Boris Sviličić, M.Sc.
Antun Kraš, Ph.D.
Faculty of Maritime Studies Rijeka
Studentska 2
51000 Rijeka

CMOS TECHNOLOGY: CHALANGES FOR FUTURE DEVELOPMENT

Since the invention of the integrated circuit technologies, there has been an unprecedented growth of the electronic industry, with significant impact on the maritime industry. In the last twenty years and so, the strongest growth area of the electronic industry has been in CMOS VLSI technology. The sustained growth in CMOS VLSI technology is fueled by a continued shrinking of transistors to ever smaller dimensions. The benefits of miniaturization are higher packing densities, higher circuit speeds, and lower power dissipation. The transistors manufactured today are 20 time faster and occupy less than 1% of the area of those built 20 years ago. It is obvious that a continued reduction of the transistor area cannot sustain forever. This paper examines issues related to the future development of CMOS technology.

Key words: CMOS, MOSFET, VLSI, integrated circuit, scaling

1. INTRODUCTION

The remarkable characteristic of transistors that fuels the rapid growth of the electronic technology industry is that their speed increases and their cost decreases as their size is reduced. The transistors manufactured today are 20 times faster and occupy less than 1% of the area of those built 20 years ago [7]. It seems intuitively obvious that a continued reduction of the area of a transistor by a factor of 2 every three years cannot be sustained forever.

CMOS circuits were invented in 1963 by Frank Wanlass. Some twenty-five years later, CMOS technology has become the predominant technology in digital integrated circuits. This is essentially because area occupation, operating speed, energy efficiency and manufacturing costs have benefited and continue to benefit from the geometric downsizing that comes with every new generation of semiconductor manufacturing processes.

2. CMOS TECHNOLOGY

Complementary-symmetry/metal-oxide-semiconductor (CMOS) is a major class of integrated circuits. CMOS chips include microprocessor, microcontroller, static RAM, and other digital logic circuits. The words "complementary-symmetry" refers to the fact that the design uses symmetrical pairs of p-type and n-type MOSFET transistors for logic functions, only one of which is switched on at any time (Figure 1.) [1].
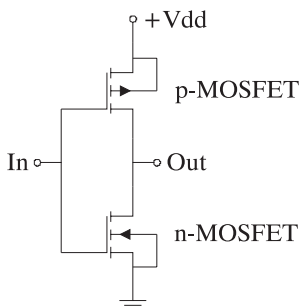
Figure 1. CMOS Inverter

The phrase "metal-oxide-semiconductor" is a reference to the nature of the fabrication process originally used to build CMOS chips (Figure 2) [3]. That process created field effect transistors having a metal gate electrode placed on top of an oxide insulator, which in turn is on top of a semiconductor material. Instead of metal, today the gate electrodes are almost always made from a different material, polysilicon, but the name CMOS nevertheless continues to be used for the modern descendants of the original process.

The performance of CMOS VLSI chip is measured by its integration density, switching speed, and power dissipation. CMOS circuits have the unique characteristic of practically zero standby power, which enables higher integra-
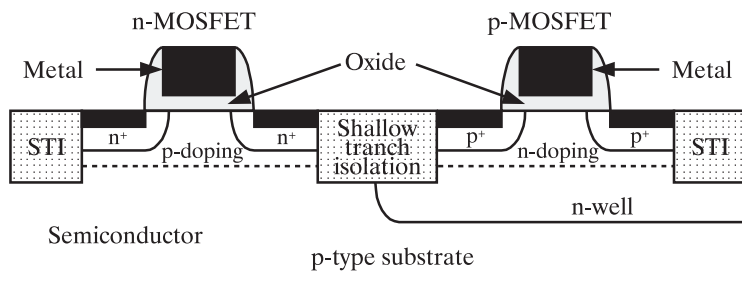
Figure 2. CMOS Inverter - schematic device cross section.

tion levels. The other important characteristic of CMOS devices is high noise immunity. Typical commercial CMOS products are integrated circuits composed of millions (or hundreds of millions) of transistors of both types on a rectangular piece of silicon of between 0.1 and 4 square centimeters.

