

# Optimal Power Switch Design Methodology for Ultra Dynamic Voltage Scaling with a Limited Number of Power Rails

Yanzhi Wang, Xue Lin, and Massoud Pedram  
University of Southern California  
Los Angeles, CA 90089 USA  
{yanzhiwa, xuelin, pedram}@usc.edu

## ABSTRACT

Many burst-mode applications require high performance for brief time periods between extended sections of low performance operation. Digital circuits supporting such burst-mode applications should work in both the near-threshold regime and the super-threshold regime for brief time periods. This work proposes the structure support of fine-grained ultra dynamic voltage scaling (UDVS) from the traditional strong-inversion region to the near-threshold region, with limitations on the number of power rails. The number, type, and size of the power switches are jointly optimized to minimize the overall energy consumption of the UDVS circuit block, meanwhile satisfying the target delay or frequency requirement at each DVS level. The proposed optimization framework properly accounts for the dynamic energy consumption as well as the leakage energy consumption through all the power switches during both the operation time and stand-by time of the circuit block. Experimental results on 22nm Predictive Technology Model demonstrate the effectiveness of the proposed optimization framework.

## Categories and Subject Descriptors

B.8.2 [Performance and Reliability]: Performance Analysis and Design Aids

## Keywords

Ultra dynamic voltage scaling (UDVS); power switch; near-threshold

## 1. INTRODUCTION

Aggressive voltage scaling from the traditional super-threshold region to the near/sub-threshold region has been shown to be very effective in reducing power consumption in digital circuits [1][2][3]. It is especially beneficial for applications such as wireless sensor processing and RFID tags where performance is not the primary concern. The operating frequency of near/sub-threshold logic is much lower than that of regular strong-inversion circuits ( $V_{DD} > V_{th}$ ) due to small transistor current, which consists mostly of leakage current. Authors of [4][5] derived analytical expressions of the optimal  $V_{DD}$  to minimize energy, i.e., the minimum energy point or MEP, and showed that the MEP for CMOS circuits typically occurs in the near-threshold region.

Many burst-mode applications require high performance for brief time periods between extended sections of low performance operation [6]. Digital circuits supporting such burst-mode applications should work in both the near-threshold region and super-threshold region (for brief time periods.) Therefore, traditional dynamic voltage scaling (DVS) method should be extended to include near-threshold operation, but the overhead of providing the necessary voltages can be large. Adjustable DC-DC converters tend to have limited efficiency over broad output voltage ranges, and they take hundreds of micro-seconds to switch between different  $V_{DD}$  supply levels especially in the near-threshold regime [6]. An alternative implementation approach called local voltage dithering (LVD) uses header power switches to connect circuit blocks to one of the several power supply rails, thereby allowing for faster switching [7][8]. The LVD approach supports application of fine-grained ultra DVS (UDVS) down to the near-threshold regime and to smaller internal circuit blocks. As the required operating frequency changes, each circuit block spends a different fraction of its operating time at different voltage levels. However, the area overhead of LVD can be significant when the number of required virtual- $V_{DD}$  levels becomes relatively large, since a separate power rail is required for each virtual- $V_{DD}$  level.

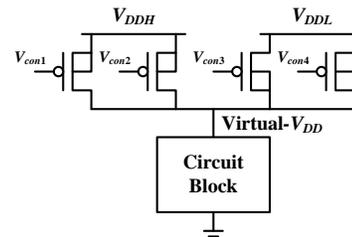


Figure 1. Architecture of fine-grained UDVS to generate six different virtual- $V_{DD}$  levels using two supply power rails.

In this paper, we propose an implementation structure for UDVS with a limited number of power rails. The proposed structure is a generalization of the LVD structure and induces less area overhead than the LVD structure when the number of required virtual- $V_{DD}$  levels is large. We use parallel, independently controllable power switches with different widths connecting between each  $V_{DD}$  supply rail and the circuit block. During circuit operation, we turn on a subset of the parallel-connected power switches and turn off the rest to vary the effective size dynamically, in order to generate an appropriate operating voltage level for the circuit block. The circuit block is therefore not constrained by the available voltage rails. Figure 1 illustrates an example of the proposed structure, which provides up to six different virtual- $V_{DD}$  levels for the circuit block from two  $V_{DD}$  supply rails if the power switches are properly sized.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

GLSVLSI'14, May 21–23, 2014, Houston, Texas, USA.

Copyright © 2014 ACM 978-1-4503-2816-6/14/05...\$15.00.

