Optimal Power Switch Design for Dynamic Voltage Scaling from High Performance to Subthreshold Operation

Kyle Craig, Yousef Shakhsheer, and Benton H. Calhoun
Dept. of Electrical and Computer Engineering
University of Virginia, Charlottesville, VA, USA

ABSTRACT
This work explores optimizing power switch design for Dynamic Voltage Scaling schemes that use headers to connect components to voltage supplies ranging from strong inversion to subthreshold values. We propose using NMOS devices with their gate controlled at the nominal voltage as power switches connected to the subthreshold voltage rail. Measured results show that an NMOS can provide the subthreshold voltage with a power switch size *>280X* smaller than a PMOS. For architectures targeting operation from subthreshold up to nominal voltage, we show that using an asymmetric transmission gate power switch provides a lower overhead way to enable this flexibility.

Categories and Subject Descriptors
1.1 Technologies and Digital Circuits

Keywords
Power gate, Low Power, Dynamic Voltage Scaling

1. INTRODUCTION
With continued technology scaling and the increase of portability in battery and battery-less electronic devices, energy efficiency has become arguably the most critical metric facing circuit designers. In order to address this metric, there exists extensive research into how to best improve energy efficiency. Some common techniques include dynamic voltage scaling (DVS), power gating for leakage reduction, and subthreshold circuit operation. Since dynamic energy is quadratically proportional to V_{DD} and frequency is proportional to V_{DD}, DVS reduces energy by dynamically reducing voltage when performance requirements allow. Power gating reduces leakage energy by using power switches to cut the circuit off from the supply when idling. Finally, in subthreshold operation the circuit operates below the threshold voltage (V_T). In subthreshold, frequency is exponentially dependent on V_{DD}.

DVS is conventionally deployed at the chip or core level, often using DC-DC converters to adjust the voltage. A different approach uses header switches to connect blocks to one of several power supply voltages, allowing for faster switching and application of DVS to smaller internal blocks. A scheme called Panoptic Dynamic Voltage Scaling (PDVS) uses this approach to achieve near optimal energy efficiency from high speed to subthreshold operation [1]. The wide range DVS chip in [1] uses multiple voltage supplies at high, middle and low values (V_{DDH},V_{DDM}, V_{DDL}) with PMOS power switches and voltage dithering to select the appropriate V_{DD} for different fine grained blocks/components in the design depending on local workload requirements. Figure 1 shows the generic PDVS architecture. This scheme supports DVS, power gating for leakage reduction, and subthreshold (V_{SUBVT}) operation when appropriate, so we select it as a platform that is representative of header based DVS schemes for our analysis of header switch optimization. In the rest of the paper, we explore power switch design to enable subthreshold operation in the PDVS architecture in the most area and energy efficient manner.

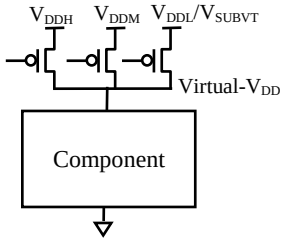

Figure 1. Architecture for using headers to implement DVS across a wide voltage range.

The paper is organized as follows: In Section 2 we will discuss current power switch sizing methodologies. In Section 3 we propose using an NMOS device as a subthreshold power switch and compare it to the conventional PMOS device. In Section 4 we propose a transmission gate based power switch for greater design flexibility. Finally, in Section 5 we discuss some physical design considerations for the proposed power switch.

2. SIZING FOR DELAY
The introduction of a power switch device between V_{DD} and a component creates an IR drop across the header resulting in a reduced virtual-V_{DD} value and performance degradation. Power switch sizing is critical to maintain low power and expected performance. An undersized power switch results in large performance degradation, however an oversized power switch results in increased leakage and increased area overhead. Power switch sizing methodologies have been examined in depth to support techniques like multi-threshold CMOS (MTCMOS), which used high V_T power switches to reduce leakage [3]. Optimal power switch design for has been extensively explored, e.g., [4]-[7]. These schemes explore ways to optimize Ion/ Ioff and develop methodologies or tools to implement these schemes. For the

ISLPED'12, July 30–August 1, 2012, Redondo Beach, California, USA.

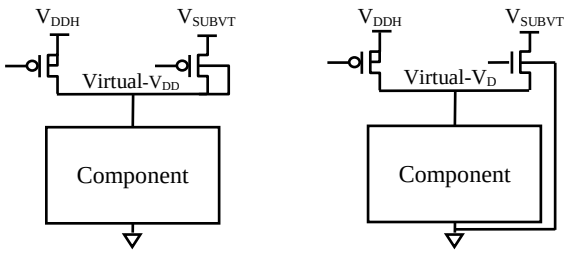

Figure 2. (*left*) Conventional PMOS subthreshold power switch (*right*) Proposed NMOS subthreshold power switch

purposes of this work, we will build from a commonly used power switch sizing methodology that sets the power switch size such that the critical path meets an acceptable delay or frequency degradation from the nominal case (i.e., no power switch). The allowable degradation is a design choice chosen by the system designer.

