared once the parameter values for \(\chi\), \(\delta\), and \(\eta\) are known. These parameters are design-dependent because of different amount and type of digital cells used as well as the logic implementation. Moreover, they are also process-technology-dependent because circuit area, performance, and body-bias sensitivity depend on technology scaling. For example, a given digital logic circuit will be smaller (lower \(\eta\), different \(\chi\)) and faster (lower |\(\delta|\), different \(\chi\)) when implemented in a next-generation CMOS technology.

B. PPA Figure-of-Merit (FOM)

Circuit performance and area are key performance metrics for digital circuit designers. We introduce a new metric (PPA) to qualify how effectively the design achieves high performance while accounting for area scaling. The PPA metric depends on the technology node, the technology’s threshold voltage option, and the standard cells available for circuit synthesis. Let \(f_{ck} = 1/T_{ck} = 1/\max(D_j)\). Then we obtain

\[
\text{PPA} = \frac{f_{ck}}{A_{\text{total}}} = \frac{1}{T_{ck} A_{\text{total}}}. \quad (9)
\]

A higher PPA value indicates that the circuit design utilizes silicon area more effectively to achieve a high performance. There exists a point in (4) with a maximum PPA. This point indicates the optimum performance without circuit over-dimensioning. By combining (4) and (9), we obtain

\[
\text{PPA} = \frac{\delta + T_{ck}}{(\chi + \eta T_{ck}) T_{ck}}. \quad (10)
\]

The clock period value at which the maximum PPA occurs \(T_{\text{best}}\), can be determined by making the derivative of PPA with respect to \(T_{ck}\) equal to zero. By solving the equation for \(T_{ck}\), we obtain a closed-form expression for \(T_{\text{best}}\), namely

\[
T_{\text{best}} = -\delta + \frac{\sqrt{\delta^2 - 4\chi \eta}}{\eta} \quad \forall T_{ck} \geq T_{\text{min}} \land \delta \chi \eta \leq 0, \quad (11)
\]

\(T_{ck} > T_{\text{best}}\) yields circuits without area over-dimensioning, and the contrary holds true for \(T_{ck} < T_{\text{best}}\). Therefore, \(T_{\text{best}}\) identifies the minimum possible clock period without circuit over-dimensioning. The maximum PPA at \(T_{ck} = T_{\text{best}}\) is obtained after substituting (11) into (10) as follows:

\[
\max(\text{PPA}) = \frac{1}{\chi - \delta \eta + 2\sqrt{\delta^2 - 4\chi \eta}} \quad \forall T_{ck} \geq T_{\text{min}} \land \delta \chi \eta \leq 0, \quad (12)
\]

Under WCD, \(T_{\text{best}}\) may be too large to meet the target frequency specification of high-performance designs. In this case, over-dimensioning cannot be avoided, thereby worsening PPA.

In the forthcoming analysis, we make use of a normalized representation of PPA. The normalization is against the highest performance under WCD \((f_{ck} = f_{\text{max}} = 1/T_{\text{min}}, A_{\text{total}} = A_{\text{max}})\):

\[
\text{PPA}_{\text{norm}} = \frac{T_{\text{min}} A_{\text{max}}}{T_{ck} A_{\text{total}}}, \quad (13)
\]

C. Power Modeling

Power consumption of a digital gate can be modeled as

\[
P_{\text{gate}} = a(x C_{\text{intr}} + C_{\text{load}}) V_{DD}^2 f_{ck} + x I_{\text{leak}} V_{DD} \quad (14)
\]

where \(a\) is the switching activity of the gate, \(C_{\text{intr}}\) and \(C_{\text{load}}\) are the intrinsic and load capacitance of a gate, respectively, and \(f_{ck}\) is the operating frequency. \(I_{\text{leak}}\) is the leakage current of a gate, which depends both \(V_{DD}\) and \(V_{th}\). From experimental results, we model the normalized leakage current dependence on body biasing by a fourth-order polynomial expression

\[
\text{leakage}_{\text{norm}} = 1 + \sum_{n=1}^{4} I_{\text{leak}}(V_{BB}) \quad (15)
\]

The leakage at various FBB conditions has been normalized to the case of nominal body bias. As before, \(V_{BB}\) represents the FBB value: \(V_{BB} = V_{\text{pred}} = V_{DD} - V_{\text{th}}\). Parameters \(I\) are the polynomial coefficients, which are different for each gate. The maximum error of expression (15) is lower than 2% for our 90-nm LP-CMOS test-structures.

