Device Scaling Limits of Si MOSFETs and Their Application Dependencies

DAVID J. FRANK, MEMBER, IEEE, ROBERT H. DENNARD, FELLOW, IEEE,
EDWARD NOWAK, MEMBER, IEEE, PAUL M. SOLOMON, FELLOW, IEEE, YUAN TAUR, FELLOW, IEEE,
AND HON-SUM PHILIP WONG, FELLOW, IEEE

Invited Paper

This paper presents the current state of understanding of the factors that limit the continued scaling of Si complementary metal-oxide-semiconductor (CMOS) technology and provides an analysis of the ways in which application-related considerations enter into the determination of these limits. The physical origins of these limits are primarily in the tunneling currents, which leak through the various barriers in a MOS field-effect transistor (MOSFET) when it becomes very small, and in the thermally generated subthreshold currents. The dependence of these leakages on MOSFET geometry and structure is discussed along with design criteria for minimizing short-channel effects and other issues related to scaling. Scaling limits due to these leakage currents arise from application constraints related to power consumption and circuit functionality. We describe how these constraints work out for some of the most important application classes: dynamic random access memory (DRAM), static random access memory (SRAM), low-power portable devices, and moderate and high-performance CMOS logic. As a summary, we provide a table of our estimates of the scaling limits for various applications and device types. The end result is that there is no single end point for scaling, but that instead there are many end points, each optimally adapted to its particular applications.

Keywords*—*CMOS, device design, discrete dopants, double-gate MOSFET, DRAM, high-k dielectrics, high-performance logic, leakage currents, limits, low power, MOSFET, nanotechnology, power density, scale length, scaling, SRAM, tunneling.

I. INTRODUCTION

In 1930, Lilienfeld [1] patented the basic concept of the field effect transistor (FET) . Thirty years later, in 1960, it was finally reduced to practice in Si–SiO$_2$ by Kahng and Attala [2]. Since that time, it has been incorporated into integrated circuits and has grown to be the most important device in the electronics industry. Progress in the field for at least the last 25 years has followed an exponential behavior that has come to be known as Moore's Law [3]. Since 1994, the semiconductor industry has been projecting these exponentials into the future to provide technology development targets. The most recent of these projections is the 1999 International Technology Roadmap for Semiconductors (ITRS99) [4]. It contains projections for complementary metal-oxide-sesmiconductor (CMOS) technology out to 2014, including 32-Gb dynamic random access memory (DRAM) entering production and processors with gate lengths down to 20 nm and 2×10^{10} FETs per chip.

But will these exponential projections come to pass or will physical limits make them impossible? Many reviews have been written about the current state and future prospects for Si MOS field-effect transistors (MOSFETs) and CMOSs [5]–[9]. In particular, many different scaling limits for MOSFETs have been proposed and discussed. In this work, we describe the current state of understanding of these scaling limits and seek to advance this state of understanding by addressing the ways in which application requirements must be intertwined with the setting of limits. The result in the end is that there will be no single "end to scaling," but rather, a wide range of limiting FET technologies, each optimally adapted to its applications.

Much of our discussion centers on bulk-like MOSFET scaling, as illustrated in Fig. 1, but this is not intended to exclude other device geometries for MOSFETs. In particular, partially depleted silicon-on-insulator (PD-SOI) MOSFETs are considered to be part of this bulk-like category, since most of the same limits apply to PD-SOIs as to bulk. Consequently, PD-SOI is not explicitly discussed except when there are significant device design differences. At the circuit level, there are, of course, some important features of SOIs, such as the floating body effects, but these are for the most part outside the scope of this paper. The scaling behavior of fully depleted silicon-on-insulator (FD-SOI) MOSFETs depends a great deal on the

Manuscript received March 30, 2000; revised September 26, 2000.

The authors are with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: djf@us.ibm.com; pwong@watson.ibm.com).

Publisher Item Identifier S 0018-9219(01)02066-7.