The introduction of a power switch device between  $V_{DD}$  and the circuit block creates an IR drop across the header, thereby resulting in a reduced virtual- $V_{DD}$  value. Power switch sizing is critical to maintain low power consumption and expected performance. An undersized power switch results in a large performance degradation, whereas an oversized power switch results in increased leakage and increased area overhead. Power switch sizing methodologies have been examined in depth to support techniques such as multi-threshold CMOS (MTCMOS), which uses high- $V_{th}$  power switches to reduce leakage [10][11][12]. In this work, we propose an optimization framework of the UDVS implementation structure. We jointly optimize the supply voltage  $V_{DD}$  levels as well as the number, type (PMOS or NMOS), and size of the power switches. We minimize the overall energy consumption of the UDVS circuit block satisfying the target delay or frequency requirement at each DVS level. We take into account the additional constraints on the number of  $V_{DD}$  power rails and the total area overhead. The proposed optimization framework also properly accounts for the dynamic energy consumption as well as the leakage energy consumption through all the power switches during both the operation time and stand-by time of the circuit block. Experimental results on HSpice simulation of 22nm Predictive Technology Model (PTM) [14] show that the proposed optimization framework achieves up to 19% reduction in energy consumption or 74% reduction in area overhead compared with the baseline method.

The rest of this paper is organized as follows. Section **Error! Reference source not found.** presents the transistor and circuit models operating in the sub/near-threshold regime. In Section **Error! Reference source not found.**, we propose the structure support for UDVS over a wide supply voltage range. Section IV discusses the design considerations and optimization variables. Section **Error! Reference source not found.** provides the optimization framework and algorithm. Experimental results and conclusion are presented in Section **Error! Reference source not found.** and Section **Error! Reference source not found.**, respectively.

## 2. NEAR-THRESHOLD COMPUTING

### 2.1 Transistor Modeling

First, we use NMOS transistors as an example. We know that the MOSFETs satisfy the  $\alpha$ -power law model in the traditional super-threshold regime [15]. On the other hand, the drain current  $I_{ds}$  of NMOS transistors operating in the subthreshold or near-threshold regime obeys an exponential dependency on the gate drive voltage  $V_{gs}$  and drain-to-source voltage  $V_{ds}$ , given by:

$$I_{ds} = \mu C_{ox} \frac{W}{L} (m-1) v_T^2 \cdot e^{\frac{V_{gs} + \lambda V_{ds} - V_{th}}{m \cdot v_T}} \left(1 - e^{-\frac{V_{ds}}{v_T}}\right), \quad (1)$$

where  $\mu$  is the mobility,  $C_{ox}$  is the oxide capacitance,  $m$  is the subthreshold slope factor,  $\lambda$  is the DIBL coefficient, and  $v_T$  is the thermal voltage  $\frac{kT}{q}$ . Given a specific technology node (e.g., the 22 nm PTM), we can rewrite Eqn. (1) as follows:

$$I_{ds} = I_0 W \cdot e^{\frac{V_{gs} + \lambda V_{ds} - V_{th}}{m \cdot v_T}} \left(1 - e^{-\frac{V_{ds}}{v_T}}\right), \quad (2)$$

where  $I_0$  is a technology-dependent parameter.

### 2.2 Circuit Modeling

In the circuit level, let  $P_{CB,dyn}(V)$  and  $P_{CB,sta}(V)$  denote the average dynamic (switching) power consumption and static (leakage) power consumption of the circuit block in the UDVS

structure, respectively, when the virtual- $V_{DD}$  value is  $V$ . Let  $P_{CB}(V)$  denote the average power consumption of the circuit block during operation time, and we have  $P_{CB}(V) = P_{CB,dyn}(V) + P_{CB,sta}(V)$ . Similarly, we define the average current values  $I_{CB,dyn}(V)$ ,  $I_{CB,sta}(V)$ , and  $I_{CB}(V)$ . Furthermore, let  $T_{CB}(V)$  denote the worst-case delay, i.e., the clock period, of the circuit block when virtual- $V_{DD}$  value is  $V$ . We characterize from ISCAS benchmarks and typical circuits and derive the corresponding functions. Figure 2 shows the measured and fitted dynamic and leakage power v.s. virtual- $V_{DD}$  of a typical circuit using 22nm PTM. Figure 2 also shows measured and fitted delay v.s. virtual- $V_{DD}$  of that circuit.

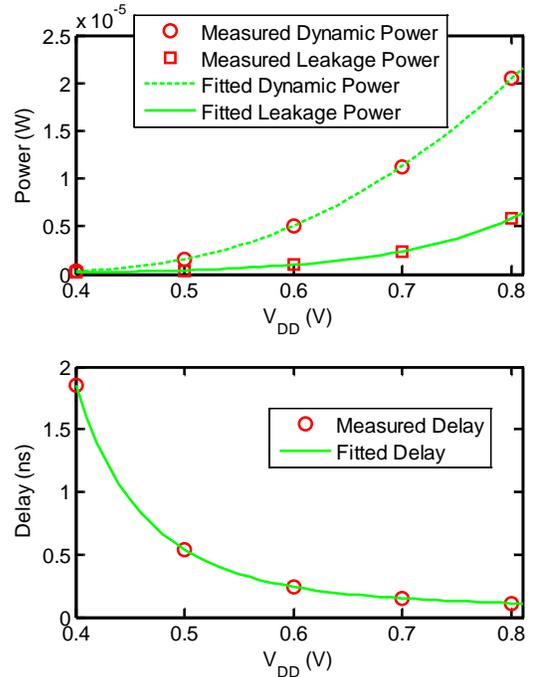


Figure 2. Characterization results of the circuit block in the UDVS structure.

