

# ACHIEVING CONTINUOUS $V_T$ PERFORMANCE IN A DUAL $V_T$ PROCESS

Kanak Agarwal, Dennis Sylvester, David Blaauw, Anirudh Devgan<sup>†</sup>

University of Michigan, <sup>†</sup>IBM Research

agarwalk@engin.umich.edu

**Abstract**— In this paper, we present a novel approach to obtain any desired intermediate threshold voltage in a dual  $V_T$  process. The intermediate threshold voltages are achieved by combining low and high threshold voltages in a device. We show that this combination can be easily implemented in layouts with negligible design and manufacturing overhead. Our results show that power-delay characteristics of the achieved intermediate thresholds match well with the ideal (but impractical) scenario that assumes that all intermediate thresholds are available in the technology.

## 1. INTRODUCTION

Due to continuous scaling of threshold voltage and oxide thickness, leakage power dissipation has become one of the most critical challenges for nanometer technologies. Over the past few years, leakage power has grown exponentially with each new technology generation and has now become a significant component of total chip power consumption [1,2].

In order to cope with the leakage problem, most advanced processes provide an option of two or more threshold voltages. Low threshold ( $V_{TL}$ ) devices are used in timing critical paths while high threshold ( $V_{TH}$ ) devices are used in paths that have sufficient timing slack or where performance is non-critical [3,4]. This allows significant savings in leakage since  $V_{TH}$  devices often exhibit an order of magnitude lower leakage current than  $V_{TL}$  devices. However, one limitation of this methodology is that low and high thresholds are fixed for a particular process and cannot be altered during design. Hence,  $V_{TL}$  and  $V_{TH}$  values must be chosen very carefully during process development. This is not a trivial task - if the high threshold is set much higher than the low threshold to enable more savings in power, then the  $V_{TH}$  devices are significantly slower and cannot be used very often. Conversely, if high threshold is set close to the low threshold to allow more extensive use of  $V_{TH}$  devices, then the savings in power are not as significant. Also, in the latter case, even after using  $V_{TH}$  devices, a design may still contain paths with significant slack. Power associated with these paths is thus wasted due to the non-availability of a threshold higher than the one provided by the process. This is especially an issue when different designs manufactured in the same technology node have different power and speed requirements and hence different optimal low and high thresholds.

One way to handle the above problem is through technology or process-based solutions. The most obvious and commonly employed solution is to increase the number of thresholds provided by the process. Many advanced processes now provide three threshold voltage options. This provides added flexibility in trading off power with performance for an individual design. Another technology-based solution is novel device structures that allow control of the device characteristics such as threshold voltage and subthreshold swing dynamically. The underlying idea behind such devices (double-gated devices or FinFETs) is to use additional gate surfaces to control a very thin channel [5,6]. These technology-based solutions can be very effective in controlling leakage but the design and manufacturing costs associated with them can be too high to justify their use and many are not yet available in production.

A second class of solutions that provide a better control of threshold voltage are design-based solutions. The most commonly employed

approach here is adaptive body biasing [7,8]. This approach exploits the fact that the threshold voltage of a device can be increased by reverse biasing the source-substrate junction. This allows control of the device threshold and leakage by appropriately biasing the body voltage. However, body biasing techniques require significant design overhead and hence can only be used to control threshold either at the full-chip level or possibly with block-level granularity [9].

In this paper, we propose a novel design approach that allows us to obtain any intermediate threshold voltage between the low and high threshold voltages available in the process. The underlying idea behind the proposed methodology is to combine characteristics of low and high threshold voltage transistors to achieve an intermediate desired threshold. The proposed approach has minimal design and manufacturing overhead and can be easily implemented in modern design flows. The ability to achieve desired intermediate threshold voltages is very useful in minimizing power as it allows us to assign threshold voltages in a continuous manner as compared to existing discrete  $V_T$  assignment approaches [10-12].

The remainder of the paper is organized as follows. In the next section, we describe the basic concept of jointly using low and high thresholds devices. Section 3 presents our methodology to obtain a desired intermediate threshold voltage. Section 4 provides experimental results and we conclude in Section 5.

## 2. COMBINING LOW AND HIGH THRESHOLDS

The underlying idea behind our approach is to combine low and high threshold behavior to achieve power-delay characteristics that are representative of an intermediate threshold voltage device. In this section, we show how low and high threshold transistors can be combined with negligible design and area overhead.

Let us consider the case of a wide device. We assume that this device is implemented using parallel fingers as shown in Figure 1. Splitting wide devices into parallel fingers is a standard practice since it reduces gate series resistance by interdigitating the polysilicon gate and it also reduces source/drain diffusion capacitances by allowing source/drain sharing [13]. Another advantage of such a layout scheme is that it helps in maintaining a certain geometric regularity as required by standard cell layouts. Usually, all fingers in such devices have the same threshold voltage adjust implant. However, in a dual  $V_T$  process, a subset of fingers can be doped with the low threshold implant while the remaining set of fingers can be made high threshold as shown in the figure. A mixed threshold device obtained in this manner creates the ability to have power and delay characteristics equivalent to that of a transistor with an intermediate threshold voltage. This is the basic idea of our approach.

