# Voltage Reduction Techniques for Portable Systems

Anantha Chandrakasan
Department of EECS,
Massachusetts Institute of Technology, Cambridge

## ABSTRACT

Supply voltage scaling to 1V and below is the key to low-power system design. Threshold voltage reduction enables aggressive supply scaling but increases leakage power. Emerging technologies such as MTCMOS and variable threshold Bulk/SOI will be essential in controlling leakage while achieving high performance levels at low supply voltages. Power can also be reduced by adaptively varying the supply voltage in applications where the computational workload varies with time. Aggressive voltage and power level scaling requires efficient DC-DC conversion circuitry and in some cases, it is necessary to embed this function in the processor.

## 1. Introduction

There are three main sources of power consumption in digital CMOS circuits: switching power, short circuit power, and leakage power. In conventional process technology, the switching component dominates and is given by $\alpha \cdot C_L \cdot V_{DD} \cdot \Delta V \cdot f_{clk}$ where $\alpha$ is the node transition activity factor, $C_L$ is load capacitance, $V_{DD}$ is the supply voltage, $\Delta V$ is the voltage swing, and $f_{clk}$ is the clock frequency. A variety of techniques have been proposed to reduce switching activity ranging from low level logic restructuring and power down techniques, to architectural restructuring and selection of data representation [1]. It is important to consider transitions that are fundamentally required to perform a given operation as well as spurious transitions arising from imbalances in the signal path.

The most efficient approach to lower energy consumption is to operate at the lowest possible power supply voltage. The individual circuit elements, however, run slower at lower supply voltages and circuit performance degrades. One approach to maintain throughput at reduced voltages is to use parallel architectures to compensate for increased gate delays [1]. Significant power reduction over conventional approaches is possible at the cost of increased silicon area. A variety of voltage scaling strategies are described here which reduce power without significantly increasing silicon area. Associated with aggressive voltage and power scaling is the need for high-efficiency regulation techniques. In many cases, embedding the power converter control in the processor can significantly reduce power dissipation.

The short circuit component of power can be kept to less than 10% of the total power through proper transistor sizing. The third component is the leakage power, resulting from reverse biased diode conduction and subthreshold operation. Scaling supply voltages below 1V requires the scaling of the threshold voltage, which unfortunately comes at the cost of increased leakage. While leakage is typically negligible when circuits are active, it can be significant during idle mode. A variety of technology and circuit solutions are discussed here which address the conflicting requirements of high-performance during active periods (favoring low thresholds) and low leakage during standby (favoring high thresholds).

## 2. Embedded Power Supply Systems

There are many systems where the amount of computation per input sample (henceforth called workload) varies with time. For example, in differential video compression algorithms, the amount of computation performed depends highly on the frame to frame image correlation. Typically each frame is divided into small blocks and only those blocks that change more than some threshold are processed. A scene change requires the entire frame to be coded while a highly correlated sequence requires only a few blocks to be coded. The wide variation in workload can be exploited to save power by shutting down the processor using gated clocks when the workload is less than the peak; this results in a linear reduction in power dissipation.

However, in such variable workload systems, power can be further reduced if a variable power supply is used in conjunction with a variable switching speed processor. Circuits with a fixed supply voltage work at a fixed speed and idle if the data sample requires less than the peak computation. Less power would be consumed if supply voltage were lowered when the workload decreased. Figure 1 shows a plot of power vs. normalized workload ($V_{DD(max)}$=2V and $V_T$=0.4) [2]. The straight line represents a fixed power supply system. If the workload is less than peak workload (i.e., *workload=1*), then the processor would compute as fast as possible at a fixed voltage and idle for the remaining fraction *(1-workload)* of the sample period. Power is reduced in a linear fashion since the energy per operation is fixed, and only the number of operations changes.

1063-0988/97/$10.00 © 1997 IEEE.