## 3. STRUCTURE SUPPORT FOR FINE-GRAINED ULTRA DYNAMIC VOLTAGE SCALINGS

In this section, we propose the structure support for fine-grained UDVS over a wide voltage range from the traditional super-threshold regime down to the near-threshold regime. The proposed structure support induces less area overhead especially when the number of required virtual- $V_{DD}$  levels is relatively large. Please note that the determination procedure of the number of required virtual- $V_{DD}$  levels is out of the scope of this paper. Figure 1 shows an example with two  $V_{DD}$  power rails and four PMOS switches with different width values. We number the four PMOS switches as shown in Figure 1. We can generate three different virtual- $V_{DD}$  levels for the circuit block using the 1<sup>st</sup> and 2<sup>nd</sup> PMOS switches and the higher supply voltage rail  $V_{DDH}$  by turning on the 1<sup>st</sup> switch only, turning on the 2<sup>nd</sup> switch only, and turning on both switches, respectively. When both power switches are activated, the effective width is the sum of the width values of the two power switches. Similar observation also applies for the 3<sup>rd</sup> and 4<sup>th</sup> switches. Therefore, we can generate six potentially different virtual- $V_{DD}$  levels, denoted by  $V_1$  through  $V_6$  in the descending order, for the circuit block using this example structure. The three

higher virtual- $V_{DD}$  levels, i.e.,  $V_1$ ,  $V_2$ , and  $V_3$ , are generated by the first two switches and  $V_{DDH}$ , whereas the rest are generated by the last two switches and  $V_{DDL}$ . Proper sizing of the header switches is critical in order to generate the appropriate virtual- $V_{DD}$  levels for the circuit block to satisfy the target delay requirement at each DVS level.

Consider the 3<sup>rd</sup> and 4<sup>th</sup> PMOS switches that are connected to the lower supply voltage rail ( $V_{DDL}$ ). The body of these PMOS switches is tied to the virtual- $V_{DD}$  to avoid forward body bias, which results in a significant increase in the leakage current through these switches when the 1<sup>st</sup> and/or the 2<sup>nd</sup> switches are activated [7]. The gate drive signals of the 3<sup>rd</sup> and 4<sup>th</sup> PMOS switches are either connected to the ground when they are activated or to  $V_{DDH}$  if they are inactivated. These signals cannot be connected to  $V_{DDL}$  to be inactivated. This is because it will result in high ON-current flowing from the virtual- $V_{DD}$  to  $V_{DDL}$  when the 1<sup>st</sup> and/or the 2<sup>nd</sup> switches are activated (it is highly likely that the virtual- $V_{DD}$  level is higher than  $V_{DDL}$  in this case.)

#### 4. DESIGN CONSIDERATIONS AND OPTIMIZATION VARIABLES

In this section, we provide the design considerations and optimization variables for UDVS in the following four aspects: the number, type, and size of the header power switches, as well as the  $V_{DD}$  levels.

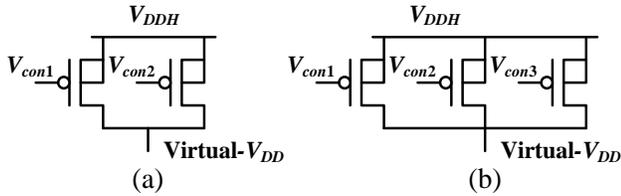


Figure 3. (a) Two parallel power switches and (b) three parallel power switches to achieve three different virtual- $V_{DD}$  levels.

**Number of Header Power Switches:** Consider only the header switches connecting to  $V_{DDH}$  as an example. Suppose that we are required to generate three different virtual- $V_{DD}$  levels, i.e.,  $V_1^{req}$ ,  $V_2^{req}$ , and  $V_3^{req}$ , using these switches and the  $V_{DDH}$  power rail in order to satisfy the corresponding frequency requirement at each DVS level. Then we may use either two parallel switches (as shown in Figure 3 (a)) or three parallel switches (as shown in Figure 3 (b)) to achieve this goal. When three parallel switches are utilized, we can achieve **exactly** the three required virtual- $V_{DD}$  levels by proper sizing of the parallel switches (even when leakage is considered.) On the other hand, when only two parallel switches are utilized, we can reduce the area overhead but may not generate exactly the three required virtual- $V_{DD}$  levels. In this case, one or two virtual- $V_{DD}$  levels generated by this structure may be inevitably higher than the required values in order to satisfy the three requirements simultaneously, which induces higher power/energy consumption. Utilization of two parallel switches will have another effect of reducing the leakage power consumption. In general, the former effect outweighs the latter effect, and therefore application of only two parallel switches will increase the overall power/energy consumption. Similar observation also applies to the header switches connected to  $V_{DDL}$ . Hence, the number of header power switches is an important design variable to achieve a desirable tradeoff between lower power/energy consumption and less area overhead.

