A Review of Low-Power VLSI Technology Developments

Nakka Ravi Kumar

Abstract Ever since the invention of integrated circuits, there has been a continuous demand for high-performance, low-power, and low-area/low-cost diversified applications from a variety of consumers. This demand has been pushing the fabrication process beyond ultra-deep, sub-micron technologies such as, 32, 22, 14 nm, and so on. In this chapter, various technological aspects for low-power applications are reviewed in detail, along with the evolution of new technology, bearing in mind the PPA (power, performance, and area). Some basic reviews of components of power consumption in CMOS are also given.

Keywords Low-power VLSI · VLSI technology
CMOS power consumption · Bulk CMOS · SOI · FINFET

1 Introduction

Electronic devices and systems have entered into almost every corner of human life. Most of these devices, at present, are portable and battery operated. The evolution of the computer moved from requiring a huge powerhouse to run one device, to a small handheld battery operated device like a palm-top. All these portable electronic devices need ICs to operate under low-power consumption. Along with the lower power consumption demand, there is a demand for high performance. In addition, all these low-power and high-performance devices are demanded at low cost. The aforementioned demands have pushed the fabrication process from a micrometer level to a nanometer level. Scaling down vertical and lateral dimensions and the voltages of the MOS transistors will increase the performance of the circuit in terms of speed, decrease the overall power dissipation of the chip, and reduce the area of the chip, thereby the cost per transistor.

N. Ravi Kumar (✉)
ECE Department, Mahaveer Institute of Science and Technology, Keshavagiri Vyasapuri, Bandlaguda, R.R.District, Hyderabad 500 005, Telangana, India
e-mail: nrkumar_2002@yahoo.co.in

H.S. Saini et al. (eds.), *Innovations in Electronics and Communication Engineering*, Lecture Notes in Networks and Systems 7,
https://doi.org/10.1007/978-981-10-3812-9_2

There are various scaling methods adopted for CMOS: constant field scaling or full scaling, constant voltage scaling, and generalized scaling [1–3].

Due to the restrictions of some material properties, such as the energy band gap, built in voltage, etc., the voltages could not be scaled down as rapidly as dimensions. Because of this, current densities and power densities will increase. An increase in power density increases heat. After a certain limit, managing heat becomes more expensive, if the power density is not addressed properly, than the actual chip fabrication cost.

In addition to this, the need of portable battery operated devices along with their reliability issues, thermal dissipation, and scalability, have demanded for low supply voltages.

CMOS is the best choice for low-power applications. The power consumption of CMOS can further be reduced to a great extent by carefully studying the parameters influencing the power consumption and then by designing the CMOS technology for power efficiency. In the text below (Sect. 2) we deal with various components of CMOS power consumption and the parameters influencing each of them.

The drive for lowering power dissipation is emphasized in Sect. 3.

Technological approaches to lowering the power consumption are described in Sect. 4. The need for low power along with high speed for certain applications has led to new technological inventions. These demands and inventions are described in Sect. 5. Section 6 concludes the chapter.

2 Components of CMOS Power Consumption

The total power consumption in a CMOS circuit consists of three components, namely, dynamic power (P_D), static power/leakage power (P_{leak}), and short circuit power (P_{SC}).

2.1 Dynamic (Active) Power Consumption (P_D)

This component of power consumption is due to switching activity of logic gates in a CMOS circuit. The gates will be continuously charging and discharging their load capacitances according to the logic-switching activity. This power consumption depends on power supply voltage (V_{DD}), switching frequency (f_s), and switching node capacitance (C_l). Dynamic power occurs when the device is 'ON'.

If the total number of gates in a chip are 'n' and the average switching frequency and the average switching node (load) capacitances are 'f_s' and 'C_l', respectively, then the total dynamic power consumption of the chip is given by:

$$P_{\mathrm{D}} = n * f_{\mathrm{s}} * 1/2 * C_{\mathrm{l}} * V_{\mathrm{DD}}^2 \quad (1)$$

2.2 Static DC Power/Leakage Power Consumption (P_{leak})

Static power occurs when the device is 'OFF'. The MOS transistors act as switches for digital applications. An ideal switch is expected to have infinite input impedance and infinite 'OFF' state resistance.

Infinite input impedance means the gate should be perfectly isolated from the substrate. However, in practical terms when the gate oxide thickness is scaled down, there will be some leakage of current from the gate (I_{gate}) to drain due to tunneling and hot carrier injection (Fig. 1).

Infinite 'OFF' state resistance means there should not be any drain to source current when the gate voltage is below its threshold. However, there will be some diffusion current when the gate voltage is below VT. This current is called sub-threshold leakage current (I_{sub}). There is another leakage current from drain to substrate, which is called junction leakage current (I_{j}). One more source of leakage current is gate induced drain leakage (GIDL; I_{GIDL}) (Fig. 1).