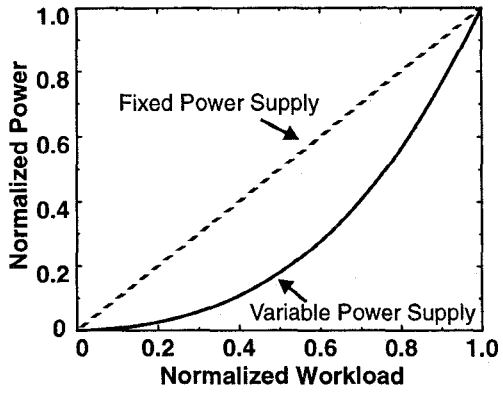


**Figure 1.** Power reduction using variable supply.

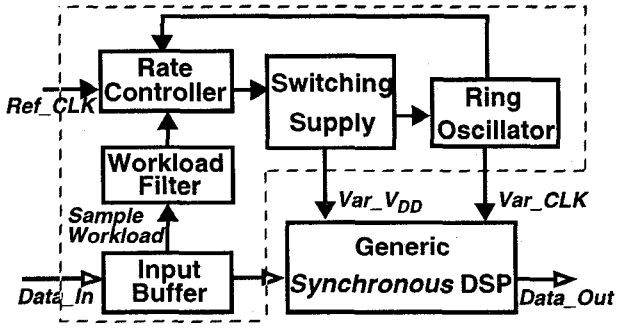


**Figure 2.** Variable supply voltage block diagram [2].

The lower curve of Figure 1 represents the case for a variable power supply system. If the workload for a given sample period is less than peak, then the delay of the processing element can be increased by a factor of *1/workload* without loss in throughput, allowing the processor to operate at a lower power supply voltage. The energy per sample therefore varies not only because fewer operations are performed, but also because each operation consumes less energy.

Self-timed circuits have been proposed to exploit variable processing times to dynamically vary the supply voltage [3]. Self-timed circuits implicitly account for variations in bit-level data dependent delay, algorithmic workload, temperature, and process parameters. Unfortunately, implementation requires the use of dual-rail coding, which comes at the cost of high switching activity. An alternative approach for implementing a variable supply system is to use standard *synchronous* design techniques with a ring-oscillator based clock and switching supply [2]. While this involves giving up the ability to exploit bit-level data dependent delay variations (e.g., data dependent variable ripple length of an adder), algorithmic workload variations can be exploited.

The supply voltage is adjusted by changing the duty cycle of the switching supply based on the workload derived from the input signal. The clock frequency is controlled by a digital phase-locked loop with the variable

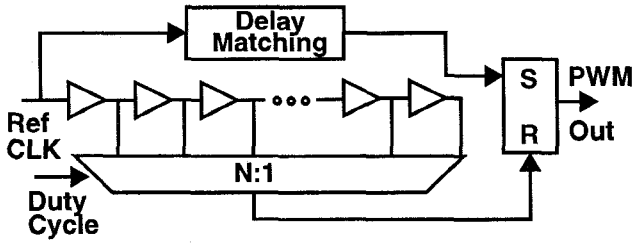


**Figure 3.** Delay Line based PWM generator [6].

$V_{DD}$ as the control signal for the ring-oscillator (Figure 2). A lookup table (in the rate controller) records the voltage needed for each processing rate, and the feedback slowly adjusts the entries in the table as needed. Thus the loop tracks process and temperature, but computation rate is changed open-loop by reading table entries.Even when workload is fixed, the supply can be minimized by tracking process and temperature variations [4].

In order to achieve high efficiencies in the embedded power supply systems described above, the fixed control overhead power should be minimized. A key circuit element is the PWM generator, which produces a square wave of desired duty cycle. Typically, in digital systems PWM signals are created by using a clock at some multiple of the switching frequency with a counter. The PWM signal is set high at the beginning of a switching period, and then reset after the counter detects that some number of cycles of the faster clock have passed. This approach has been applied to digital power supply controllers [5].