**Type of Header Power Switches:** Consider the four power switches in Figure 1. We may replace some PMOS switches by NMOS switches and reduce area overhead while maintaining the same performance and power consumption, as illustrated in [7]. The 1<sup>st</sup> and 2<sup>nd</sup> PMOS switches cannot be replaced by NMOS ones. This is because an NMOS switch with a much larger size is required due to the relatively minor difference (less than the threshold voltage  $V_{th,n}$ ) between  $V_{DDH}$  and the required virtual- $V_{DD}$  level when the 1<sup>st</sup> and/or the 2<sup>nd</sup> power switch are activated. On the other hand, the 3<sup>rd</sup> and 4<sup>th</sup> PMOS switches, which are connected to  $V_{DDL}$ , may be potentially replaced by NMOS power switches as shown in Figure 4. In general, an NMOS switch induces less area overhead and is more desirable than its PMOS counterpart when

$$V_{DDH} - \text{Virtual-}V_{DD} - V_{th,n} > V_{DDL} - 0 - |V_{th,p}| \quad (3)$$

where  $V_{DDH} - \text{Virtual-}V_{DD}$  is the  $V_{gs}$  value when the NMOS switch is turned on, whereas  $V_{DDL} - 0$  is the  $|V_{gs}|$  value when the PMOS switch is turned on. Please note that Eqn. (3) is an approximate criterion since some secondary effects, such as the effect of body biasing or DIBL (drain-induced barrier lowering), are not accounted for. Moreover, utilizing NMOS switches will have another benefit of reducing the leakage power consumption mainly due to the reverse body biasing at any operation mode. Detailed discussions are omitted in this paper due to space limitation.

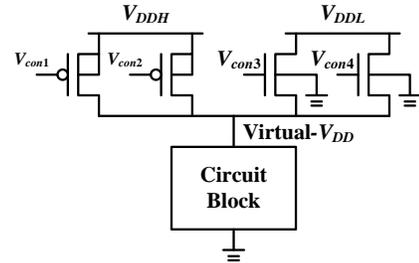


Figure 4. UDVS structure support with NMOS power switches.

**Sizing of Header Power Switches:** Appropriate sizing of the header power switches is crucial to maintain low power consumption and expected performance. Generally speaking, an undersized power switch results in a large performance degradation, whereas an oversized power switch results in increased leakage and increased area overhead. We need to perform joint sizing optimization of all the power switches in the proposed structure of UDVS, since the sizes of those power switches affect the virtual- $V_{DD}$  values in an interleaved manner. Let us consider the structure for UDVS in Figure 1 or Figure 4 again. Then we have the following two cases:

**Case I:** In this case the 1<sup>st</sup> and/or 2<sup>nd</sup> switches are active and  $V_1$ ,  $V_2$ , or  $V_3$  are generated as the virtual- $V_{DD}$  value. Increasing the size of the 1<sup>st</sup> or the 2<sup>nd</sup> power switch will result in an increase in the virtual- $V_{DD}$  level, whereas increasing the size of the 3<sup>rd</sup> or the 4<sup>th</sup> switch will result in a decrease. This is because current flows from  $V_{DDH}$  through virtual- $V_{DD}$  to  $V_{DDL}$  in this case (virtual- $V_{DD}$  is higher than  $V_{DDL}$ .)

**Case II:** In this case the 3<sup>rd</sup> and/or 4<sup>th</sup> switches are active and  $V_4$ ,  $V_5$ , or  $V_6$  are generated as the virtual- $V_{DD}$  value. Increasing the size of any power switch will result in an increase in the virtual- $V_{DD}$  level. This is because virtual- $V_{DD}$  is lower than both  $V_{DDH}$  and  $V_{DDL}$  in this case.

Because we need to satisfy the corresponding required virtual- $V_{DD}$  value at each DVS level, we should perform elaborate optimization on the sizes of power switches.

**Supply Voltage Levels in the Power Rails:** The supply voltage levels in the power rails, i.e.,  $V_{DDH}$  and  $V_{DDL}$  in Figure 1 or Figure 4, need to be jointly optimized with the power switches to achieve the globally optimal UDVS structure. A higher  $V_{DDH}$  or  $V_{DDL}$  value will reduce the required total width of power switches but incur higher power consumption, whereas a lower  $V_{DDH}$  or  $V_{DDL}$  value will have the opposite effect.