As can be seen from the above discussion, this technique is applicable only to wide devices that are split into parallel fingers. However, this is not a significant drawback since leakage in small devices is typically small as compared to the leakage in large timing critical devices. Generally, small devices are not critical for timing and can be easily doped with the high threshold implant. Another point to note here is that the parallel fingers in the interdigitated layout structure are separated by a contacted pitch. As a result, the two adjacent fingers are sufficiently spaced to enable manufacturability of different threshold implants with zero or negligible area overhead.

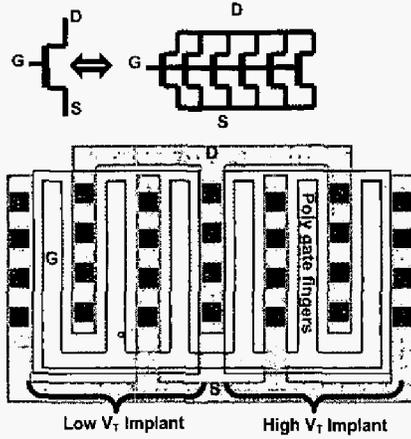


Figure 1: Layout diagram demonstrating splitting of wide devices into parallel fingers. In a dual threshold process, the fingers can be doped with different implants.

### 3. ACHIEVING INTERMEDIATE THRESHOLDS

In this section, we provide a theoretical framework to show how an intermediate threshold can be obtained through a combination of low and high threshold voltages.

#### 3.1 Delay and Power Models

We use the alpha power law based delay model for our analysis [14]. According to this model, delay of a gate with threshold voltage  $V_T$  is given by:

$$\text{Delay} = \frac{KV_{DD}C_L}{W(V_{DD} - V_T)^\alpha} \quad (1)$$

Here  $K$  and  $\alpha$  (velocity saturation index) are technology constants,  $V_{DD}$  is the supply voltage,  $C_L$  is the load capacitance seen by the gate, and  $W$  is the width of the gate.

The simplified power model used for our analysis assumes that power is proportional to the widths of the devices. This holds true for leakage (subthreshold and gate) as well as dynamic power. According to this model, total power of a gate with threshold voltage  $V_T$  is given by:

$$\text{Power} = W \times P_{VT} \quad (2)$$

Here,  $P_{VT}$  is the power per unit width of the gate.

#### 3.2 Realizing Intermediate Thresholds

Let us assume that low and high thresholds available in the process are represented by  $V_{TL}$  and  $V_{TH}$  respectively. The objective is to obtain any intermediate desired threshold voltage  $V_{TD}$  (where  $V_{TL} \leq V_{TD} \leq V_{TH}$ ) by a combination of  $V_{TL}$  and  $V_{TH}$  transistors. Obtaining any desired threshold voltage  $V_{TD}$  requires realizing a gate that behaves like a gate at threshold voltage  $V_{TD}$ . In other words, the realized gate should have both delay and power less than or equal to the delay and power of a  $V_{TD}$  gate.

We first examine a simple circuit shown in Figure 2. We point out here that Figure 2 shows two gates in parallel only for illustration purposes. In reality, a combination of low and high threshold gates would be achieved by using the methodology discussed in Section 2. Here, we are trying to obtain a  $V_{TD}$  gate of width  $W$  with a parallel combination of  $V_{TH}$  and  $V_{TL}$  gates. Let us assume that a fraction  $f$  ( $0 \leq f \leq 1$ ) of the total width  $W$  is assigned to the  $V_{TH}$  gate while the remaining width is given to the  $V_{TL}$  gate.

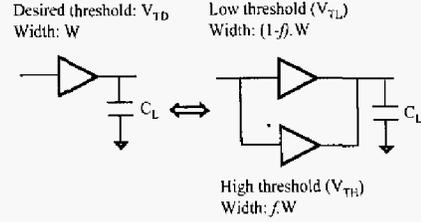


Figure 2: Achieving intermediate threshold by splitting device width into low and high thresholds.

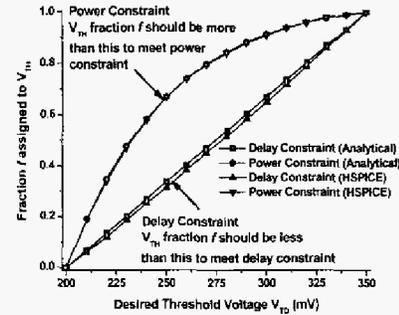


Figure 3: Fraction of total width that must be assigned to the high threshold gate in order to achieve an intermediate threshold.

The delay of the parallel combination of  $V_{TH}$  and  $V_{TL}$  gates can be expressed as:

$$\text{Delay} = \frac{KV_{DD}C_L}{fW(V_{DD} - V_{TH})^\alpha + (1-f)W(V_{DD} - V_{TL})^\alpha} \quad (3)$$

In order for the parallel gates to behave identically to a single gate of threshold  $V_{TD}$ , the following constraints must be satisfied.