Ultra-fast-clocked counters are not particularly well suited for low power operation. The clock frequency at which the counter operates can be extremely high (e.g., a 1MHz switching waveform with 256 discrete levels of duty cycle requires a 256Mhz clock!). As a result of the short delay requirement, such a circuit does not lend itself to voltage scaling. A second approach to transforming a digital word to a pulse width modulated signal is to use a tapped delay line [6]. The delay line (Figure 3) can be constructed with current starved buffers, and controlled with a feedback loop so that the delay of the delay line matches the period of the switching clock. The number of taps on the delay line is chosen to be the number of desired discrete duty cycle levels. The taps of the delay line are then addressed with a multiplexer. The PWM signal is set high as a pulse is input to one end of the delay line, and reset when that pulse arrives at the selected tap. Lacking any need for short delays, the tapped delay line PWM can tolerate deep voltage scaling. Such a PWM circuit has been fabricated, and demonstrates a power dissipation on the order of tens of microwatts [6].

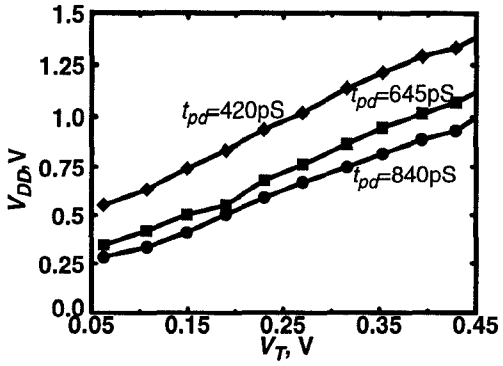


**Figure 4:** Experimental $V_{DD}$ vs. $V_T$ for a fixed delay.

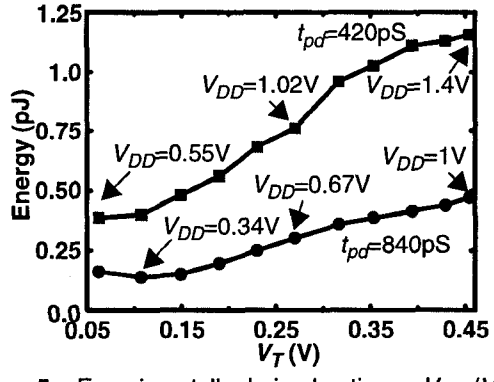


**Figure 5.** Experimentally derived optimum $V_{DD}/V_T$ point using the dual-gate SOI technology described in [7].

## 3. Threshold Voltage Scaling

Reducing the threshold voltage ($V_T$) of the devices allows the supply voltage to be scaled down without loss in performance. Figure 4 shows an experimentally obtained plot of $V_{DD}$ vs. $V_T$ while keeping performance (gate delay) constant. These experimental plots were obtained from ring oscillator structures by adjusting the $V_T$ and $V_{DD}$ for a fixed delay. Figure 5 shows a plot of energy vs. $V_T$ for two different speeds of a 101 stage ring oscillator (i.e., all points on each curve have a fixed delay). Here, the power supply voltage is allowed to vary to keep the performance fixed.

For a fixed delay, the supply voltage and therefore the switching component of power can be reduced while reducing the threshold voltage. However, at some point, the threshold voltage and supply reduction is offset by an increase in the leakage current, resulting in an optimum threshold voltage and power supply voltage. That is, the optimum threshold voltage must compromise between lower switching energy and higher sub-threshold leakage energy at low supply voltages. It is interesting to note that the optimum supply voltage is significantly lower than 1V, even with a very low activity factor.

## 4. Technologies for Burst Mode Applications

For continuously operated circuits, the increased leakage due to reduced threshold voltages is typically not significant (as shown in the previous section). However, for "event-driven" computation in which intermittent computation activity triggered by external events is separated by long periods of inactivity (e.g., X-server), leakage power is critical. To satisfy the contradicting requirements of high-performance during active periods and low-standby leakage, several device technologies and circuit styles have been developed. This includes the dynamic control of threshold voltages in triple-well CMOS using backgate effect, Multiple Threshold CMOS, variable threshold CMOS and dual-gated SOI technology. All these technologies provide a knob to reduce leakage power during idle periods while enabling low-threshold high-speed operation during active periods.