The intrinsic (or junction) capacitance of a gate is dependent on the applied body bias [8]. In our experiments, we have extracted the junction capacitance values from the dynamic power consumption measurement results. Hence, we model the normalized junction capacitance by a second-order polynomial expression. As before, the normalization has been done against the nominal body bias case:

\[
C_{\text{intr,norm}} = 1 + \sum_{n=1}^{2} m_{n} V_{BB}^{n} \quad (16)
\]

Parameters \(m\) are the polynomial coefficients, which are different for each gate. The maximum error of expression (16) is lower than 1.5% for our 90-nm LP-CMOS test-structures when used to model the body-bias impact on dynamic power.

By combining (14), (15), and (16), we model the total power consumption of a generic CMOS digital logic circuit as

\[
P_{\text{total}} = \sum_{i=1}^{m} a_{i} x_{i} C_{\text{intr,i}} \left(1 + \sum_{n=1}^{2} m_{n,i} V_{BB}^{n} \right) + C_{\text{load,i}} \times V_{DD}^2 f_{ck} + \sum_{i=1}^{m} x_{i} V_{DD} I_{\text{leak,i}} \left(1 + \sum_{n=1}^{4} I_{\text{leak}}(V_{BB}) \right) \quad (17)
\]

where \(i\) is an index that runs over all gates in the circuit. Observe that we are assuming that WCD and BBD circuits use the same power supply voltage and operate at the same temperature. WCD and BBD circuits have different power consumption depending on differences in circuit dimensions, circuit activity, and operating frequency. Moreover, BBD circuits utilize FBB that increases power.

IV. OPTIMUM PPA DESIGN SPACE

Here, we explore area, performance, and power trends for WCD and BBD design styles by using the previously presented models. For this purpose, we take a generic digital logic circuit with calibrated technology parameters for 90-nm LP-CMOS. The analysis was done at \(V_{DD} = 1.1\) V and \(T = 85\) °C. For BBD design, we utilized a maximum FBB of 0.5 V to explore the limits of PPA driven design. All results relate to the slow-
process corner. Finally, we discuss technology-scaling implications by analyzing the same circuit in 65- and 45-nm LP-CMOS. The same process and operating conditions have been used as before with the exception of $V_{DD} = 1$ V for the 45-nm case.

A. PPA Trends

Fig. 4 shows the design exploration space for circuit area, clock period, and PPA. The area–clock period trend curves are plotted for WCD (solid line) and BBD (dashed line) design. The iso-PPA curves are plotted as overlay. The intersection with the area–clock period curves represents the normalized PPA ratio of the design as defined by (13). Logic synthesis usually aims at achieving a given target speed. As way of example, all PPA values of Fig. 4 have been normalized to the maximum frequency of operation under WCD ($T_{ch} = T_{min}$). This reference point is highlighted by the triangle symbol in Fig. 4. The triangle is located at a clock period of $T_{min}$, while the circles relate to $T_{test}$ which are the corresponding best PPA points. Observe from Fig. 4 that, for a given circuit area, BBD design achieves higher performance than WCD counterparts. Alternatively, BBD design enables lower area designs for a given clock period. Any FBB of less than 0.5 V results in area–clock period curves located in between the two curves plotted in Fig. 4. Therefore, it makes most sense to use BBD design with the maximum possible FBB to obtain the best PPA ratio.