*Delay Constraint*

$$\frac{KV_{DD}C_L}{fW(V_{DD} - V_{TH})^\alpha + (1-f)W(V_{DD} - V_{TL})^\alpha} \leq \frac{KV_{DD}C_L}{W(V_{DD} - V_{TD})^\alpha}$$

*Power Constraint*

$$fW(P_{VTH}) + (1-f)W(P_{VTL}) \leq W(P_{VTD}) \quad (4)$$

Here,  $P_{VTH}$ ,  $P_{VTL}$ , and  $P_{VTD}$  represent power per unit width for  $V_{TH}$ ,  $V_{TL}$ , and  $V_{TD}$  gates. The first condition in Equation 4 represents that the delay of the parallel combination should be less than the delay of a single  $V_{TD}$  gate. The second condition represents a similar constraint on total power. These constraints result in the following delay and power constraints on the fraction  $f$  ( $0 \leq f \leq 1$ ) of the total width  $W$  that can be assigned to the  $V_{TH}$  gate.

$$\text{Delay Constraint} \quad f \leq \frac{(V_{DD} - V_{TL})^\alpha - (V_{DD} - V_{TD})^\alpha}{(V_{DD} - V_{TL})^\alpha - (V_{DD} - V_{TH})^\alpha}$$

$$\text{Power Constraint} \quad f \geq \frac{P_{VTL} - P_{VTD}}{P_{VTL} - P_{VTH}} \quad (5)$$

Figure 3 plots the conditions shown in Equation 5 for an arbitrarily chosen threshold voltage range of 200 mV to 350 mV. An industry 90nm SOI process was used in the analysis. In order to verify the accuracy of our model (and hence the accuracy of the underlying delay and power models as expressed in Equations 1 and 2), the figure also shows the same constraints obtained through HSPICE simulations. To generate constraints from simulations, we consider a single gate at each threshold and measure its delay and power. Next we simulate a

circuit containing a combination of low and high threshold gates as shown in Figure 2 and sweep the value of fraction  $f$  from 0 to 1 in small steps. The value of fraction  $f$  at which delay of the combination becomes equal to that of the single gate is then measured to find the delay constraint at this threshold. Similarly, the value of  $f$  at which power becomes the same in two configurations is measured to find the power constraint.

Some important conclusions can be drawn from this experiment. The most important observation is that it is not possible to realize a gate at an intermediate threshold  $V_{TD}$  by simple partitioning of its total width  $W$  between  $V_{TH}$  and  $V_{TL}$  gates. This is because delay is a relatively weak function of threshold voltage compared to power (i.e., delay is roughly linear with  $V_T$  while leakage power varies exponentially with threshold voltage). Hence at a desired threshold  $V_{TD}$ , power constraints cause the  $V_{TH}$  fraction to be significantly higher than is acceptable from a delay standpoint.

The above experiment shows the combination gate is worse than a single gate in terms of power-delay. This implies that total gate width of the combination must be increased in order to meet both delay and power constraints. We examine a circuit similar to the circuit shown in Figure 2, however, this time we allow the widths of  $V_{TH}$  and  $V_{TL}$  gates to change independently. The new setup is shown in Figure 4. For this setup, the delay and power constraints can be written in the following manner.

*Delay Constraint*

$$\frac{KV_{DD}C_L}{W_{VTH}(V_{DD}-V_{TH})^\alpha + W_{VTL}(V_{DD}-V_{TL})^\alpha} = \frac{KV_{DD}C_L}{W(V_{DD}-V_{TD})^\alpha}$$

$$W_{VTH}(V_{DD}-V_{TH})^\alpha + W_{VTL}(V_{DD}-V_{TL})^\alpha = W(V_{DD}-V_{TD})^\alpha$$

*Power Constraint*

$$W_{VTH}(P_{VTH}) + W_{VTL}(P_{VTL}) = W(P_{VTD}) \quad (6)$$

Equation 6 contains two equations in terms of two unknowns  $W_{VTH}$  and  $W_{VTL}$ . Solving Equation 6 for  $W_{VTH}$  and  $W_{VTL}$  gives

$$\frac{W_{VTH}}{W} = \frac{(V_{DD}-V_{TD})^\alpha P_{VTL} - (V_{DD}-V_{TL})^\alpha P_{VTD}}{(V_{DD}-V_{TH})^\alpha P_{VTL} - (V_{DD}-V_{TL})^\alpha P_{VTH}} \quad (7)$$

$$\frac{W_{VTL}}{W} = \frac{(V_{DD}-V_{TH})^\alpha P_{VTD} - (V_{DD}-V_{TD})^\alpha P_{VTH}}{(V_{DD}-V_{TH})^\alpha P_{VTL} - (V_{DD}-V_{TL})^\alpha P_{VTH}}$$

Figure 5 shows  $(W_{VTH}/W)$  and  $(W_{VTL}/W)$  for threshold voltage range of 200-350 mV. Power per unit width ( $P_{VTH}$ ,  $P_{VTL}$  and  $P_{VTD}$ ) at each threshold was computed using SPICE simulations and includes dynamic power (activity factor = 0.05 and frequency = 1 GHz) along with leakage power.