In Multiple Threshold CMOS process (MTCMOS), the logic circuits are implemented using low threshold devices and the low-$V_T$ transistors are "gated" using series connected high threshold switches [8]. During idle periods, the high threshold devices are cut-off, significantly reducing the subthreshold leakage. During active periods, the high threshold switches are turned on and circuits resume normal low threshold high speed operation, provided that the devices are sized properly.

A key challenge in designing with MTCMOS is transistor sizing of high $V_T$ devices which is very strongly dependent on data dependent current profiles [9]. For example, consider an 8x8 bit carry save multiplier. For simplicity, Figure 6 shows only a 4x4 version with a worst case delay path highlighted. It is easy to see that one critical path (many others exist) lies along the diagonal and bottom row. However, two distinct input vectors that give the same delay in a CMOS implementation can give very different results in an MTCMOS circuit. The transition

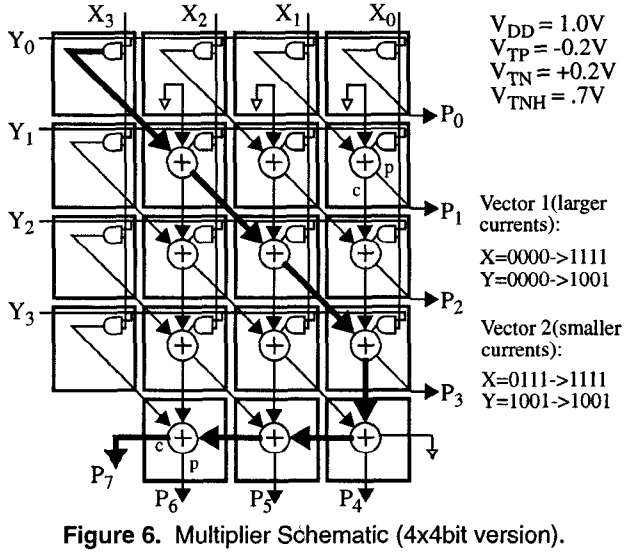


**Figure 6.** Multiplier Schematic (4x4bit version).

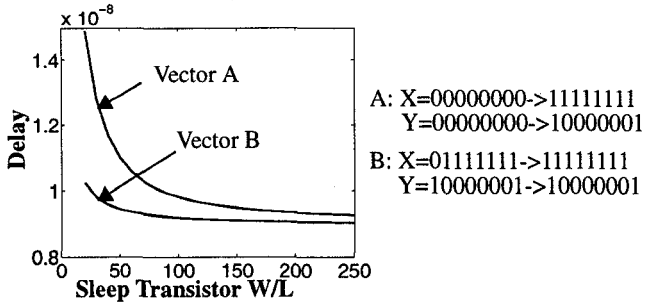


**Figure 7.** Multiplier delay vs. W/L for different vectors [9].

from (x:00,y:00) -> (x:FF,y:81) for example causes many more internal transitions in adjacent cells and thus is more susceptible to ground bounce than the (x:7F, y:81) -> (x:FF, y:81) transition. The second input causes a rippling effect through the multiplier, where only a few blocks are discharging current at the same time. Figure 7 shows how delay varies with the W/L ratio of the sleep transistor for these two cases.

Another proposal for idle mode leakage control involves changing the threshold voltage dynamically by changing the substrate bias - that is, use a lower threshold voltage to operate at low power supply voltages during active periods and raise the threshold voltage during idle periods to lower leakage power [10]. One potential disadvantage is that the threshold voltage changes in a square root fashion with respect to source to bulk voltage and therefore a large voltage may be required to change $V_T$ by a few hundred mV. A high body factor is required which results in increased parasitic capacitance.