## 5. OPTIMIZATION FRAMEWORK

In this section, we propose the optimization framework of UDVS. We jointly optimize the supply voltage  $V_{DD}$  levels as well as the number, type (PMOS or NMOS), and size of the header power switches. We minimize the overall energy consumption of the UDVS circuit block, subject to the constraints on the number of supply power rails and the total area overhead. We account for both the dynamic energy consumption and leakage energy consumption through all the power switches during both the operation time and stand-by time of the circuit block. We formally describe the design optimization problem for UDVS as follows:

**Given:**  $M$  supply power rails (we use  $M = 2$  in the experiments);  $N$  different required virtual- $V_{DD}$  values, i.e.,  $V_1^{req}, V_2^{req}, \dots, V_N^{req}$ , which correspond to the  $N$  different required frequency/latency values at different DVS levels (we use  $N = 6$  in the experiments)<sup>1</sup>; the circuit block characteristics obtained from our characterization procedure.

**Find:** Number ( $K$ ), type (PMOS or NMOS), and width ( $W_1, W_2, \dots, W_K$ ) of all power switches, as well as the voltage supply levels  $V_{DDH}$  and  $V_{DDL}$ .

**Objective Functions:** We define two objective functions for minimization as follows. Let  $V_1, V_2, \dots, V_N$  denote the **actually generated** virtual- $V_{DD}$  levels using the UDVS structure. Let  $P_{DVS}(V_i)$  denote the (average) power consumption of the whole UDVS structure (including PMOS headers) during operation time when the generated virtual- $V_{DD}$  level is  $V_i$ , and let  $T_{DVS}(V_i)$  denote the corresponding latency value (clock period.) We know that  $P_{DVS}(V_i) \cdot T_{DVS}(V_i)$  is the energy consumption of the UDVS structure in one clock cycle, which has accounted for the conduction loss in PMOS headers. Then the first objective function, named the *weighted energy consumption*, is given as follows:

$$\sum_{i=1}^N \alpha_i \cdot P_{DVS}(V_i) \cdot T_{DVS}(V_i) + \alpha_0 \cdot P_{DVS,sta} \quad (4)$$

where  $\alpha_i$  ( $1 \leq i \leq N$ ) are the number of clock cycles when the circuit block operates at the  $i^{\text{th}}$  DVS level;  $\alpha_0$  is the idle time of the circuit block;  $P_{DVS,sta}$  is the leakage power consumption value of the UDVS structure.

For the second objective function, we know that the energy consumption per clock cycle of the circuit block (the power switches are not considered here) is given by  $P_{CB}(V_i^{req})$ .

$T_{CB}(V_i^{req})$ , when the supply voltage is  $V_i^{req}$ . Then the second objective function for minimization, named the *maximum energy overhead*, is given as follows:

$$\max_i \frac{P_{DVS}(V_i) \cdot T_{DVS}(V_i)}{P_{CB}(V_i^{req}) \cdot T_{CB}(V_i^{req})} \quad (5)$$

**Subject to:**

(i) *Virtual- $V_{DD}$  constraints:*  $V_i \geq V_i^{req}$  for  $1 \leq i \leq N$ .

(ii) *Area overhead constraint:*  $\sum_{i=1}^K W_i \leq W_{max}$ , where  $W_{max}$  is the maximum total width of the power switches.

Consider a UDVS structure with  $K_1$  power switches connected to  $V_{DDH}$  and  $K_2$  power switches connected to  $V_{DDL}$ , satisfying  $K_1 + K_2 = K$  and  $K \leq N$ . We name  $(K_1, K_2)$  a *configuration* of the  $N$ -level UDVS structure. For example, the UDVS structure with six required virtual- $V_{DD}$  levels has four configurations (2, 2), (2, 3), (3, 2), and (3, 3), if we generate the same number of virtual- $V_{DD}$  levels from  $V_{DDH}$  and  $V_{DDL}$ .

The proposed joint optimization algorithm consists of an outer loop and a kernel algorithm. The outer loop finds the best-suited configuration of UDVS structure as well as values of  $V_{DDH}$  and  $V_{DDL}$ . The kernel algorithm finds the optimal type and size of each power switch. The general procedure of the proposed joint optimization algorithm is shown in Algorithm 1.

---

Algorithm 1: Brief procedure of the joint optimization algorithm.

---

**For each** configuration of the UDVS structure:

**Perform** ternary search to find the optimal  $V_{DDH}$  and  $V_{DDL}$ :

*The kernel algorithm:*

Step I: Generate initial sizing of all power switches.

Step II: Generate feasible sizing of all power switches.

Step III: Determine the types of power switches

Step IV: Refine the sizing of all power switches

Find the optimal configuration and values of  $V_{DDH}$  and  $V_{DDL}$ , such that the objective function is minimized and constraints are satisfied.