The above analysis shows that any intermediate threshold voltage between  $V_{TH}$  and  $V_{TL}$  can be realized by using a combination of high and low threshold voltages if the total device width is allowed to change. For example, in order to realize a gate with a threshold voltage of 250 mV, we can compute widths of low and high threshold devices from Figure 5 (Equation 7). These high and low threshold devices are then connected in parallel as shown in Figure 4. The resulting combination has the same delay and power as that of a single gate with 250 mV threshold voltage. In other words, the resulting combination behaves like a gate at desired threshold voltage of 250 mV.

Realizing an intermediate threshold causes an increase in the total device width. Our results show that this increase in device width is not large. For example, in Figure 5 the maximum increase in the total width is only 10%. We emphasize here that in the formulation shown in Equation 6,  $P_{VTH}$ ,  $P_{VTL}$  and  $P_{VTD}$  represent total (dynamic + leakage) power per unit width. Hence, even though there was an increase in total width while realizing an intermediate threshold through  $V_{TL}$ - $V_{TH}$  combination, total power (including dynamic power) stays the same.

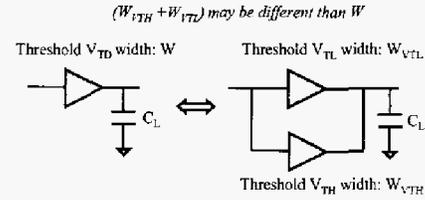


Figure 4: Achieving intermediate threshold by combination of low and high thresholds.

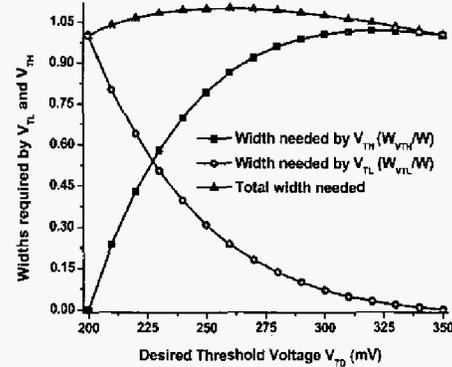


Figure 5 Widths required by low and high threshold gates in order to achieve an intermediate threshold.

The only impact of increase in device width is that it may increase the loading capacitance seen by the preceding gate. So far, in our analysis, we have assumed an independent gate driving a constant load. In reality, the gate will be embedded in a larger circuit and any increase in its device width may change the loading capacitance seen by other gates. This impact of change in load capacitance with device width is discussed in Section 3.4.

An interesting observation here is that the widths of low and high threshold devices required to realize an intermediate voltage ( $W_{VTL}/W$  and  $W_{VTH}/W$  in Equation 7) depend only on technology parameters and are independent of the circuit analyzed. This implies that these widths can be characterized once for each value of  $V_{TD}$  between  $V_{TH}$  and  $V_{TL}$  (either by simulation or by proposed analytical model) and can be used whenever needed.

### 3.3 Realizability of Intermediate Thresholds

In this section, we discuss the realizability of all intermediate threshold voltages by the methodology proposed in Section 3.2. In particular, we highlight the cases when it may not be possible to realize all threshold voltages in the low to high threshold range. Finally, we show that such non-realizable threshold voltages are of little or no interest as one would rarely want to operate at these thresholds.

In order to be able to realize all intermediate threshold voltages, both  $W_{VTL}$  and  $W_{VTH}$  should be positive for all values of  $V_{TD}$  between  $V_{TL}$  and  $V_{TH}$ . The denominator in the expressions of  $W_{VTL}$  and  $W_{VTH}$  given in Equation 7 is independent of  $V_{TD}$ . This denominator is a positive number for all practical values of  $V_{TL}$  and  $V_{TH}$ . It is due to the fact that  $P_{VTL}$  is usually much higher as compared to  $P_{VTH}$  while the difference between  $(V_{DD}-V_{TH})^\alpha$  and  $(V_{DD}-V_{TL})^\alpha$  is relatively smaller. Hence, in order to have realizable values of  $W_{VTL}$  and  $W_{VTH}$ , the numerators should also be positive. This in turn requires that the following equations must be satisfied for all values of  $V_{TD}$  between  $V_{TL}$  and  $V_{TH}$ .

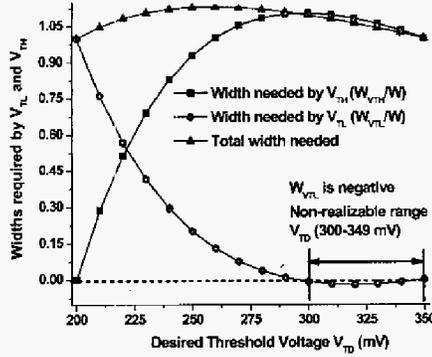


Figure 6: Widths required by low and high threshold gates in order to achieve an intermediate threshold. Threshold voltage range (300-349 mV) is non-realizable in this case.

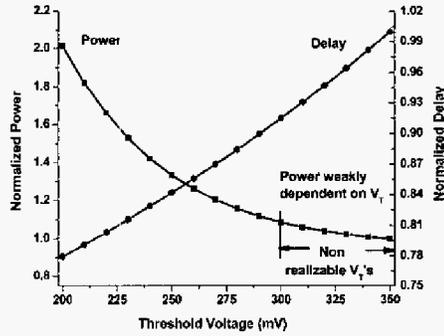


Figure 7: Total power and delay as function of threshold voltage for the example considered in Figure 7.