3. NMOS as a Subthreshold Header

In this paper we propose using an NMOS power switch to connect to a subthreshold voltage rail instead of a PMOS power switch. Figure 2 shows the convention subthreshold power switch architecture as well as the proposed NMOS power switch architecture. The conventional architecture uses a PMOS power switch with the body tied to virtual-V_{DD} to avoid reverse body bias [2]. Since the PDVS architecture has multiple supplies, the power switch control signals are full swing, up to V_{DDH}, providing strong turn off of the off headers. In the proposed alternative, an NMOS device with its body tied to ground is used as the header between the subthreshold rail and the component. During subthreshold operation (i.e., only V_{SUBVT} enabled) the conventional PMOS power switch has a $|V_{GS}| = V_{SUBVT}$, however the proposed NMOS power switch has a $|V_{GS}|=V_{DDH}-V_{SUBVT}$. The higher $|V_{GS}|$ on the NMOS device provides a much higher current than the PMOS, since the NMOS is in the linear region of operation while the PMOS is in the cutoff/subthreshold region of operation. The higher current from the NMOS device provides a more stable virtual-V_{DD} while potentially using a much smaller device, decreasing the area overhead associated with the power switch.

3.1 Comparison to Conventional

3.1.1 Comparison Circuit Setup

We used a commercial 130nm bulk process to simulate, measure, and compare the conventional and proposed subthreshold header topologies. To provide a flexible, representative load circuit, we used ten 27-stage ring oscillators (ROs) in parallel, and each RO was capable of being enabled independently. To simplify the comparison, each block of parallel ROs only had two header power switches as shown in Figure 2. In simulation, we swept the widths of the headers to examine the impact of size on header behavior, and in the test chip we describe later, we included programmable sized headers for flexible measurements.

3.1.2 Simulation Results

Figure 3 demonstrates the impact of power switch width on virtual-V_{DD} for two different activity factors. An activity factor of 1.0 corresponds to all 10 ring oscillators enabled in parallel while 0.1 corresponds to only 1 ring oscillator enabled. These two activity factors represent the upper and lower bounds in this design.

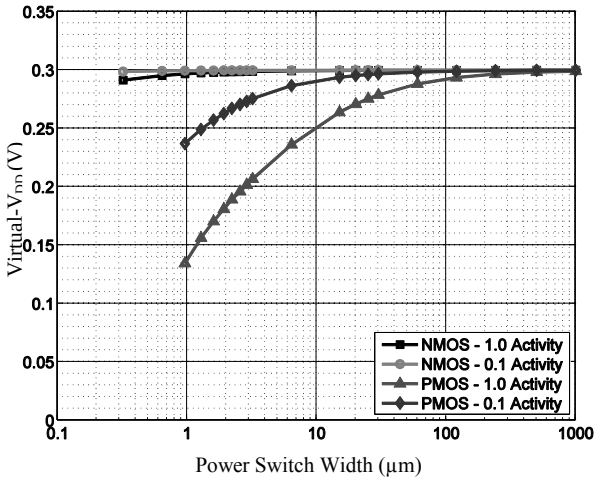

Figure 3. Simulation Virtual-V_{DD} at 0.3Vfor the conventional PMOS and proposed NMOS

V_{SUBVT} was set to 0.3V, well below the threshold voltage (V_T) in the technology. Across the wide range of sizes used, the NMOS is able to keep virtual-V_{DD} at the target 0.3V due to the NMOS being in the linear operating region. The PMOS, however, is unable to keep virtual-V_{DD} at the target 0.3V for small widths since it is in the subthreshold operating region. It is necessary to keep the virtual-V_{DD} near the target V_{DD} because frequency depends exponentially on the virtual-V_{DD} voltage in subthreshold, so voltage droop leads to huge slow downs.

The impact of the power switch width on oscillator frequency is shown in Figure 4. The frequency has been normalized to the frequency at 0.3V without power switches. With near minimum sizing at the lowest activity factor the NMOS has a worst case frequency degradation of only *3%*, while the minimum PMOS has a worst case frequency degradation of *88%*. At the highest activity factor with near minimum sizing the NMOS has a worst case frequency degradation of *16%*, while the smallest PMOS has a worst case frequency degradation of *93%*. Using the traditional sizing methodology and a target delay degradation of 10%, the required NMOS size is approximately *280X smaller* than a PMOS for the same target degradation at the same worst case activity factor, with sizes of 640nm and 180µm respectively.