The total leakage current (I_{leak}) is given by:

$$I_{\mathrm{leak}} = I_{\mathrm{gate}} + I_{\mathrm{sub}} + I_{\mathrm{j}} + I_{\mathrm{GIDL}} \quad (2)$$

Leakage power (P_{leak}) is given by: $P_{\mathrm{leak}} = I_{\mathrm{leak}} * V_{\mathrm{DD}}$. Though leakage power is negligible compared to active power, for a long channel behavior device, it becomes comparable for devices with short channel effects. Short channel effects are the result of generalized scaling, where voltages are not scaled down as rapidly as dimensions are. As shown in Fig. 2, the power consumption due to leakage current becomes a significant factor for the total power consumption.

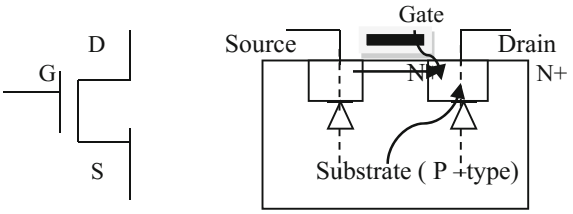

Fig. 1 Sources of leakage currents in an NMOS transistor

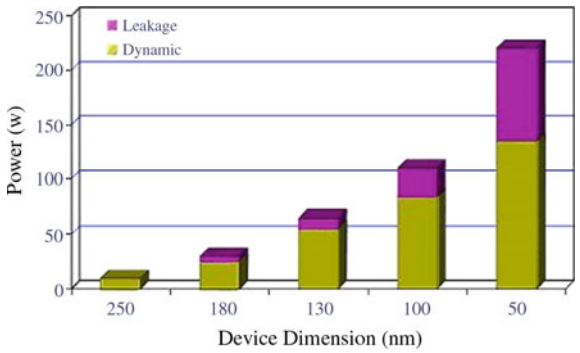

Fig. 2 Trend of dynamic and leakage power against device dimensions

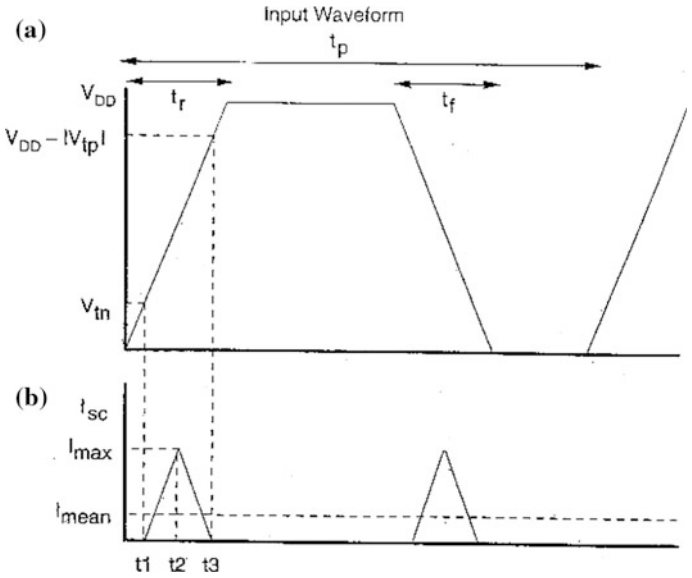

Fig. 3 **a** Input waveform; **b** Short circuit current spikes

2.3 Short Circuit Power Consumption (P_{SC})

This component of power consumption comes about due to the momentary 'ON' state of both NMOS and PMOS of a CMOS gate during transitions. Due to a finite rise time and fall time of the input waveform of a gate, both NMOS and PMOS transistors are 'ON' for short periods of time. As shown in Fig. 3, this results in spikes of short circuit currents during transitions.

The short circuit power (P_{SC}) is given by: $P_{SC} = I_{mean} * V_{DD}$. This short circuit current can be totally eliminated by making $V_{DD} < |V_{tp}| + V_{tn}$.

3 Drive to Lower Power Dissipation

Even if one can afford to pay for the power, one cannot get along with it because, as the chip size shrinks, the power dissipation per unit area increases. This will increase the temperature of the chip. The increased temperature will deteriorate the carrier mobility and V_T. Therefore, the performance of the chip will be greatly deteriorated. Moreover, as the temperature rises, the minority carrier concentration also rises. This leads to increased leakage current. As we have seen in the previous section, increased leakage current leads to increased static power dissipation. This will further increase the chip temperature and eventually lead to chip breakdown.

Though scaling of CMOS technologies reduces the overall power dissipation of a chip, the power density increases (Fig. 4) for conventional CMOS processes. The power density increases with decreasing minimum feature size of CMOS technology for generalized scaling.

The projection of power density is not far from that exhibited on a nuclear reactor surface. This has driven the requirement to invent new technologies and design methodologies for low-power dissipation density.