A third approach is to use a silicon-on-insulator-with-active-substrate (SOIAS), to achieve dynamically variable threshold voltages [7]. Figure 8 shows the cross section of the SOIAS device. The back gate controls the threshold voltage ($V_T$) of the front gate device since the surface potentials at the front and back interfaces are coupled in fully depleted SOI devices. Leakage power and on-currents are controlled by lowering $V_T$ when a circuit is active and raising $V_T$ when the circuit is idle. This addresses the opposing requirements of high performance and low power, particularly at low power supply voltages. A 250mV change in threshold voltage results in a 3.5-4 decade reduction in off current and an 80% switching current increase at 1V operation [7].

## 5. Conclusion

Extremely low power operation can be achieved for digital circuits by aggressively reducing the supply voltage. New technologies can be exploited to enhance the efficiency of low supply voltage circuits, such as MTCMOS and variable threshold bulk-CMOS/SOI. Variable supply voltage circuits can be used to increase power savings for variable workload applications. Low voltage and low power systems also require designers to investi-

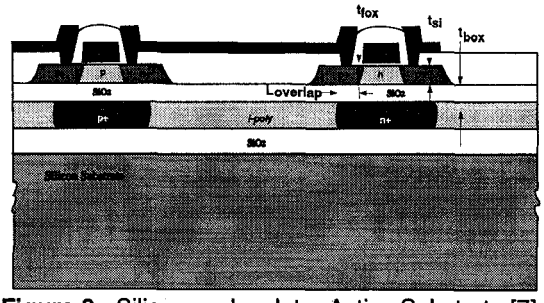


**Figure 8:** Silicon-on-Insulator Active Substrate [7].

gate power converter circuits to create low supply voltages and variable supply voltages efficiently.

### Acknowledgments

This work is sponsored by DARPA, under contracts #DAAL-01-95-K3526 and #DABT63-95-C-0088. The author would like to thank A. Dancy, V.Gutnik, J. Kao, I. Yang and Prof. D. Antoniadis for their contributions to this paper.

### References

[1] A. Chandrakasan, R. Brodersen, *Low Power Digital CMOS Design*, Kluwer Academic Publishers, July 1995.

[2] V. Gutnik, A. Chandrakasan, "An Efficient Controller for Variable Supply-voltage Low Power Processing," *IEEE VLSI Circuits Symposium*, pp. 158-159, June 1996.

[3] L. Nielsen, C. Niessen, J. Sparso, K. van Berkel, "Low-Power Operation Using Self-Timed Circuits and Adaptive Scaling of Supply Voltage," *IEEE Transactions on VLSI systems*, pp 391-397, December 1994.

[4] P. Maken, M. Degrauwe, M. Van Paemel, H. Oguey, "A Voltage Reduction Technique for Digital Systems," *IEEE ISSCC*, pp. 238-239, February 1990.

[5] G. Wei and M. Horowitz, "A Low Power Switching Power Supply for Self-Clocked Systems," *1996 International Symposium on Low Power Electronics and Design*, pp. 313-318, 1996.

[6] A. Dancy, A. P. Chandrakasan, "Ultra Low Power Control Circuits for PWM Converters," *IEEE Power Electronics Specialists Conference*, pp. 21-27, June 1997.

[7] I. Yang, C. Vieri, A. Chandrakasan, and D. Antoniadis, "Back Gated CMOS on SOIAS for Dynamic Threshold Control," *IEEE 1995 IEDM*, December 1995.

[8] M. Mutoh, T. Douskei, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V Power Supply High-Speed Digital Circuit Technology with Multi-threshold Voltage CMOS," *IEEE JSSC*, pp. 847-854, August 1995.

[9] J. Kao, A. Chandrakasan, D. Antoniadis, "Transistor Sizing Issues and Tool For Multi-Threshold CMOS Technology", *IEEE/ACM Design Automation Conference*, pp.409-414, June 1997.

[10] T. Kuroda et. al., "A 0.9V 150MHz 10mW 4mm$^2$ 2-D discrete cosine transform core processor with variable threshold-voltage ($V_T$) scheme," *IEEE JSSC*, pp. 1770-1779, November 1996.