Fig. 5 highlights the PPA and clock period trends under WCD and BBD design. Notice that BBD design achieves a better PPA ratio than WCD under all circumstances. From large clock periods towards smaller ones, the PPA increases to a maximum value irrespective of the chosen design style. The increasing PPA is because the decrease in clock period is greater than the increase in circuit area. This trend is reversed after the maximum PPA has been reached due to area over-dimensioning. Observe from Fig. 5 that the maximum PPA can significantly be higher than the PPA of the maximum frequency under WCD. At large clock periods, the PPA of the BBD and WCD circuits is similar (not shown in Fig. 5).

B. Power Consumption Trends

Fig. 6 shows the design exploration space for circuit area, clock period and power consumption. The trend lines for total power and dynamic power have been indicated under body bias conditions. The total power includes both dynamic and leakage power consumption. The iso-power curves are plotted as overlay, and the solid and dotted lines correspond to the total power and dynamic power, respectively. The intersection of the total power curves with the area–clock period curves represents the power consumed by the design. Observe in Fig. 6 that the power increases for decreasing clock periods due to a larger circuit area, and higher frequency of operation. The dynamic power increases linearly proportional to operating frequency. Recall that the same power supply voltage of 1.1 V is used for both WCD and BBD design. Only when FBB is applied, the leakage power becomes noticeable in the total power consumption. In this case, a difference occurs between the iso-total-power and iso-dynamic-power curves due to FBB under BBD design. This difference becomes larger for larger FBB values. The snapback point of the total power trends...
defines the maximum FBB value to be applied from a power point of view. In our case, this point occurs at an FBB value of 0.5 V. Notice that BBD enables lower power operation at a constant clock period. This is because of the lower circuit area for BBD design. For a given power target, BBD design offers better performance and area figures. However, BBD design consumes more power for the same circuit area as WCD. This is not only because of the higher operating frequency, but also due to the higher junction capacitance and leakage power associated to the application of FBB.

C. Impact of Technology Scaling

Fig. 7 shows the design exploration space for circuit area, clock period and total power for the same generic digital logic circuit in different process technology nodes. Three groups of iso-power curves are plotted as overlay, each representing a given technology node. The symbols represent the maximum PPA designs for which the results are summarized in Table II. All values have been normalized to the maximum PPA design under WCD for 90-nm LP-CMOS. Observe in Fig. 7 the same area–clock period trends for each technology node. BBD design consistently outperforms WCD. The maximum PPA design is faster and smaller in a next-generation technology. Consequently, the PPA increases with technology scaling, as illustrated in Table II. BBD design achieves a similar PPA increase in each technology, because the performance increase with FBB is nearly constant [4], [15]. Also observe in Table II the opposing total power trends under WCD and BBD design. For WCD, the maximum PPA design operates at lower power in a scaled technology, despite the higher clock speed and the increasing leakage. This is no longer the case for BBD design, mainly due to the amplified leakage with FBB which is more pronounced in BBD.

V. BBD DESIGN SYNTHESIS

From the previous sections, we saw that the PPA point indicates the optimum area–delay tradeoff for the circuit under consideration. Yet, it is necessary to construct this area–delay curve to find this point with minimum overhead effort. Here, we discuss the implementation of BBD design using a commercial logic synthesis tool and present an algorithm that enables fast reconstruction of the area–clock period tradeoff curve of the design.

A. BBD Synthesis With Commercial Tools

Commercial synthesis tools can target area optimization subject to delay constraints. To validate our approach, we have implemented BBD synthesis in Cadence RTL Compiler.1 Digital cell libraries have been recharacterized to account for FBB in 90-nm LP-CMOS using Altos’ Liberate library characterizer.2 The library characterization uses the effective current source model (ECSM) for timing, noise, and power modeling. To enable BBD synthesis, FBB-characterized timing views have been created, utilizing 0.5-V FBB for pMOS and nMOS transistors. Both WCD and BBD digital cell libraries have been characterized for slow process conditions, $V_{DD} = 1.1$ V, and $T = 125^\circ$C settings. Such digital cell libraries also enable static timing verification of BBD circuits.