The total energy per operation while operating at V_{SUBVT} is defined by the following equation:

$$E_{OP} = E_{DYN_{VSUBVT}} + E_{LEAK_{VSUBVT}} + E_{LEAK_{VDDH}} \qquad (1)$$

When we compute this energy equation, we include the overheads of the PDVS architecture associated with having multiple V_{DD}s and power switch devices. Simulated energy per operation versus power switch width for an activity of 1.0 is shown in Figure 5. For both NMOS and PMOS, the energy is normalized to the same value; the energy per operation with no power switches. The shift in energy above the nominal for each of these designs is due to overheads inherent in the PDVS architecture. Specifically, the increase in energy comes from E_{LEAK} through the off V_{DDH} power switch while the V_{SUBVT} power switch is on. This V_{DDH} leakage contributes to an energy overhead of 3% and exists whether the NMOS or PMOS power switch is used. The decrease

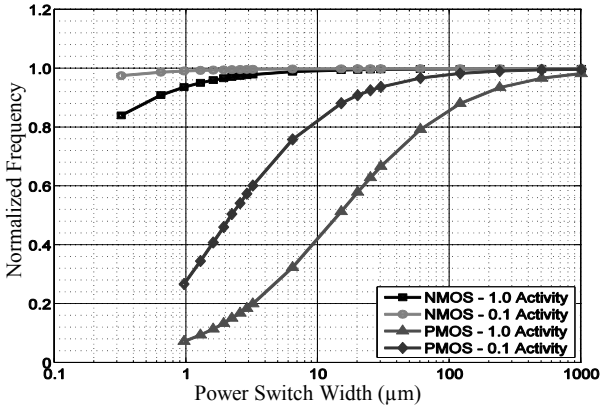

Figure 4. Simulated Frequency at 0.3Vfor the conventional PMOS and proposed NMOS

in energy as power switch size is reduced is due virtual-V_{DD} drop, which lowers the dynamic energy. However for the PMOS design, the energy starts to increase at the lower widths due leakage becoming the dominate factor because of lower operating frequencies. At these lower widths, the virtual-V_{DD} becomes much lower, resulting in initially lower dynamic and total energy as discussed, As the width is reduced further the frequency becomes so slow which results in E_{LEAK} through the V_{DDH} and V_{SUBVT} power switches becoming dominate causing an increase in total energy. The use of an NMOS power switch does not adversely increase the overheads associated with the PDVS architecture or have a higher energy per operation than a PMOS power switch. Gate leakage for this technology was not a concern; at the largest power switch size of 1mm, the gate leakage energy was only 0.04% of the total energy.

3.1.3 Measurement Results

A test chip was fabricated in a 130nm bulk commercial process to verify the simulation results. Figure 6 shows the normalized frequency with an activity of 0.1, and Figure 7 shows the normalized frequency at the activity of 1.0. The simulated data were normalized as described early, while the measured data were normalized to the largest power switch value for NMOS and PMOS at the largest power switch width possible. Even though the range of width values for the measured data is not as large as the simulated, the same trend is observed for both activity factors. At the lowest activity factor of 0.1, the NMOS has a worst case

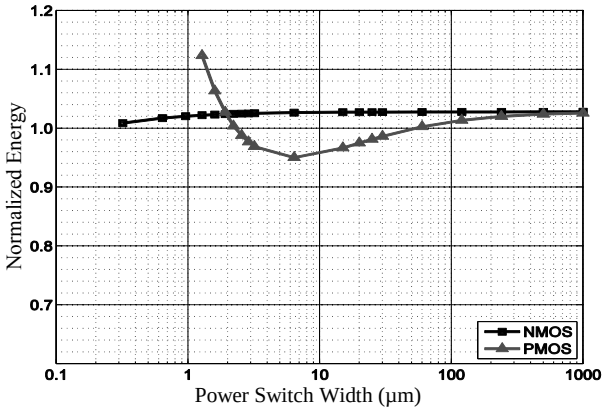

Figure 5. Normalized Energy per op vs. power switch width

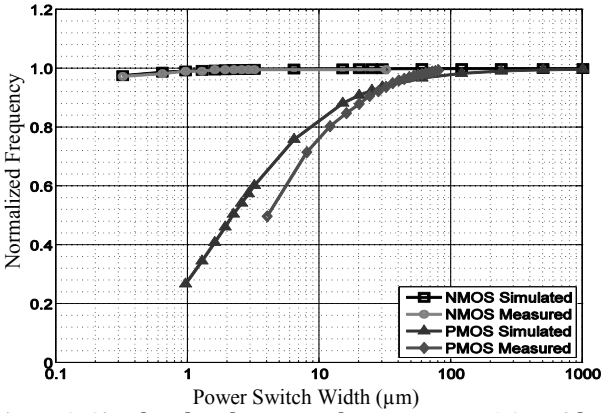

Figure 6. Simulated and Measured Frequency at 0.3V with an activity factor of 0.1