3. MOSFET SCALING

The steady downscaling of MOSFET devices dimensions over the past two decades has been the main stimulus to the growth of CMOS integrated circuits. Scaled MOSFETs are desirable for a few main reasons. Scaled MOSFETs allow more current to pass. Conceptually, MOSFETs are like resistors in the on-state, and shorter resistors have less resistance. Furthermore, since scaled MOSFETs have lower gate capacitance, and since the amount of charge on a gate is proportional to its capacitance, logic gates incorporating scaled MOS-FETs have less charge to move. Indeed, these two factors combined contribute to lower switching times, and thus higher operating speeds.

Another important reason for MOSFET scaling is a reduced area, leading to a reduced cost. Scaled MOSFETs can be packed more densely, resulting in either smaller chips or chips with more computing power in the same area. Since the cost of fabricating a semiconductor wafer is relatively fixed, the cost of producing integrated circuits is mainly related to the number of chips that can be produced per wafer. Hence, smaller integrated circuits allow more chips per wafer, reducing the price per chip.

The basic principle of MOSFET scaling lies in scaling the device voltages and device dimension (both horizontal and vertical) by the same factor so that the electric field remains unchanged (Figure 3) [1]. This assures that reliability of the scaled device is not worse than that of the original device.

Figure 4 shows the scaling trend of a power-supply voltage (V_{dd}), threshold voltage (V_t), and gate-oxide thickness (t_{ox}) as a function of CMOS channel.
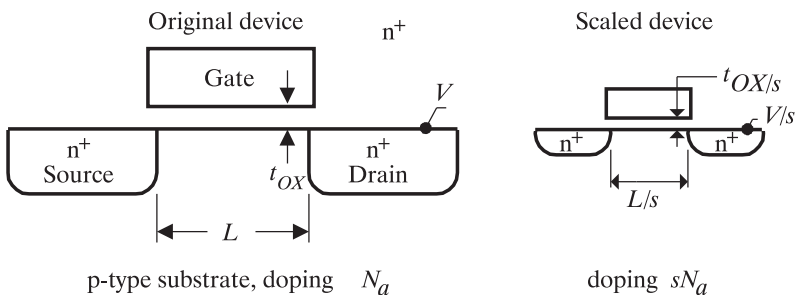

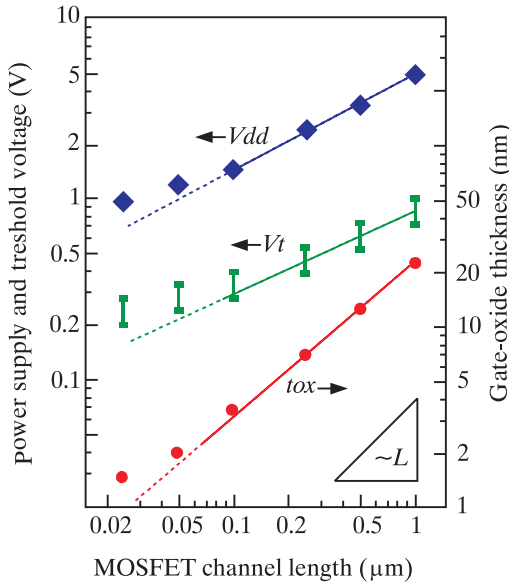Figure 3. Principles of MOSFET constant electric field scaling.

Figure 4. History and trends of a power-supply voltage (V_{dd}), threshold voltage (V_t), and gate-oxide thickness (t_{ox}) vs. channel length for CMOS logic technologies.