For  $W_{V_{TH}}$  to be positive

$$\frac{P_{V_{TL}}}{P_{V_{TD}}} \geq \frac{(V_{DD} - V_{TL})^\alpha}{(V_{DD} - V_{TD})^\alpha} \quad \text{or} \quad \frac{P_{V_{TL}}}{P_{V_{TD}}} \geq \frac{(Delay)_{V_{TD}}}{(Delay)_{V_{TL}}} \quad (8)$$

Similarly, for  $W_{V_{TL}}$  to be positive

$$\frac{P_{V_{TD}}}{P_{V_{TH}}} \geq \frac{(V_{DD} - V_{TD})^\alpha}{(V_{DD} - V_{TH})^\alpha} \quad \text{or} \quad \frac{P_{V_{TD}}}{P_{V_{TH}}} \geq \frac{(Delay)_{V_{TH}}}{(Delay)_{V_{TD}}} \quad (9)$$

If we examine these equations carefully, the first equation implies that when threshold voltage is increased with respect to  $V_{TL}$ , the power ratio ( $P_{V_{TL}}/P_{V_{TD}}$ ) should be higher than the delay ratio ( $(Delay)_{V_{TD}}/(Delay)_{V_{TL}}$ ). Equation 9 implies that when threshold voltage is decreased with respect to  $V_{TH}$ , the power ratio ( $P_{V_{TD}}/P_{V_{TH}}$ ) should be higher than the delay ratio ( $(Delay)_{V_{TH}}/(Delay)_{V_{TD}}$ ). These equations are usually satisfied since power is a stronger function of threshold voltage than delay. Hence, any change in threshold voltage impacts power more strongly than it impacts delay. However, there may exist scenarios when these constraints may be violated, thereby leading to non-realizable values of  $V_{TD}$  lying between  $V_{TL}$  and  $V_{TH}$ .

We explain this by an example shown in Figure 6. Here, we consider the same setup as Figure 5 but we increase dynamic power significantly (activity factor = 0.15 and frequency = 1.5 GHz). Figure 7 shows total power and delay as a function of threshold voltage. We draw the following conclusions from Figures 6 and 7.

- $W_{V_{TH}}$  is always positive because the power ratio with respect to  $V_{TL}$  ( $P_{V_{TL}}/P_{V_{TD}}$ ) is always higher than the corresponding delay ratio ( $(Delay)_{V_{TD}}/(Delay)_{V_{TL}}$ ) for all practical values of  $V_{TL}$ . This is due to the strong exponential dependence of power on  $V_T$  at low threshold voltages when leakage dominates. This causes power consumption at  $V_{TL}$  to

be significantly higher than that at any other threshold and hence the ratio ( $P_{V_{TL}}/P_{V_{TD}}$ ) is always larger than the ratio ( $(Delay)_{V_{TD}}/(Delay)_{V_{TL}}$ ).

- $W_{V_{TL}}$  can become negative for some range of threshold voltages close to  $V_{TH}$ . This occurs since power is not very sensitive to  $V_T$  in this region (as shown in Figure 8). This can happen because near  $V_{TH}$ , the main component of leakage is gate leakage, which is independent of threshold voltage. Also, since dynamic power is not sensitive to  $V_T$ , in dynamic power dominated designs ( $P_{V_{TD}}/P_{V_{TH}}$ ) can become very small in the region close to  $V_{TH}$ . On the other hand, delay continues to vary in a similar manner with  $V_T$ . This results in violation of Equation 9 in the region close to  $V_{TH}$ , thereby resulting in a non-realizable  $W_{V_{TL}}$ .

We have shown that all threshold voltages in  $V_{TL}$  to  $V_{TH}$  range may not be realizable by proposed approach. However, it can be readily seen that it is rarely beneficial to operate in non-realizable regions. In these regions, the improvements in power are minor while delay continues to worsen. When delay is not a concern, assigning gates to operate at  $V_{TH}$  is the best design point. On the other hand, if delay is tightly constrained, operating in a non-realizable region does not provide any gains. This is explained by the fact that we can simply operate at a lower limit of the non-realizable range (for example 300 mV in Figure 7) with nearly the same power as any threshold in the non-realizable region but with a much better performance. The ability to achieve intermediate thresholds provides a continuous trade-off between power and delay. Our approach is useful for problems where we want to minimize power under a given delay constraint. Any arbitrary  $V_T$  can be obtained through our methodology, thereby allowing us more flexibility than provided by a simple two  $V_T$  process.

### 3.4 Intermediate Thresholds for Non-Constant Loads

We have seen that achieving an intermediate threshold voltage results in an increase in total width. The analysis in Section 3.2 assumed an independent gate driving a constant load. In reality, the gates are embedded in a larger circuit and any increase in their total width may change the loading capacitance seen by other gates. The impact of changing input capacitance due to larger total device width is discussed in this section.

