

CMOS Scaling into the Nanometer Regime

YUAN TAUR, SENIOR MEMBER, IEEE, DOUGLAS A. BUCHANAN, MEMBER, IEEE,
WEI CHEN, MEMBER, IEEE, DAVID J. FRANK, MEMBER, IEEE, KHALID E. ISMAIL,
SHIH-HSIEN LO, GEORGE A. SAI-HALASZ, FELLOW, IEEE, RAMAN G. VISWANATHAN,
HSING-JEN C. WANN, SHALOM J. WIND, MEMBER, IEEE, AND HON-SUM WONG

Invited Paper

Starting with a brief review on 0.1- μm (100 nm) CMOS status, this paper addresses the key challenges in further scaling of CMOS technology into the nanometer (sub-100 nm) regime in light of fundamental physical effects and practical considerations. Among the issues discussed are: lithography, power supply and threshold voltage, short-channel effect, gate oxide, high-field effects, dopant number fluctuations, and interconnect delays. The last part of the paper discusses several alternative or unconventional device structures, including silicon-on-insulator (SOI), SiGe MOSFET's, low-temperature CMOS, and double-gate MOSFET's, which may take us to the outermost limits of silicon scaling.

Keywords—CMOS integrated circuits, integrated circuits, MOSFET logic devices, MOSFET's, transistors, very-large-scale integration.

I. INTRODUCTION

After more than two decades of relentless scaling to ever smaller dimensions for higher packing density, faster circuit speed, and lower power dissipation, CMOS technology has become the prevailing technology for very large scale integration (VLSI) applications today. These advances led to computers and networks with far superior performance and dramatically reduced cost per function. Currently, 0.35- μm CMOS technology with 0.25- μm channel length is being used in the manufacturing of 64-Mb DRAM's and 200-MHz microprocessors with the number of transistors per chip in the 10^7 – 10^8 range. Meanwhile, CMOS devices of 0.1- μm (100 nm) channel length have recently been fabricated in research laboratories. Fig. 1 shows the trends of power supply voltage (V_{dd}), threshold voltage (V_t), and gate oxide thickness (t_{ox}) versus channel length (L) for CMOS logic technologies. A key question at this point is: are we approaching the limit of silicon scaling? This paper addresses the challenges in further scaling of MOSFET's

into the sub-100-nm regime in light of fundamental physical effects and practical considerations.

Section II reviews the current status of 0.1- μm (100 nm) CMOS devices in terms of their performance. Section III examines key issues in scaling bulk CMOS below 100 nm. These include: lithography, power supply and threshold voltage, short-channel effect, gate oxide, high-field effects, dopant number fluctuations, and interconnect delays. Section IV discusses several alternative or unconventional device structures that may take us to the outermost limit of silicon scaling. They are: silicon-on-insulator (SOI), SiGe MOSFET, low-temperature CMOS, and double-gate MOSFET. Section V concludes the paper.

II. 0.1- μm CMOS RESULTS

Fig. 2 shows a transmission-electron micrograph (TEM) cross section of a 0.1- μm gate MOSFET with 30-Å-thick gate oxide. This device operates at a power supply voltage of 1.5 V for high-performance logic application [1]. The measured saturation transconductances versus the effective channel length are shown in Fig. 3, where the highest values are 640 mS/mm for nMOSFET and 360 mS/mm for pMOSFET, respectively. The effective channel lengths are extracted from a series of low-drain-bias $I_{\text{ds}}-V_g$ curves using the "Shift and Ratio" method [2].

The ac performance of the 0.1- μm CMOS devices is shown in Fig. 4, where the measured unity-current-gain frequencies (f_T) are plotted versus channel length. The highest f_T 's obtained are 118 GHz for nMOSFET and 67 GHz for pMOSFET [1]. These are comparable to the highest numbers reported for bipolar devices. The gate delay of a 101-stage unloaded CMOS-inverter ring oscillator is shown in Fig. 5 as a function of the supply voltage. At 1.5 V, the delay is 22 ps/stage, which is more than a factor of two faster than the 2.5-V, 0.25- μm channel CMOS devices currently being manufactured. The measured delays agree well with model simulations also plotted in Fig. 5 [1].

Manuscript received September 5, 1996; revised February 10, 1997.
The authors are with the IBM Thomas J. Watson Research Center,
Yorktown Heights, NY 10598 USA.
Publisher Item Identifier S 0018-9219(97)03191-5.