

# The Impact of Transistor Sizing on Power Efficiency in Submicron CMOS Circuits

Robert Rogenmoser, Hubert Kaeslin, and Norbert Felber

Integrated Systems Laboratory  
Swiss Federal Institute of Technology  
CH-8092 Zürich, Switzerland

## Abstract

*Transistor size optimization is one method to reduce the power dissipation of CMOS VLSI circuits. Analysis shows that parasitic capacitances and velocity saturation of submicron technologies favor wider than minimum transistor sizes. The reason is that they allow for a larger reduction of the supply voltage which results in more substantial power savings. Spice simulation of prescalers with differently scaled transistors confirm the analysis. The same prescaler has been implemented in a 1.0  $\mu\text{m}$  technology with minimum sized and with optimized transistors for high speed. Measurements confirm that power dissipation is reduced for optimized transistor sizes.*

## 1 Introduction

A general concern in VLSI design is power efficiency. It is indeed very obvious that battery operated equipment, such as hand-held cellular phones, laptop computers, etc. impose stringent limits on the acceptable power dissipation. It is perhaps less obvious that power efficiency is also of interest when the power supply is from the mains. The reason for this is the removal of heat from high-performance high-density VLSI circuits and boards, such as CPU's, workstations, and broadband telecommunication equipment, for instance.

The power dissipation of CMOS circuits is determined by decisions at different levels. On the system/architecture level, pipelining, replication, retiming, and bit-serial operation can result in power savings [1, 2]. Special number systems may reduce power dissipation on the algorithm level. Moreover, new technologies with smaller feature sizes and lower supply voltages contribute to lowering power dissipation. However, by optimizing transistor sizes, the power dissipation can be further reduced.

It is generally believed that low power designs need to have minimum size transistors. While this is true for long channel devices, parasitic capacitances become more important in today's submicron CMOS technologies. In addition, the channel length is reduced to a degree where velocity saturation occurs, changing first order MOS equations. Considering these facts, transistor sizes can be increased while lowering the supply voltages, which results in a reduced total power dissipation.

In this work we show how transistors can be sized to lower the power dissipation of CMOS circuits. As a circuit example, we will look at a simple asynchronous prescaler composed of four toggle-flip-flops [3], a circuit that can neither be replicated, pipelined, retimed or otherwise transformed. In the next section we will look at the analytical calculation of power dissipation for long and short channel CMOS technologies. Then we will compare SPICE simulation results for prescalers implemented with various transistor sizes before we present experimental results of a minimum sized prescaler compared to the same circuit optimized for high speed.

## 2 Analytical Power Calculation of CMOS Circuits

### 2.1 Long Channel Devices

For this analysis, we take as example an inverter driving another inverter of the same size. To charge the input of the second inverter, the first one has to deliver a current of  $I = C \cdot \frac{dV}{dt}$ . Assuming linear ramps, the duration of this charging process can be calculated as  $t = C \cdot \frac{V}{I} (\propto \frac{1}{f})$ . The current in the transistor of the inverter can be approximated by using the Sah-model for saturation (with a full step at the input) as  $I = \frac{KP}{2} \cdot \frac{W}{L} \cdot (V - V_t)^2$  (1).

The capacitance  $C$  which has to be charged by the first inverter is composed of the gate capacitances of the second inverter  $C_{gate}$ , the interconnect capacitance  $C_{int}$ , and the output capacitance of the first inverter  $C_{diff}$  ( $C = C_{gate} + C_{diff} + C_{int}$ ). If we take  $C_o$  as the input capacitance of a minimum sized inverter and  $d$  as the relative contribution from the diffusion areas we can now calculate the capacitance of a unit sized inverter as:  $C_U = C_o + d \cdot C_o + C_{int}$  with:  $C_{int} = w \cdot C_o$  ( $w$  is the wiring factor). For an inverter whose transistor widths are sized with  $n$  we get  $C_S = n \cdot C_o + n \cdot d \cdot C_o + C_{int}$ .

As a constraint, we demand that the speed of both circuits should be the same. Therefore we can now reduce the supply voltage of the scaled circuit  $V_S$  so that  $f_S = f_U$ :

$$\frac{KP \cdot W}{2 \cdot C_U \cdot V_U \cdot L} \cdot (V_U - V_t)^2 = \frac{KP \cdot n \cdot W}{2 \cdot n \cdot C_S \cdot V_S \cdot L} \cdot (V_S - V_t)^2$$

This results in:

$$V_S = V_t + \frac{(w + n \cdot (1 + d)) \cdot \sqrt{\frac{n \cdot (V_U - V_t)^2}{(w + n \cdot (1 + d)) \cdot (w + 1 + d)}}}{n}, \quad (2)$$

and with  $P \propto C \cdot f \cdot V^2$  the relative power dissipation can be calculated as:

$$\frac{P_S}{P_U} = \frac{(w + n \cdot (1 + d))}{n \cdot (w + 1 + d) \cdot V_U} \cdot \left( n \cdot V_t + (w + n \cdot (1 + d)) \cdot \sqrt{\frac{n \cdot (V_U - V_t)^2}{(w + n \cdot (1 + d)) \cdot (w + 1 + d)}} \right) \quad (3)$$

These equations correspond to the results in [1]. The dashed lines in Fig. 1 represent (2) and (3) for  $d = 0.25$  and  $w = 0, \dots, 5$ .