---

The proposed kernel algorithm consists of four steps as shown in Algorithm 1. Without losing generality, we describe these four steps using configuration (3, 2) of the 6-level UDVS structure as an example. In this example, the 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> power switches are connected to  $V_{DDH}$ , whereas the 4<sup>th</sup> and 5<sup>th</sup> power switches are connected to  $V_{DDL}$ . Similar optimization steps can also be applied to the other configurations.

**Step I** (Generating the initial sizing of all power switches): In the first step, we generate the initial sizing of all power switches only considering the ON-currents of the power switches that are turned on, while neglecting the leakage currents of the other power switches. We continue with the above-mentioned example.

Let  $I_{PMOS}(V_s, V_d, V_g, V_b)$  denote the source-to-drain current of a unit-size PMOS switch with voltage levels at source, drain, gate, and body given by  $V_s, V_d, V_g$ , and  $V_b$ , respectively. Then for the three switches connected to  $V_{DDH}$  ( $1 \leq i \leq 3$ ), we generate an initial sizing as follows:

$$W_i = \frac{I_{CB}(V_i^{req})}{I_{PMOS}(V_{DDH}, V_i^{req}, 0, V_{DDH})} \quad (6)$$

<sup>1</sup> Please note that the focus of this work is to reduce the energy and area overhead in implementing the UDVS requirements. In the proposed framework,  $M$  and  $N$  can be general values. Necessities of multiple power rails and virtual- $V_{DD}$  values as well as the derivation procedure of the optimal  $M$  and  $N$  values are out of the scope of this paper.

In this way, we have  $V_i = V_i^{req}$  for  $1 \leq i \leq 3$  because the current flowing through the  $i^{\text{th}}$  PMOS switch matches the current flowing through the circuit block (leakage currents through other PMOS switches are ignored here.) On the other hand, the initial sizing of the two switches connected to  $V_{DDL}$  is more involved because the initial sizing should satisfy the following three constraints simultaneously:

$$W_4 + W_5 \geq \frac{I_{CB}(V_4^{req})}{I_{PMOS}(V_{DDL}, V_4^{req}, 0, V_4^{req})} \quad (7)$$

$$W_4 \geq \frac{I_{CB}(V_5^{req})}{I_{PMOS}(V_{DDL}, V_5^{req}, 0, V_5^{req})} \quad (8)$$

$$W_5 \geq \frac{I_{CB}(V_6^{req})}{I_{PMOS}(V_{DDL}, V_6^{req}, 0, V_6^{req})} \quad (9)$$

If (8) and (9) are the dominant constraints, we can set (8) and (9) to be equalities and  $W_4$  and  $W_5$  achieve the minimal possible value in this case. However, if (7) is the dominant constraint, we need to find the optimal  $W_4$  and  $W_5$  values such that the objective function (4) or (5) is minimized and constraints (7) – (9) are satisfied. Details are omitted due to space limitation.

**Step II** (Generating a feasible sizing of all power switches): In this step, we generate a feasible sizing of all power switches in the sense that the virtual- $V_{DD}$  constraints, i.e.,  $V_i \geq V_i^{req}$  for  $1 \leq i \leq N$ , are satisfied simultaneously. We consider both the ON-currents of the turned on switches and the leakage currents of the other power switches in this step. We continue with the above-mentioned example.

This step is based on the following observation from Section **Error! Reference source not found.**: Increasing the width of the 1<sup>st</sup>, 2<sup>nd</sup>, or 3<sup>rd</sup> switch can only increase the  $V_i$  values for  $1 \leq i \leq 6$ , whereas increasing the width of the 4<sup>th</sup> or 5<sup>th</sup> switch will increase  $V_4$ ,  $V_5$ , and  $V_6$  but decrease  $V_1$ ,  $V_2$ , and  $V_3$ . Hence when we check the virtual- $V_{DD}$  constraints taking into account the leakage currents, only the constraints on  $V_1$ ,  $V_2$ , and  $V_3$  may be violated. After we identify the virtual- $V_{DD}$  constraints that are violated, we increase the corresponding width of switches until there is no violation. Detailed procedure is omitted due to space limitation. The proposed procedure guarantees to find a feasible sizing of all power switches with no violation on virtual- $V_{DD}$  constraints. This is because increasing the width of the 1<sup>st</sup>, 2<sup>nd</sup>, or 3<sup>rd</sup> power switch will only increase the  $V_i$  values.

**Step III** (Determining the type of each power switch): In this step, we determine the type (NMOS or PMOS) of each power switch. Originally we set each power switch as PMOS switch. We continue with the above-mentioned example. We know from Section **Error! Reference source not found.** that the 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> power switch, which are connected to  $V_{DDH}$ , can only be implemented using PMOS switch. On the other hand, the 4<sup>th</sup> and 5<sup>th</sup> power switch can be potentially replaced by NMOS switch. For the 4<sup>th</sup> or 5<sup>th</sup> power switch, if we find out that NMOS power switch can achieve the same current driving capability with less width value, we conclude that NMOS is more suitable. We replace the original PMOS power switch by the NMOS one.