4. SCALING LIMITS

Producing MOSFETs with source-to-drain spacing i.e. channel length (Figure 3, value *L*) smaller than micrometer is a challenge, and the difficulties of semiconductor device fabrication are always a limiting factor in advancing integrated circuit technology. Recently, MOSFET scaling was propelled by the rapid advancement of high-resolution lithographic techniques and ion implantation techniques [2], and the small size of the MOSFET has created operational problems.

4.1. Subthreshold leakage

Because of small MOSFET geometries, the voltages that can be applied to the gate must be reduced as well to maintain reliability. As threshold voltage is reduced as well, the transistor cannot be completely turned off. The transistor operates in a weak-inversion mode, with a subthreshold leakage between source and drain.

The standby power of a CMOS chip due to source-to-drain subthreshold leakage is given by

$$P_{off} = W_{tot} V_{dd} I_{off} = W_{tot} V_{dd} I_0 \exp\left(-\frac{qV_t}{mkT}\right)$$

where W_{tot} is the total turned-off device width with V_{dd} across the source and drain, I_{off} is the average off-current per device, I_0 is the extrapolated current per width at threshold voltage, m is a dimensionless ideality factor, and V_t is the threshold voltage. The off-state leakage current would increase by about ten times for every 0.1 V reduction of V_t. For a chip with an integration level of 100 million transistors, the average leakage current of turned-off devices should not exceed a few times 10^{-8} A. This constraint holds the threshold voltage to a minimum of about ~0.2 V at the operating temperature (100°C worst case).

4.2. Heat production

As more transistors are integrated on a chip to enable more functions and a higher frequency is used to obtain an increased performance, the total power consumption increases and results in more heat. The active power of a CMOS chip (crossover currents are usually negligible), is given by

$$P_{ac} = C_{tot} V_{dd}^2 f$$

where C_{tot} is the total node capacitance being charged and discharged in a clock cycle, and f is the clock frequency. As CMOS technology advances, clock frequency goes up. The total switching capacitance is likely to increase as well, as one tries to integrate more circuits into the same or an even larger chip area. The active power of today's high-end microprocessors is already in the 50 - 100 W range.

Integrated circuits operate slower at high temperatures, and have reduced reliability and shorter lifetimes. Heat sinks and other cooling methods are now required for many integrated circuits including microprocessors.

4.3. Gate dielectric tunneling

The gate electrode turns the transistor on and off, and a gate dielectric is an insulator beneath it (Figure 3, value t_{OX}). The voltage of the gate electrode above it controls the flow of electric current across the transistor. The gate dielectric should be made as thin as possible to increase the channel conductivity and to keep short-channel effects under control when a transistor is on, and to reduce subtreshold leakage when a transistor is off. To keep the electric field constant during the MOSFET scaling, the gate dielectric thickness should be shrink proportional to the scaling factor:

$$F_{OX} = \frac{V_{GS} - \Phi_S}{t_{OX}} = F'_{OX} = \frac{V'_{GS} - \Phi'_S}{t'_{OX}}$$

where F_{OX} is the standard MOSFET electric field, V_{GS} is the standard MOS-FET gate voltage, and t_{OX} is the standard MOSFET gate dielectric thickness. F'_{OX}, V'_{GS}, , and t'_{OX} are the electric field, gate voltage, and gate dielectric thickness for the scaled MOSFET device.

The gate dielectric (currently made of silicon dioxide) thickness of today's high-end microprocessors in a 65 nm transistor is 1.2 nm, which represents a thickness of only five atomic layers [6]. While it is amazing that a silicon dioxide (SiO_2) can carry us this far without being limited by extrinsic factors such as defect density, surface roughness, or large-scale thickness and uniformity control, oxide films, this thin are subject to a quantum-mechanical tunneling, giving rise to a gate leakage current that increases exponentially as the oxide thickness is scaled down [4]. These tunneling currents can initiate damage, leading to previously unexpected reliability concerns in very thin gate dielectrics. New device structures or a new gate oxide with a higher dielectric constant will have to be developed to enhance performance without further reduction of the gate-oxide thickness. No scaling theory exists to guide either development. In this sense, we will indeed soon reach the limits of CMOS scaling.