B. Minimizing the Number of Synthesis Runs

Finding the maximum PPA design starts with an iterative process for collecting sufficient data points to reconstruct the area–clock period trade-off curve. Collecting data points evenly spaced across the clock period range requires many synthesis runs, which are time-consuming for large designs. Fortunately, the trade-off curve is a continuous function over a closed clock period interval. A reconstruction is possible with three specific data points only, namely a first point at a small clock period and large circuit area, a second point at a large clock period and small circuit area, and a third point when the slope of the curve is $-1$ at $T_{ck} = T_A$, see (5).

Our approach to curve reconstruction is based on a greedy search algorithm. It makes use of a kind of Newton–Raphson iteration, which is known for its fast convergence [18]. The proposed algorithm searches for the clock period value at which the slope of the area–clock period tradeoff curve equals $-1$ ($T_{ck} = T_A$). Instead of calculating the derivative of the area-clock period explicitly, we made use of (5) to determine $T_{ck} = T_A$. Let us now address our algorithm as described in Fig. 8. As a first step, the design is synthesized at the minimum clock period bound, $T_{clk}$. The synthesis tool returns the actual clock period $T_1$ and circuit area $A_1$. Note that $T_1 > T_{ck}$ when $T_{ck} < T_{\text{min}}$, for other cases $T_1 \approx T_{ck}$. Next, the design is synthesized at the maximum clock period bound, $T_{clk}$. $T_{\text{high}}$ should be chosen large enough to ensure the clock period range at which area

\[\begin{array}{c|c|c|c}
\hline
\text{Technology} & \text{Relative PPA} & \text{Relative Clock Period} & \text{Relative Total Power} \\
\hline
\text{WCD} & \text{90nm LP-CMOS} & 1.45 & 0.85 & 1.07 \\
 & \text{65nm LP-CMOS} & 2.49 & 0.61 & 0.74 \\
 & \text{45nm LP-CMOS} & 3.17 & 0.46 & 1.21 \\
\hline
\text{BBD} & \text{90nm LP-CMOS} & 1.27 & 0.79 & 1.75 \\
 & \text{65nm LP-CMOS} & 1.82 & 0.68 & 1.81 \\
 & \text{45nm LP-CMOS} & 3.17 & 0.46 & 2.12 \\
\hline
\end{array}\]
The synthesis tool returns the actual clock period \( C. BBD \). If this condition is met, the clock period at which for driving digital circuits of 1 mm, is determined at which the slope of and the new \( 3n_s, \) is \( 1.1 \) V, and \( . \) The \( 30 \) ns, and \( . \) Again, the least-squares method place-and-routed in cell rows as shown in [19], while inserting dedicated tap cells have been added to the library. The circuit is prepared for BBD place-and-route support. N-well and P-well are separated from the P-substrate. Since the nMOS devices share the same P-well, the deep N-well isolation is added to (part of) the design for minimizing the overhead of the deep N-well. Only 2 \( \mu m \) extra is needed at each side of the body biased circuit part.

over-dimensioning occurs is captured. We have used a \( T_{\text{high}} \) value of \( 10T_j \). The synthesis tool returns the actual clock period \( T_2 \) and circuit area \( A_2 \). The third synthesis point is chosen based on the bi-section to ensure proper conditioning with three points for curve fitting based on (4). We apply then the least-squares method to determine the fitting parameters \( \chi, \delta, \) and \( \eta \). The clock period \( T_{\text{next}} = T_j \), is determined at which the slope of area–clock period tradeoff curve equals \(-1\) from (5). When the difference between the current clock period \( T_{\text{cur}} \) and the new one \( T_{\text{next}} \) is larger than the tolerated error \( \varepsilon \), a new synthesis run is executed at \( T_{\text{cur}} = T_{\text{next}} \). Again, the least-squares method is used to recalculate the fitting parameters using the available synthesis results. This process is repeated until \( T_{\text{next}} = T_{\text{cur}} \) is smaller than \( \varepsilon \). If this condition is met, the clock period at which the maximum PPA occurs can be calculated by expression (11), as shown before. A final synthesis run is required at \( T_{\text{ck}} = T_{\text{next}} \) to obtain the maximum PPA design.