Let us first consider a simple case as shown in Figure 8. We explain our theory for this simple circuit. Later, we will show how the same methodology can be extended to larger circuit topologies. We assume that the desired thresholds of first and second gates are  $V_{TD}$  and  $V_{TD1}$  respectively. Our goal is to realize this circuit using low and high threshold gates.

In order to realize the above circuit, we begin by modifying the second gate. This gate is driving a constant load and can be easily realized using the methodology discussed in Section 3.2 (Equation 7). Once the second gate has been transformed, the load capacitance seen by the first gate in the intended circuit ( $C_L^{new}$ ) increases relative to the load seen by the first gate in the original circuit ( $C_L$ ). The values of  $C_L^{new}$  and  $C_L$  can be calculated as shown in the figure. Once the values of load capacitances in the original and intended circuits are known, then first gate can be realized using Equation 11 as derived below.

For a single gate (threshold voltage  $V_{TD}$  and load  $C_L$ ) and a  $V_{TL}$ - $V_{TH}$  combination (load  $C_L^{new}$ ), the delay and power constraints can be written in the manner similar to Equation 6.

Delay Constraint

$$\frac{KV_{DD}C_L^{new}}{W_{V_{TH}}(V_{DD} - V_{TH})^\alpha + W_{V_{TL}}(V_{DD} - V_{TL})^\alpha} = \frac{KV_{DD}C_L}{W(V_{DD} - V_{TD})^\alpha}$$

$$W_{V_{TH}}(V_{DD} - V_{TH})^\alpha + W_{V_{TL}}(V_{DD} - V_{TL})^\alpha = \frac{C_L^{new}}{C_L} W(V_{DD} - V_{TD})^\alpha$$

Power Constraint

$$W_{V_{TH}}(P_{V_{TH}}) + W_{V_{TL}}(P_{V_{TL}}) = W(P_{V_{TD}}) \quad (10)$$

Equation 10 contains two equations in terms of two unknowns  $W_{V_{TH}}$  and  $W_{V_{TL}}$ . Solving Equation 10 for  $W_{V_{TH}}$  and  $W_{V_{TL}}$  gives

$$\frac{W_{V_{TH}}}{W} = \frac{\frac{C_L^{new}}{C_L}(V_{DD}-V_{TD})^\alpha P_{V_{TL}} - (V_{DD}-V_{TL})^\alpha P_{V_{TD}}}{(V_{DD}-V_{TH})^\alpha P_{V_{TL}} - (V_{DD}-V_{TL})^\alpha P_{V_{TH}}} \quad (11)$$

$$\frac{W_{V_{TL}}}{W} = \frac{(V_{DD}-V_{TH})^\alpha P_{V_{TD}} - \frac{C_L^{new}}{C_L}(V_{DD}-V_{TD})^\alpha P_{V_{TH}}}{(V_{DD}-V_{TH})^\alpha P_{V_{TL}} - (V_{DD}-V_{TL})^\alpha P_{V_{TH}}}$$

Equation 11 is similar to Equation 7 other than additional terms representing ratio of load capacitances. The ratio of  $C_L^{new}$  to  $C_L$  is usually greater than one due to an increase in width while realizing an intermediate threshold. Figure 9 shows  $(W_{V_{TH}}/W)$  and  $(W_{V_{TL}}/W)$  as computed using Equation 11 for a  $C_L^{new}$  to  $C_L$  ratio of 1.1.

The methodology discussed above can be extended to implement larger circuits in a similar manner. The general procedure is outlined below.

- Start with gates at sink nodes and realize their desired thresholds using Equation 7 (constant load case).
- Move towards fan-in nodes of the gates realized in the above step. Compute new load capacitances ( $C_L^{new}$ 's) seen by these gates based on the knowledge of increase in width of the fan-out gates. Use Equation 11 to implement desired thresholds in this logic level.
- Continue moving towards source node in a manner similar to the previous step (breadth-first search, realizing one logic level at a time) until the entire circuit has been mapped to the desired threshold voltages.

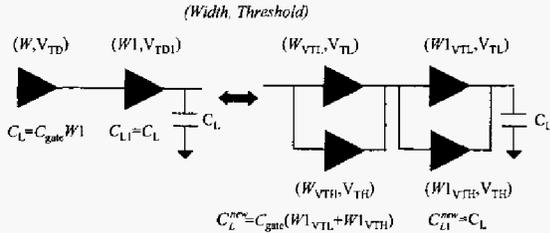


Figure 8: Achieving intermediate thresholds in general logic networks.

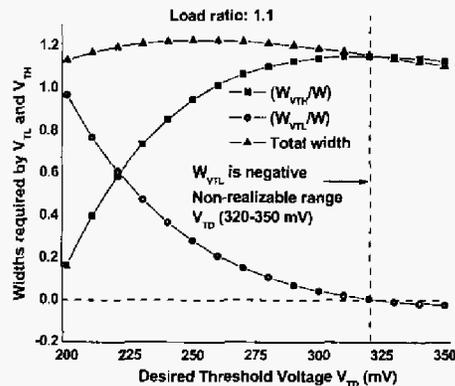


Figure 9: Widths required to achieve intermediate threshold voltages for a load ratio of 1.1.