**Step IV** (Refining the sizing results of all power switches): Please note that we have the opportunity of refining, i.e., reducing, the sizing results of all power switches due to two reasons: (i) Some  $V_i$  values (such as the  $V_4$ ,  $V_5$ , and  $V_6$  values in the above-mentioned example) are higher than those calculated in Step I due

to the effect of leakage; (ii) Potential width increase of some power switches in Step II will further increase those  $V_i$  values. If no violation of virtual- $V_{DD}$  constraint will be resulted in, refining/reducing the width of a power switch will have two benefits: (i) reducing the ON-current and hence the power/energy consumption and (ii) reducing the leakage power consumption. In this step, we find and exploit the opportunity in reducing the sizing results of power switches derived from the previous steps. The detailed procedure is shown in Algorithm 2.

Algorithm 2: Refining the sizing results of power switches.

**Do** the following procedure:

Identify the set of power switches where reducing width by  $\Delta\%$  (a small amount) will not cause violation of virtual- $V_{DD}$  constraint.

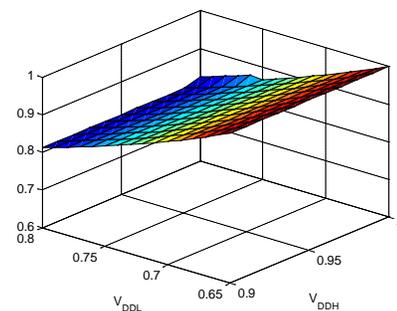
Identify the power switch from the set where reducing width will result in the minimal objective function value.

Reduce the width of the identified power switch by  $\Delta\%$ .

**Until** the sizing results cannot be further reduced, i.e., any further size reduction will cause violation in virtual- $V_{DD}$  constraint.

## 6. EXPERIMENTAL RESULTS

We test the proposed optimization framework of UDVS on the 22nm PTM [14]. We consider two supply power rails  $V_{DDH}$  and  $V_{DDL}$ , and six required virtual- $V_{DD}$  values, i.e.,  $V_1^{req} = 0.85$  V,  $V_2^{req} = 0.8$  V,  $V_3^{req} = 0.7$  V,  $V_4^{req} = 0.6$  V,  $V_5^{req} = 0.5$  V,  $V_6^{req} = 0.4$  V. Our proposed optimization framework finds the number, type and width of power switches as well as the values of  $V_{DDH}$  and  $V_{DDL}$  for the UDVS structure. The baseline UDVS structure also generates the same six required virtual- $V_{DD}$  values from two supply power rails  $V_{DDH}$  and  $V_{DDL}$ . In the baseline structure, the configuration is fixed at (3, 3), PMOS switches are used, and the values of  $V_{DDH}$  and  $V_{DDL}$  are fixed. We generate feasible sizes of power switches in baseline UDVS structure using the kernel algorithm up to Step II.

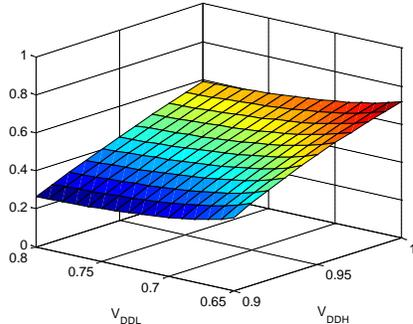


**Figure 5. The ratio of “maximum energy overhead” of the proposed UDVS structure to that of the baseline structure under the same area overhead.**

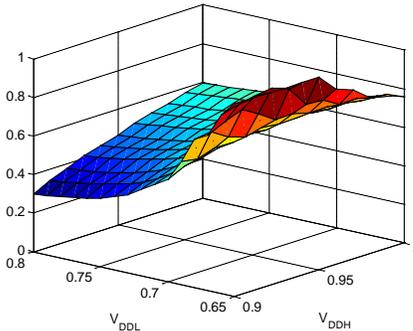
We compare the baseline UDVS structure with different pairs of ( $V_{DDH}$ ,  $V_{DDL}$ ) values with the proposed UDVS structure. We plot the ratio of the maximum energy overhead of the proposed UDVS structure to that of the baseline UDVS structure under the same area overhead in Figure 5. The proposed optimization framework can reduce the maximum energy overhead by up to 19% (occurs when  $V_{DDH} = 0.9$  V and  $V_{DDL} = 0.8$  V) compared to the baseline.