C. Physical Design Aspects and Body Bias Generation

The backend views of the digital standard cells have been prepared for BBD place-and-route support. N-well and P-well taps have been removed from the digital cell layouts. Instead, dedicated tap cells have been added to the library. The circuit is place-and-routed in cell rows as shown in [19], while inserting tap cells in columns at a maximum pitch of \( 60 \mu m \). A two-layer routing grid for connecting the tap cells to the body bias supplies has been utilized.

The body-biased cells share the same N-well and P-well. Deep N-well isolation is added to (part of) the design for separating the P-well of the body biased nMOS devices from the P-substrate. Since the nMOS devices share the same P-well, the overhead of the deep N-well is minimized. Only 2 \( \mu m \) extra is needed at each side of the body biased circuit part.

Dynamic FBB requires a voltage generator circuit to generate the N-well and P-well bias voltages. A 90-nm LP-CMOS solution has been presented in [20]. This FBB generator occupies 0.03 \( \text{mm}^2 \) for driving digital circuits of 1 mm\(^2\), translating into \( \sim 3\% \) area overhead. The generator’s size increases by 0.01 \( \text{mm}^2 \) for each additional \( \text{mm}^2 \) digital circuit size. For digital circuits smaller than 1 \( \text{mm}^2 \), the size of the generator needs to be adapted to minimize area overhead.

VI. MODEL VALIDATION

WCD and BBD design have been analyzed and compared for a commercial microprocessor design in 90-nm LP-CMOS. The circuit contains 3764 flip-flops and about 31 K combinational gates. It makes use of SVT devices only. This section presents correlated results obtained from logic synthesis and the presented models. As before, the analysis has been performed for slow-process conditions, \( V_{\text{DD}} = 1.1 \text{ V} \), and \( T = 85 \text{ °C} \). BBD design makes use of a maximum FBB of 0.5 V.

A. Design Synthesis Approach

Design synthesis targeted reconstruction of the area–clock period tradeoff curve. We made use of the greedy algorithm as presented in Section VI-B. For each design style, only four synthesis runs were required for reconstruction. The algorithm received the following inputs: \( T_{\text{kw}} = 3 \text{ ns}, T_{\text{high}} = 30 \text{ ns}, \) and \( \varepsilon = 250 \text{ ps} \). Table III summarizes the fitting parameters. \( T_\chi \) and \( T_{\text{next}} \) for WCD and BBD design styles. One more synthesis run was required to obtain the optimum PPA design.

B. Area, Clock Period, and PPA Comparison

Fig. 9 shows the design exploration space for circuit area and clock period for the given microprocessor design. The synthesis results have been indicated by circles and triangles for WCD and BBD design, respectively. The filled symbols correspond to the four synthesis cases based on our algorithm. The open symbols are additional synthesis cases for trend verification purposes only. The solid and dotted lines show the calculated tradeoff curves for WCD and BBD design, respectively, by using (4). The fitting parameters of the model are given in Table III.

![Fig. 9. Area versus clock period for the microprocessor design in 90-nm LP-CMOS. Lines: WCD (solid) and BBD (dashed) model; symbols: synthesis results. The normalized PPA ratio is indicated for each design.](image-url)
Observe from Fig. 9 the close match between the modeled and the synthesized area–clock period trends. The rms error between the calculated curves and the location of each synthesis result is within 1.5%. After completing the fourth synthesis run, we calculated $T_A$ and $T_{\text{test}}$ values which are given in Table III. The PPA value for each synthesis point has been indicated in Fig. 9 normalized to a $T_{\text{min}}$ of 5.5 ns under WCD. Observe the existence of a maximum PPA design for both WCD ($\text{PPA} = 1.01 \pm 0.01$ ns) and BBD design ($\text{PPA} = 1.27 \pm 0.7$ ns). The calculated $T_{\text{test}}$ values match within 5% of the values obtained through synthesis (WCD: 6 ns; BBD: 4.9 ns). As expected, BBD design not only gives a better performance but also better area utilization as indicated by the PPA value.