measured frequency degradation of only *3%*, while the smallest PMOS available in hardware has a worst case frequency degradation of *50%*. At the highest activity factor with near minimum sizing, the NMOS has a worst case measured frequency degradation of *13%*, while the minimum PMOS has a worst case frequency degradation of *84%*. The power switch size range in hardware was not large enough to meet the 10% frequency degradation at the highest activity for the PMOS. Using the traditional sizing methodology and a target degradation of 23% (the best achievable by the PMOS in hardware), the required NMOS size is approximately *280X smaller* than a PMOS for the same target degradation at the same worst case activity factor, with values of 320nm and 90µm respectively.

4. Flexible Transmission Gate Design

4.1.1 Motivation

If the V_{SUBVT} rail is always kept at a subthreshold voltage, using an NMOS power switch is optimal since a near minimum sized transistor would provide the target frequency requirement. However, if the rail connecting the header to the component needs to be a flexible and encompass a wide range of V_{DD}s, the NMOS fails as a power switch above V_T due to the V_T drop seen across the NMOS. For designs that require a wide range of voltages on the V_{SUBVT} rail, we propose to use a transmission gate architecture shown in Figure 8. When V_{SUBVT} is near- or subthreshold, the NMOS will be the dominate device. Conversely, when in voltages

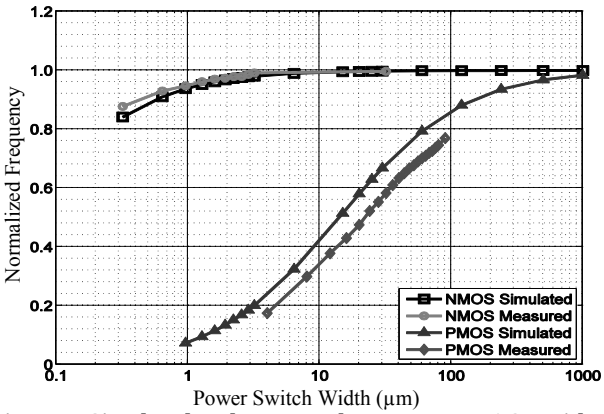

Figure 7. Simulated and Measured Frequency at 0.3V with an activity factor of 1.0

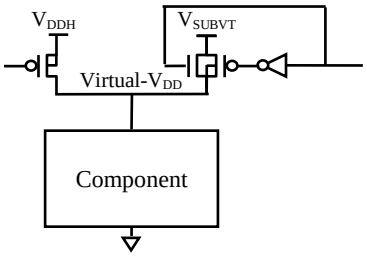

Figure 8. Proposed transmission gate power switch

above V_T, the PMOS will be the dominate device.

4.1.2 Sizing Methodology

Both PMOS and NMOS devices should be sized independently for a given frequency degradation. This will lead to the transmission gate power switch having asymmetric sizing. The PMOS device would be relatively large to meet the frequency target at high voltage; in our design it is in the 10s-100s of µm depending on the frequency target. However, the NMOS device would be in the <1 µm range and still provide the target frequency for the component with the supply at low voltages. For our test chip, the NMOS actually provides a frequency degradation target better than designed.

4.1.3 Measured Results

Measured results were taken from the test chip verifying the ability to use a transmission gate for a power switch for a wide range on the supply voltage rail. The V_{SUBVT} rail voltage was varied from 0.3V to 1.2V.The the transmission gate PMOS device was limited to a max width of 40µm resulting in a frequency degradation of 30% resulting in a PMOS size of 40µm and a much-smaller NMOS size of 320nm. Figure 9 shows the measured Energy-Delay (ED) curve for an NMOS-only, PMOS-only, and transmission gate power switch. The NMOS-only ED curve reaches a frequency limitation due to the V_T drop across the NMOS above about 0.6V. The PMOS-only curve becomes slower than the NMOS at about 0.4V due to the virtual-V_{DD} drop causing the increase in delay. Finally, the transmission gate merges the best from the NMOS and PMOS and is able to provide the lowest delay and the lowest energy. In Figure 10, the measured energy delay product (EDP) across the range of V_{SUBVT} is shown for the same three cases described above. Since EDP is a measure of energy efficiency, it is desirable to have a lower EDP. As expected, the NMOS has a lower EDP at lower V_{DD}s, but suffers from a higher EDP at higher V_{DD}s.