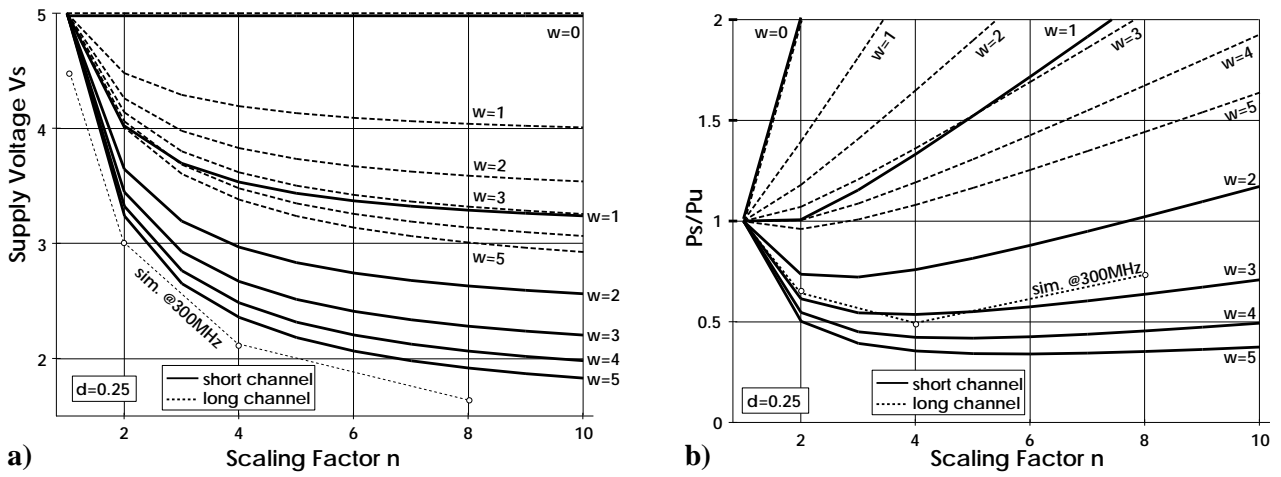


Fig. 1: a) Supply voltage reduction and b) relative power dissipation reduction of scaled circuits

## 2.2 Submicron Devices – Velocity Saturation

By shortening the channels of transistors, the carriers reach velocity saturation due to the high electrical field. The critical electrical field for velocity saturation to occur is around  $E_{crit} \sim 10^6$  V/m. For  $1 \mu\text{m}$  CMOS transistors with effective channel length of  $0.85 \mu\text{m}$  this translates to a Drain-Source voltage of  $0.85$  V above which velocity saturation occurs. Equation (1) then can be approximated as  $I = K \cdot \frac{W}{L} \cdot (V - V_t)$  (4) [4].

Equations (2) and (3) change to:

$$V_S = V_t + (V_U - V_t) \cdot \frac{(w + n \cdot (1 + d))}{n \cdot (w + 1 + d)} \quad (5)$$

$$\frac{P_S}{P_U} = \frac{(V_U \cdot (w + n \cdot (1 + d)) - w \cdot V_t \cdot (1 - n))^2 \cdot (w + n \cdot (1 + d))}{V_U^2 \cdot (w + 1 + d)^3 \cdot n^2} \quad (6)$$

The solid lines in Fig. 1 depict (5) and (6). By simply scaling all transistor sizes it is possible to reduce the supply voltage to even lower levels, which results in a larger decrease of power dissipation.

## 3 Simulation Based Analysis of Different Scaled Prescalers

The analysis of the previous section was verified using an asynchronous prescaler as a circuit example [3]. This prescaler, consisting of four identical toggle flip-flops, was implemented with minimum sized transistors ( $W_{PFET} = 2 \cdot W_{NFET} = W_{min}$ ), and with all transistors sized (W/L) by a constant factor  $n=2,4,8$ . The resulting layout (in a  $1.0 \mu\text{m}$  technology) was then used for extraction of all parasitic capacitances and simulation using HSPICE (MOS model level 28, BSIM). For each circuit and input frequency, the supply voltage was lowered until the circuit failed to operate (divide) correctly. Figure 2a shows the results. The optimum scaling factor for this circuit and technology was found to be 4, reaching a power dissipation 2 to 4 times lower than the minimum sized circuit.\*

## 4 Experimental Results of Minimum Sized vs. Optimized Prescaler

The same asynchronous prescaler has been implemented and fabricated in a  $1.0 \mu\text{m}$  CMOS technology. Two different versions of that circuit have been designed: the first one uses minimum sized transistors