4.4. Wiring and interconnections

The wiring required to interconnect transistors must scale at the same rate as the transistors in order to take advantage of the improvements in size and speed [5]. The traditionally switching time was roughly proportional to the gate capacitance of gates. However, with transistors becoming smaller and more transistors being placed on the chip, the capacitance of the wires connecting different parts of the chip is becoming a large percentage of capacitance. Reducing the wiring resistance and the inter-layer film capacitance is crucial to propagating the high-speed pulses generated in a CMOS device with little delay and waveform distortion.

As the dimensions of the interconnect structures continue to shrink, aluminium-based inter-connects are being replaced by lower-resistance copper metallurgy, which can decrease both the wiring resistance and the capacitance. A research is also underway to move from the CVD oxide/nitride interlayer dielectrics between wiring levels to various low-dielectric-constant insulators, which can further decrease the wiring capacitance. Despite these major changes in materials, there is a concern that owing to higher resistivity and capacitance, the extremely small wires will be unable to support performance enhancements.

5. CONCLUSION

Conventional CMOS fabrication processes have supported many generations of shrinking, with each new generation bringing predictable advantages in terms of cost, speed and power consumption. However, they are rapidly approaching a brick wall where cost of further progress will outweigh the benefits. Semiconductor manufactures have been investigating alternative transistor architectures such as DG (double-gated) MOSFETS, SOI (Silicon on Insulator) technology, and Tri-gate technology, but a major problem is the manufacturability of these processes. The CMOS technology will soon have a lower rate of performance increase as their size is reduced.

At present, it does not appear that there is a viable contender to displace the CMOS technology. No alternative logic technology is evolving within cost/performance products on the market today to threaten the CMOS dominance. And in light of the continuing CMOS performance evolution, an even steeper evolution and learning curve would be necessary to displace CMOS. There will likely be variations on CMOS circuits, but no radical shift in circuit type seems to be on the horizon.

BIBLIOGRAPHY

1. Biljanović, P, Poluvodički elektronički elementi, Školska knjiga Zagreb, 1996.
2. Chang, C. Y., S. M. Sze, ULSI technology, New York, McGraw-Hill, 1996.
3. Taur, Y., T. H. Ning, Fundamentals of Modern VLSI Devices, Cambridge, Cambridge University Press, 1998.
4. Taur, Y., /et al./, CMOS Scaling into the Nanometer Regime, Proc. IEEE 85, 1997., 486–504.
5. Theis, T. N., The Future of Interconnection Technology, IBM J. Res. Develop. 44 (2000), 3, 379–390.
6. http://www.intel.com/technology/silicon.
7. http://www.intel.com/research.

Sažetak

CMOS TEHNOLOGIJA: IZAZOVI ZA RAZVOJ U BUDUĆNOSTI

Od pojave tehnologije integriranog sklopa, elektronička industrija bilježi nezapamćen razvoj, vršeći snažan utjecaj na pomorstvo. Posljednjih dvadesetak godina razvoj elektroničke industrije zasnovan je na CMOS VLSI tehnologiji. Neprekidni napredak CMOS VLSI tehnologije omogućen je kontinuiranim smanjivanjem dimenzija MOS tranzistora.

Rezultat skaliranja je veća gustoća pakiranja komponenata po elektroničkom sklopu, veća brzina rada i manja disipacija snage po tranzistoru. Današnji su tranzistori 20 puta brži te zauzimaju 1% prostora u odnosu na tranzistore proizvedene prije 20 godina. Očito je da smanjivanje površine tranzistora ne može ići u beskonačnost. Ovim radom se istražuju potencijalna ograničenja razvoja CMOS tehnologije.

Ključne riječi: *CMOS, MOSFET, VLSI, integrirani sklop, skaliranje*

Pomorski fakultet u Rijeci
Studentska 2
51000 Rijeka