The above procedure can be used to realize any circuit with arbitrary intermediate threshold voltages. However, there are the following limitations to this methodology.

- The total width required to achieve an intermediate threshold increases with logic depth. This is due to the fact that as the ratio of  $C_L^{new}$  to  $C_L$  increases, additional width is required to implement a desired threshold, which in turn results in further increases in the ratio of load capacitances seen by the preceding gates. We point out here that although there is an increase in gate width with logic depth, total power (sum of static and dynamic power) and delay of the realized circuit is still matched with the ideal desired case. This is because the effect of increases in width on total power and delay are taken into account while setting the power-delay constraints in Equation 10.

- A second limitation of the above methodology is that the realizable range of threshold voltages shrinks with logic depth. If we compare Equation 11 to Equation 7, we can see that the inclusion of the

$(C_L^{new}/C_L)$  ratio causes the high threshold width to increase while the width allocated to the low threshold reduces. In Section 3.3, we showed that  $W_{V_{TH}}$  is always positive while  $W_{V_{TL}}$  may become negative in the region close to  $V_{TH}$ . With load capacitances, the new realizability constraint on  $W_{V_{TL}}$  can be reformulated. For  $W_{V_{TL}}$  to be positive:

$$\frac{P_{V_{TD}}}{P_{V_{TH}}} \geq \frac{C_L^{new}}{C_L} \frac{(V_{DD}-V_{TD})^\alpha}{(V_{DD}-V_{TH})^\alpha} \quad \text{or} \quad \frac{P_{V_{TD}}}{P_{V_{TH}}} \geq \frac{C_L^{new}}{C_L} \frac{(\text{Delay})_{V_{TH}}}{(\text{Delay})_{V_{TD}}} \quad (12)$$

As the ratio of  $C_L^{new}$  to  $C_L$  increases, Equation 12 is violated for a larger range of  $V_{TD}$ 's, thereby shrinking the realizable window. As a consequence, there may be cases where an arbitrarily chosen threshold may not be achievable by the proposed approach. Hence, any mixed threshold based power optimization algorithm must take realizability limitations into account during threshold assignment.

## 4. RESULTS

Section 3 demonstrates how to achieve a desired intermediate threshold voltage within a standard dual- $V_T$  process. In this section, we compare the theoretical results developed in that section against HSPICE simulations and show that power-delay response of the achieved intermediate thresholds match well with the ideal (but impractical) scenario that assumes that all intermediate thresholds are available in the technology.

As a first exercise, we verify the theory proposed in Section 3.2. In order to test Equation 7 (Figure 5), we consider a single gate driving a fixed load capacitance and sweep its threshold voltage from 200 mV to 350 mV. At each intermediate threshold, power and delay of the gate were measured for the ideal case (assuming that the desired intermediate threshold is available in the technology) and for the realized case. For the realized case, the intermediate thresholds were obtained by mixing high and low threshold voltages in the ratio given by Equation 7. Figure 10 shows the comparison in power and delay as obtained from HSPICE simulations. The figure shows that the curves match very well. This result not only verifies the analytical equations but it also confirms the proposed theory that a desired intermediate threshold voltage can be achieved by mixing low and high thresholds.

Next, we focus on generic circuits and show that the desired intermediate threshold voltages in such circuits can be easily achieved by the methodology proposed in Section 3.4. To verify the theory proposed in that section, we consider a logic path consisting of ten fanout-of-four inverter (FO4) stages. We assume that all gates in the setup have a desired intermediate threshold voltage of 250 mV whereas  $V_{TL}$  and  $V_{TH}$  are 200mV and 350mV respectively.

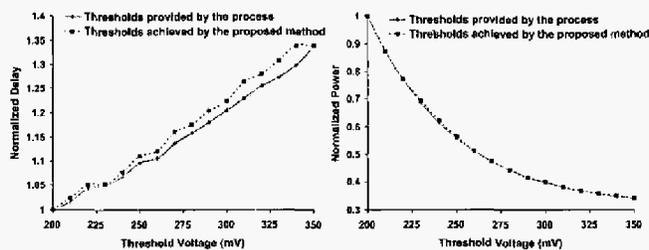


Figure 10: Comparing power-delay response of the ideal thresholds (intermediate thresholds are available in the technology) with the proposed method.

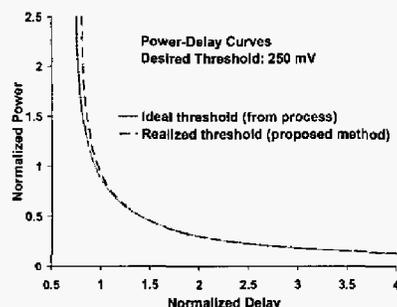


Figure 11: Power-delay curves for the ideal and the realized cases for the ten FO4 testcase match very well.