C. Power Consumption Comparison

Fig. 10 presents the same area and clock period curve as before, but now the symbols indicate the normalized power consumption of each synthesis run. For the given microprocessor design, BBD design provides lower power operation than WCD at the same clock period. Contrarily, BBD design consumes more power at the same circuit area.

VII. BENCHMARKED RESULTS

Here, we present BBD and WCD results for three industrial processor designs in 90-nm LP-CMOS. Logic synthesis, physical implementation and power analysis has been done using Cadence’s RTL Compiler, First Encounter, and Encounter Timing System, respectively. All results have been obtained for a slow process corner, $V_{\text{DD}} = 1.1$ V, and $T = 85$ °C. BBD design utilizes a maximum FBB of 0.5 V. All area results account for layout effects including the overhead for deep-N-well isolation.

Each processor design has been implemented in both SVT and HVT flavors. Table IV shows the gate count summary. Two synthesis cases have been investigated, namely: 1) a maximum PPA design and 2) a maximum frequency design under WCD. In the latter case, BBD design is utilized to operate at the same speed at a lower area cost to improve the PPA ratio.

A. Design Synthesis for Maximum PPA

Table V presents the processor design results targeting a maximum PPA design. Five circuit parameters are presented, namely clock period, circuit area, PPA, and dynamic and leakage power. The BBD design results are presented relative to the WCD results. The PPA ratio has been normalized to the maximum performance under WCD ($T_{\text{clk}} = T_{\text{min}}$).

Let us first consider SVT results. Observe that the PPA ratio differs for each design. This depends on circuit characteristics such as circuit size, path delay distribution, and logic depth. Under WCD, the PPA ratio ranges from 1.01 to 1.10. The maximum PPA point for small circuits (low $\eta$ value) tends to be located at larger clock period values ($T_{\text{test}} > T_{\text{min}}$), as can be inferred from (11). This explains the high PPA value of 1.10 for the digital signal processor. For large circuits (high $\eta$ value), the maximum PPA value is located closer to, or equal to the minimum clock period, $T_{\text{test}} \approx T_{\text{min}}$. The path delay distribution of the multimedia processor is the reason for the better PPA value w.r.t. the microprocessor design (1.03 i.s.o 1.01). The multimedia processor has many (nearly) critical delay paths which are largely responsible for the area over-dimensioning of the design when requiring high performance. BBD design enables significant improvements in maximum PPA as compared with WCD, mainly due to higher clock speeds. The maximum PPA of the BBD designs ranges between 1.25–1.38. Let us look now into HVT results. The same PPA trends are observed as in the SVT case, but the increase in maximum PPA is much larger (maximum PPA: 1.52–1.90). This is because FBB has a larger impact on circuit speed for HVT. Worth noticing is that the HVT BBD processors can operate at the same speed of the SVT WCD equivalents. However, their PPA values are slightly lower due to a higher circuit area. Irrespective of the $V_{\text{th}}$ option used, BBD design provides always a higher maximum PPA ratio than WCD. All BBD circuits operate faster than their WCD counterparts, while circuit area is comparable.

Table V also shows the dynamic power and leakage power consumption for each processor design. Notice that dynamic power dominates leakage power, even at a high operating temperature of 85°C and when FBB is applied. The ratio between dynamic and leakage power is in the range of 100–300 for SVT WCD (800–2100 for HVT WCD) for the considered processor designs. Under BBD design, this ratio is reduced to 10–30 for SVT BBD design, and 5–20 for HVT BBD design. Observe that the dynamic power for BBD is generally higher than under WCD. There are two reasons for this, namely: 1) the higher clock speed and and 2) the higher junction capacitance due to FBB. Next, the BBD leakage power is significantly higher than the WCD leakage when FBB is utilized. FBB turns on the transistor’s junction diodes, which leads to a high additional leakage current, especially at higher temperature operation. This will be