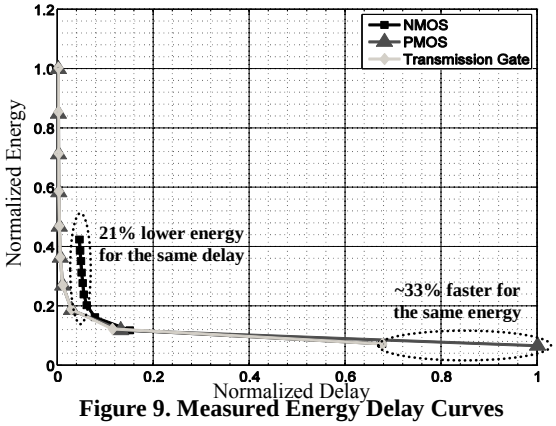

Figure 9. Measured Energy Delay Curves

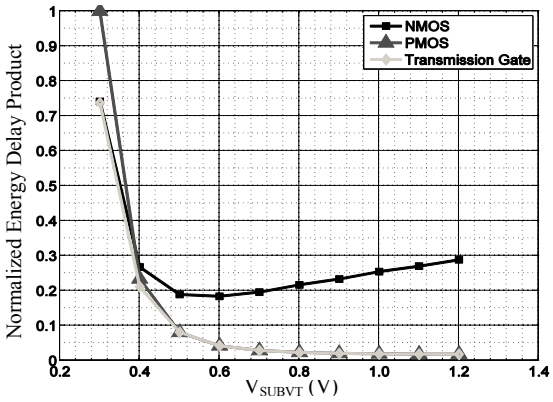

Figure 10. Measured Energy Delay Product

Conversely, the PMOS has a lower EDP at higher V_{DD}s, but suffers a higher EDP at lower V_{DD}s. The transmission gate is the pareto optimal curve of the NMOS and PMOS having the lowest EDP across all V_{DD}s.

5. Physical Design Consideration

Tradition methodologies of physical power switch design generally distribute the power switch as standard cell rows, columns, or as rings around design [6] [7]. These help the electrical properties of the power delivery system. Physical implementation of the NMOS device in these methodologies would lead to using above minimum sizing to prevent power delivery system electrical problems, potentially slightly increasing the area associated with the NMOS power switch. This would slightly alter the presented results by decreasing the NMOS area savings prevent compared to the PMOS. This would be especially true for smaller designs where a minimum NMOS could be used, but would need to be increased for the power delivery system. However, for larger designs which require larger than minimum NMOS sizing the impact of potentially needing to increase the physical size would be reduced.

6. Conclusion

We have proposed using an NMOS transistor as a power switch for low rail voltages. Simulations and measurements have shown that an NMOS device with a nominal gate swing ($0 - V_{DDH}$) is able to provide the target frequency degradation with a size over *280X* smaller than a PMOS. For flexible designs that have a wide range of V_{DD} rail values we propose using an imbalanced transmission gate header with a near minimum sized NMOS device in parallel with a large PMOS. The transmission gate will provide the targeted frequency degradation at the nominal V_{DD} and provide a better than target frequency in subthreshold with minimal additional area.

7. REFERENCES

[1] Shakhsheer, Y., et al., A 90nm data flow processor demonstrating fine grained DVS for energy efficient operation from 0.25V to 1.2V. *CICC*. Sept. 2011

[2] Calhoun, B.H., et al. Flexible Circuits and Architectures for Ultralow Power. *Proceedings of the IEEE* , vol.98, no.2, Feb. 2010

[3] Mutoh, S., et al. 1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS. *IEEE Journal of Solid-State Circuits.*

[4] Shi, K., Howard, D. Challenges in sleep transistor design and implementation in low-power designs. *DAC.* 2006

[5] Kao, J., Chandrakasan, A., Antoniadis, D. Transistor Sizing Issues And Tool For Multi-threshold Cmos Technology. *DAC.* 1997

[6] Shi, K., Howard, D. Sleep Transistor Design and Implementation - Simple Concepts Yet Challenges To Be Optimum. *VLSI-DAT.* April 2006

[7] Pakbaznia, E., Pedram, M. Coarse-Grain MTCMOS Sleep Transistor Sizing Using Delay Budgeting. *DATE 2008.* 2008