\*The power required for the clock input is not included. However, as its contribution varies between 13 to 18 % of  $P_S$  and  $P_U$ , it has negligible impact on our analysis

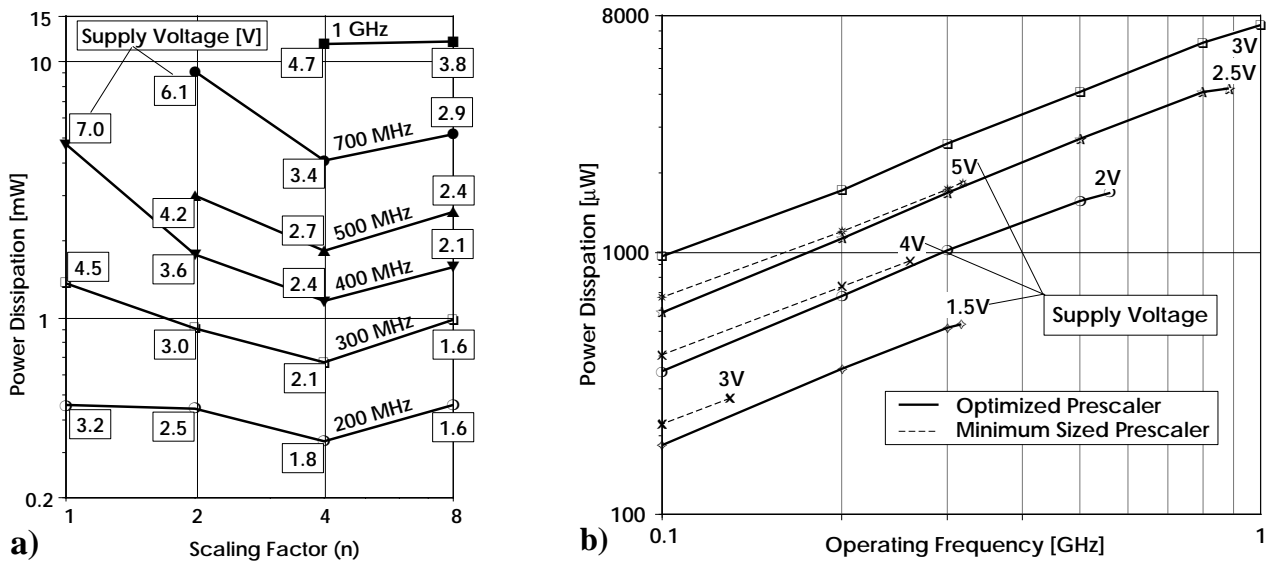


Fig. 2: Power dissipation of prescalers at different operating frequencies. **a)** Simulation with different scaling factors **b)** Measurements of minimum and high-speed prescalers

while the second one has large transistors individually optimized for high speed. Both circuits have been integrated side-by-side on the same wafer. Figure 2b shows the measurements of these prescalers. At any frequency, the high-speed prescaler can be operated with a lower power dissipation than the minimum sized prescaler. At an operating frequency of 300 MHz, for example, the high-speed version can be operated at 1.5 V supply voltage, dissipating  $575 \mu\text{W}$ , while the minimum sized version has to be operated at 5 V supply voltage and dissipates  $1740 \mu\text{W}$ , which is 3 times as much. At 200 MHz these numbers change to 1.5 V/ $359 \mu\text{W}$  for the high speed version and to 4 V/ $740 \mu\text{W}$  for the minimum version, which is still a factor of 2.

## 5 Conclusions

Simple models, simulation, and measurement have shown that substantial power savings can be attained by transistor size optimization for circuits having short channel devices. The reasons are the relatively larger parasitic capacitances and the channel velocity saturation effect of submicron CMOS technologies. We recommend a two-step procedure: In a first step, all transistors are scaled by a uniform factor. Even larger savings are then obtained in a second step where transistors are sized individually. Clearly, those transistors which are not part of the critical path can be left minimal.

## References

- [1] A. P. Chandrakasan and R. W. Brodersen, "Minimizing Power Consumption in Digital CMOS Circuits", *Proceedings of the IEEE*, vol. 83, no. 4, pp. 498–523, Apr. 1995.
- [2] R. F. Lyon, "Cost, Power, and Parallelism in Speech Signal Processing", in *Proceedings of the IEEE 1993 Custom Integrated Circuits Conference 1993*, May 1993, pp. 15.1.1–15.1.9.
- [3] Q. Huang and R. Rogenmoser, "Speed Optimization of Edge-Triggered CMOS Circuits for Gigahertz Single-Phase Clocks", *IEEE Journal of Solid-State Circuits*, vol. 31, no. 3, Mar. 1996.
- [4] K. R. Laker and W. M. C. Sansen, *Design of analog integrated circuits and systems*, McGraw-Hill, Singapore, 1994.