---

**Example Gate Count—Three Industrial Processor Designs**

<table>
<thead>
<tr>
<th>Processor Design</th>
<th>#flip-flops</th>
<th>#logic gates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital signal processor</td>
<td>227</td>
<td>4416</td>
</tr>
<tr>
<td>Microprocessor</td>
<td>3764</td>
<td>34390</td>
</tr>
<tr>
<td>Multimedia processor</td>
<td>41749</td>
<td>252759</td>
</tr>
</tbody>
</table>
TABLE V  
DESIGN SYNTHESIS FOR MAXIMUM PPA—INDUSTRIAL PROCESSOR DESIGNS IN 90-nm LP-CMOS RELATIVE VALUES ARE SHOWN W.R.T. WCD FOR THE GIVEN V_TH OPTION. CONDITIONS: SLOW-PROCESS CORNER, V_DD = 1.1 V AND T = 85 °C

<table>
<thead>
<tr>
<th>Design Unit</th>
<th>Clock period</th>
<th>Area</th>
<th>PPA</th>
<th>Dynamic Power</th>
<th>Leakage Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WCD</td>
<td>BBD</td>
<td>WCD</td>
<td>WCD, BBD</td>
<td>WCD, BBD</td>
</tr>
<tr>
<td></td>
<td>[ns]</td>
<td>[ns]</td>
<td>[µm²]</td>
<td>rel.</td>
<td>[µW]</td>
</tr>
<tr>
<td>Digital signal processor</td>
<td>6.5</td>
<td>6.0</td>
<td>32656</td>
<td>1.04</td>
<td>1.10</td>
</tr>
<tr>
<td>Microprocessor</td>
<td>6.0</td>
<td>4.9</td>
<td>342705</td>
<td>0.97</td>
<td>1.01</td>
</tr>
<tr>
<td>Multimedia processor</td>
<td>3.8</td>
<td>3.0</td>
<td>3047184</td>
<td>0.95</td>
<td>1.03</td>
</tr>
<tr>
<td>Digital signal processor</td>
<td>10.5</td>
<td>6.5</td>
<td>34386</td>
<td>1.01</td>
<td>1.19</td>
</tr>
<tr>
<td>Microprocessor</td>
<td>8.5</td>
<td>6.0</td>
<td>344440</td>
<td>0.93</td>
<td>1.15</td>
</tr>
<tr>
<td>Multimedia processor</td>
<td>5.8</td>
<td>3.9</td>
<td>3447298</td>
<td>0.89</td>
<td>1.02</td>
</tr>
</tbody>
</table>

TABLE VI  
DESIGN SYNTHESIS FOR OPTIMUM AREA—INDUSTRIAL PROCESSOR DESIGNS IN 90-nm LP-CMOS RELATIVE VALUES ARE SHOWN W.R.T. WCD FOR THE GIVEN V_TH OPTION. CONDITIONS: SLOW-PROCESS CORNER, V_DD = 1.1 V AND T = 85 °C

<table>
<thead>
<tr>
<th>Design Unit</th>
<th>Clock period</th>
<th>Area</th>
<th>PPA</th>
<th>Dynamic Power</th>
<th>Leakage Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WCD, BBD</td>
<td>WCD</td>
<td>BBD</td>
<td>WCD, BBD</td>
<td>WCD, BBD</td>
</tr>
<tr>
<td></td>
<td>[ns]</td>
<td>[µm²]</td>
<td></td>
<td>[µW]</td>
<td>[µW]</td>
</tr>
<tr>
<td>Digital signal processor</td>
<td>5.5</td>
<td>42457</td>
<td>0.74</td>
<td>1</td>
<td>1.35</td>
</tr>
<tr>
<td>Microprocessor</td>
<td>5.5</td>
<td>370932</td>
<td>0.85</td>
<td>1</td>
<td>1.18</td>
</tr>
<tr>
<td>Multimedia processor</td>
<td>3.6</td>
<td>3233781</td>
<td>0.89</td>
<td>1</td>
<td>1.17</td>
</tr>
<tr>
<td>Digital signal processor</td>
<td>8.8</td>
<td>48621</td>
<td>0.60</td>
<td>1</td>
<td>1.72</td>
</tr>
<tr>
<td>Microprocessor</td>
<td>8.5</td>
<td>374440</td>
<td>0.80</td>
<td>1</td>
<td>1.50</td>
</tr>
<tr>
<td>Multimedia processor</td>
<td>5.7</td>
<td>2857787</td>
<td>0.84</td>
<td>1</td>
<td>1.36</td>
</tr>
</tbody>
</table>