Figure 11 compares power-delay curves for this testcase for the ideal case (assuming that desired intermediate threshold is available in the technology) and the realized case as obtained from HSPICE simulations. For the realized case, the desired threshold voltages were achieved using the methodology discussed in Section 3.4. The power-delay curves were obtained by varying gate widths and measuring power at each value of delay. Figure 11 shows that the two power-delay curves match very well, implying that even for general logic networks the power-delay behavior of intermediate thresholds can be achieved with the proposed method. However, as discussed in Section 3.4, total width required to achieve an intermediate threshold increases with logic depth. For this testcase, as we moved from the first to the tenth logic stage, total width required to implement the desired 250 mV threshold increased from 1.1X to 1.85X. With logic depths reducing due to heavy pipelining in high-performance designs, the amount of upsizing required in our technique is bounded and scales well into future technologies. We point out here that the width penalty can be substantially reduced if the optimization algorithm based on our approach performs continuous threshold assignment while keeping a global perspective of the width penalty. In this work, we focus on realizing a desired threshold for an individual gate and describe the penalties associated with it. How to use this methodology optimally in larger blocks with minimum penalty is beyond the scope of this work.

The ability to achieve intermediate threshold voltages can be very useful in delay-constrained power optimizations. It is due to that fact that the intermediate threshold capability provides a more continuous trade-off between power and delay than that obtained by two discrete thresholds. This is essentially the reason why advanced processes provide more than two threshold options. Dual threshold processes essentially provide two discrete power-delay points - if we cannot meet delay requirement by using high thresholds, then we are forced to use low threshold devices even though these devices may be significantly faster than required by the application. The continuous threshold option allows us to trade delay with power in a continuous

manner, thereby allowing more savings in power while still meeting the delay requirement. In other words, for a delay constrained design, there exists an optimal threshold voltage that results in minimum power. The ability to achieve continuous thresholds will allow us to operate at that optimal threshold and hence it would be very useful in reducing power.

## 5. CONCLUSIONS

We describe a simple technique to achieve arbitrary intermediate thresholds in a dual  $V_T$  process. The methodology is based on the observation that power-delay behavior of an intermediate threshold voltage can be achieved through a combination of low and high thresholds. We develop an analytical framework to realize intermediate thresholds and verify it against HSPICE simulations. Our results demonstrate good accuracy against hypothetical case that assumes that all desired threshold voltages are provided by the technology. The ability to achieve intermediate threshold voltages can be very effective in reducing leakage power as it provides a more continuous trade-off between power and delay than the one obtained with two discrete thresholds.

## REFERENCES

- [1] International Technology Roadmap for Semiconductors, 2003.
- [2] G. Sery, S. Borkar, and V. De, "Life is CMOS: Why Chase the Life After," *Design Automation Conf.*, pp. 78-83, 2002.
- [3] J. Tschanz *et al*, "Design Optimizations of a High-Performance Microprocessor using Combinations of Dual-Vt Allocation and Transistor Sizing," *Intl. Symp. VLSI Circuits*, pp. 218-219, 2002.
- [4] L. Wei, J. Chen, K. Roy, M.C. Johnson, Y. Ye and V. De, "Design and Optimization of Dual-Threshold Circuits for Low-Voltage Low-Power Applications," *IEEE Trans on VLSI Systems*, vol. 7, pp. 16-24, March 1999.
- [5] X. Huang *et al*, "Sub-50nm P-Channel FinFET," *IEEE Trans on Electron Devices*, vol. 48, pp. 880-886, 2001.
- [6] S. Cristoloveanu, F. Allibert and A. Zaslavsky, "Double-gate MOSFETs: Performance and Technology Options," *Intl Semiconductor Device Research Symposium*, pp. 459-460, 2001.
- [7] A. Keshvari *et al*, "Effectiveness of Reverse Body-Bias for Leakage Control in Scaled Dual Vt CMOS ICs," *Intl. Symp. Low Power Electronics Design*, pp. 207-212, 2001.
- [8] S. Thompson, I. Young, J. Greason and M. Bohr, "Dual Threshold Voltages and Substrate Bias: Keys to High-Performance Low-Power 0.1um Logic Designs," *Intl. Symp. VLSI Technology*, pp. 69-70, 1997.
- [9] J. Tschanz *et al*, "Dynamic Sleep Transistor and Body Bias for Active Leakage Control of Microprocessors," *IEEE Journal of Solid State Circuits*, vol. 38, pp. 1838-1844, Nov 2003.
- [10] P. Pant, K. Roy and A. Chatterjee, "Dual-threshold Voltage Assignment with Transistor Sizing for Low-Power CMOS Circuits," *IEEE Trans. on VLSI*, vol. 9, pp. 390-394, April 2001.
- [11] M. Ketkar and S. Sapatnekar, "Standby Power Optimization via Transistor Sizing and Dual Threshold Voltage Assignment," *Intl. Conf. Computer-Aided Design*, pp.375-378, 2002.
- [12] T. Karnik *et al* "Total Power Optimization by Simultaneous Dual-Vt Allocation and Device Sizing in High-Performance Microprocessors", *Design Automation Conf.*, pp. 486-491, 2002.
- [13] N. Weste and K. Eshraghian, "Principles of CMOS VLSI Design," Addison-Wesley, 1993
- [14] T. Sakurai and A.R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584-594, Apr. 1990.