Figure 6 plots the ratio of area overhead of the proposed UDVS structure to that of the baseline UDVS structure, when they have the same maximum energy overhead. As can be seen in Figure

6, the proposed optimization framework reduces the area overhead of UDVS structure by up to 74% (occurs when  $V_{DDH} = 0.9$  V and  $V_{DDL} = 0.8$  V.) Figure 7 plots the ratio of the area overhead of the proposed UDVS structure to that of the baseline UDVS structure, when they have the same weighted energy consumption. In this case, the proposed optimization framework reduces the area overhead of UDVS structure by up to 70%.



**Figure 6. The ratio of area overhead of the proposed UDVS structure to that of the baseline structure under the same constraint of “maximum energy overhead”.**



**Figure 7. The ratio of the area overhead of the proposed UDVS structure to that of the baseline structure under the same constraint of “weighted energy consumption”.**

## 7. CONCLUSION

In this paper, we propose a structure support of fine-grained ultra dynamic voltage scaling (UDVS) with a limited number of power rails. The proposed structure support induces less area overhead than the reference methods especially when the number of supply voltage rails is relatively large. Moreover, we provide an optimization framework to jointly optimize supply voltage levels as well as the number, type (PMOS or NMOS), and size of the power switches. We minimize the overall energy consumption of the UDVS circuit block satisfying the target delay or frequency requirement at each DVS level. We take into account the additional constraints on the number of supply power rails and the total area overhead. The proposed optimization framework also properly accounts for the dynamic energy consumption as well as the leakage energy consumption through all the power switches

during both the operation time and stand-by time of the circuit block.

## 8. ACKNOWLEDGMENTS

This research is sponsored in part by grants from the PERFECT program of the Defense Advanced Research Projects Agency and the Software and Hardware Foundations of the National Science Foundation.

## 9. REFERENCES

- [1] R. Dreslinksi, M. Wiekowski, D. Blaauw, D. Sylvester, and T. Mudge, “Near-threshold computing: reclaiming Moore’s law through energy efficient integrated circuits,” *Proc. of IEEE*, Feb. 2010.
- [2] D. Markovic, C. Wang, L. Alarcon, T. Liu, and J. Rabaey, “Ultralow-power design in near-threshold region,” *Proc. of IEEE*, Feb. 2010.
- [3] A. Wang and A. Chandrakasan, “A 180 mV FFT processor using subthreshold circuit techniques,” *ISSCC*, Feb. 2004.
- [4] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, “The limit of dynamic voltage scaling and insomniac dynamic voltage scaling,” *IEEE Trans. on VLSI*, vol. 13, no. 11, pp. 1239 – 1252, Nov. 2005.
- [5] B. H. Calhoun, A. Wang, and A. Chandrakasan, “Modeling and sizing for minimum energy operation in subthreshold circuits,” *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1178 – 1786, Sept. 2005.
- [6] B. H. Calhoun, A. Wang, N. Verma, and A. Chandrakasan, “Sub-threshold design: the challenges of minimizing circuit energy,” *Proc. of International Symposium on Low Power Electronic Design*, 2006.
- [7] K. Craig, Y. Shakhsheer, and B. H. Calhoun, “Optimal power switch design for dynamic voltage scaling from high performance to subthreshold operation,” *Proc. of International Symposium on Low Power Electronic Design (ISLPED)*, 2012.
- [8] K. Craig, Y. Shakhsheer, S. Khanna, S. Arrabi, J. Lach, B. H. Calhoun, and S. Kosonocky, “A programmable resistive power grid for post-fabrication flexibility and energy tradeoffs,” *Proc. of International Symposium on Low Power Electronic Design (ISLPED)*, 2012.
- [9] Y. Shakhsheer, S. Khanna, K. Craig, S. Arrabi, J. Lach, and B. H. Calhoun, “A 90nm data flow processor demonstrating fine grained DVS for energy efficient operation from 0.25V to 1.2V,” *CICC*, 2011.
- [10] E. Pakbaznia and M. Pedram, “Coarse-grain MTCMOS sleep transistor sizing using delay budgeting,” *Proc. of Design, Automation, and Test in Europe (DATE)*, 2008.
- [11] M. Seok, S. Hanson, D. Sylvester, and D. Blaauw, “Analysis and optimization of sleep modes in subthreshold circuit design,” *Proc. of Design Automation Conference (DAC)*, 2007.
- [12] K. Shi and D. Howard, “Challenges in sleep transistor design and implementation in low-power designs,” *DAC*, 2006.
- [13] D. M. Harris, B. Keller, J. Karl, and S. Keller, “A transregional model for near-threshold circuits with application to minimum-energy operation,” in *International Conference on Microelectronics*, 2010.
- [14] W. Zhao and Y. Cao, “New generation of Predictive Technology Model for sub-45nm early design exploration,” *IEEE Transactions on Electronic Devices*, vol. 53, no. 11, pp. 2816 – 2823, Nov. 2006.
- [15] T. Sakurai and R. Newton, “Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas,” *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584 – 594, April 1990.