also reflected in the total power which is the sum of dynamic and leakage power components. However, recall that FBB is only applied to those chip samples with a lower frequency than the targeted one due to the process outcome. Such slow samples have already an intrinsic low leakage current. Slow-process-corner samples receive the maximum FBB, while the other slow samples receive a lower FBB. When no FBB is applied, the BBD leakage power is proportional to the circuit area scaling (not shown in Table V). In addition, recall that we apply dynamic FBB during chip operation. In this way we avoid the leakage penalty associated to FBB during standby operation.

B. Design Synthesis for Optimum Area

Table VI presents the processor design results targeting a maximum WCD performance. The BBD circuits were designed to match the WCD performance. In this case, BBD circuits can enable significant area savings, irrespective of the V_TH option used. For the SVT BBD designs, we observed area reductions between 11% and 26% w.r.t. their WCD versions. The benefits for the HVT processors are larger due to the stronger FBB dependence (14%–40% area savings). The reduced circuit area comes mostly from the area scaling of the combinatorial logic. In general, BBD circuits have fewer logic gates than WCD ones, while the amount of flip-flops is the same. The largest area savings has been obtained for the digital signal processor, which has about 19 × more logic gates than flip-flops. The ratio between logic gates and flip-flops is lower for the other circuits, as can be derived from Table IV. The lowest ratio is found for the multimedia processor, namely about 6× more logic gates than flip-flops. This explains the area scaling trends observed. The PPA for the BBD processors is not optimal, because the BBD operating frequency is not fully utilized. However, it is significantly higher than for their WCD equivalents irrespective of the V_TH option used.

BBD design renders consistently lower dynamic power than WCD does when operating at the same maximum WCD frequency. The power reduction comes from the reduced circuit area despite the increasing junction capacitances with FBB. The dynamic power reduces up to 7% for the SVT processors, while HVT processors achieve up to 25% dynamic power reductions. We noticed that BBD design primarily affects logic gates in the data path; the clock power is not much reduced. Thus, the dynamic power savings are larger for circuits with higher data activities.

As before, the BBD leakage power is much higher than the WCD leakage when FBB is utilized. The leakage power increases up to 20× for the SVT processors, and up to 219× for the HVT ones. Recall that this leakage increase is of no concern since FBB is disabled during standby operation. The leakage power for BBD design without FBB enabled decreases by the same factor as the circuit area (not shown in Table VI). For samples that do not need FBB to achieve performance, a leakage reduction up to 26% and up to 40% is possible in case of SVT and HVT, respectively.
VIII. CONCLUSION

We presented a design synthesis strategy for digital CMOS integrated circuits that makes use of FBB. Our approach renders consistently a better PPA ratio by constraining circuit over-dimensioning without sacrificing circuit performance. An in-depth analysis of the BBD design analysis was provided, which enables designers to predict the design’s optimum performance per area with a minimum number of synthesis runs. We validated these new concepts through industrial processor designs in 90-nm LP-CMOS. For SVT implementations, we observed PPA improvements up to 40%, area and leakage reductions up to 40%, and dynamic power savings of up to 10% without performance penalties. The benefits are larger for HVT implementations. In this case, we observed PPA improvements up to 90%, area and leakage reductions up to 40%, and dynamic power savings of up to 25% without performance penalties as a benefit from our proposed BBD design strategy.

ACKNOWLEDGMENT

The authors would like to thank A. Kumar, Central Research and Development, NXP Semiconductors, Eindhoven, The Netherlands, for his support regarding timing library generation.

REFERENCES