In addition to these thermal limitations, there is a constant demand for low-power consumption from battery operated portable devices. Low-power dissipation ensures long battery life.

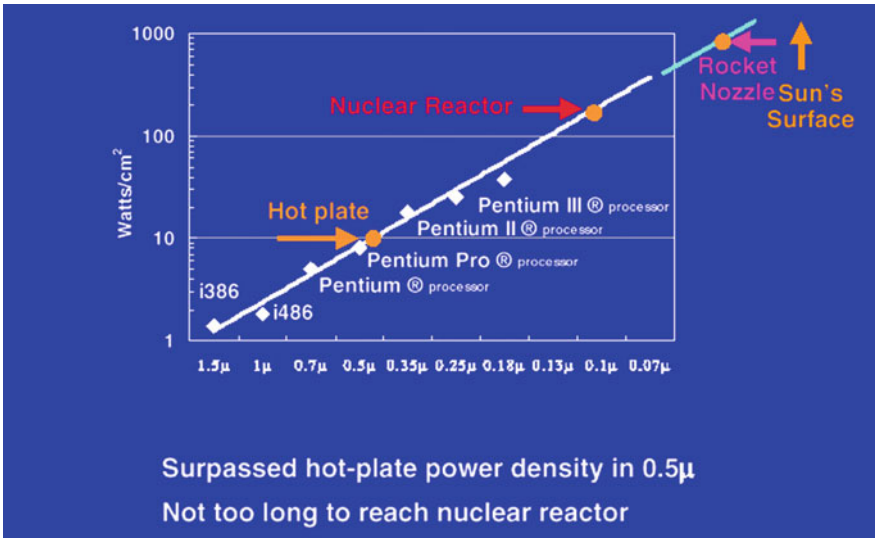

Fig. 4 Power density against feature size of CMOS technology (Courtesy of Fred Pollack, Intel CoolChips tutorial Micro-32)

4 Approaches for Lowering Power Consumption

Power consumption has to be addressed both at technological level as well as at circuit design level. For ultra-low-power applications, optimization and adoption of innovative circuit design techniques at all levels of abstraction, along with innovations in new technologies, have become mandatory.

As we have seen in Sect. 2, the overall power consumption consists of three components, from which short circuit power consumption can be eliminated.

We shall now identify methods of reducing dynamic power and leakage power.

4.1 Ways to Reduce Dynamic Power (P_D)

The dynamic power in CMOS, given by (1), depends on switching frequency (f_s), load capacitance (C_L), and power supply voltage (V_{DD}). The dynamic power can be reduced by reducing any, some or all of these parameters.

Dynamic power has a quadratic dependence on V_{DD}. Therefore, reducing V_{DD} reduces P_D drastically. However, V_{DD} cannot be reduced arbitrarily, because, it has to be compatible with the system in which the chip is assembled. However, the power supply voltage of CMOS technologies has been scaled down (as shown in Table 1) after a global agreement among major VLSI industries.

Some circuit design techniques adopt a multiple power supply. They generate a low supply voltage internally for power sensitive blocks and retain a high V_{DD} for critical path circuits [4].

Switching frequency can be reduced by using encoding and decoding techniques.

The load capacitance, C_L, constitutes gate oxide capacitance, C_{ox}, interconnect capacitance, C_I, and junction capacitance, C_J. By optimizing the gate areas and interconnect areas, load capacitance can be reduced. At the technological level, the interconnect capacitance can be reduced by low '*k*' (low dielectric constant) dielectric layers underneath the metal layers. The junction capacitance can be reduced greatly in SOI and FINFET technologies.

Table 1 Typical scaling scenario of V_{DD}

Year of introduction	Technology node	Power supply (V_{DD})
1974	5 μm	10 V
1984	1 μm	5 V
1994	0.35 μm	3.5 V
2004	90 nm	1 V

4.2 Ways to Reduce Static Power

As discussed in Sect. 2.2, as the technology shrinks, static power consumption becomes comparable to dynamic power consumption.

The static power in CMOS, given by (3), depends on I_{leak} and V_{DD}. Reducing V_{DD} reduces static power. I_{leak} should also be reduced for ultra-low-power applications.

Dealing with these four components of leakage current, as given in (2), is becoming more and more of a prime concern as technology shrinks.

Sub-threshold Leakage Current (I_{sub})
Low V_{DD} requires a low threshold voltage. Leakage current increases as the threshold voltage decreases. For a particular target threshold voltage, the sub-threshold leakage current can be minimized by making the sub-threshold slope close to its minimum value. The sub-threshold slope is given by:

$$S = (\ln 10) * (kT/q) * ((C_{ox} + C_d + C_{it})/C_{ox}) \tag{3}$$

where C_{ox} is the gate oxide capacitance; C_{it} is the interface trap capacitance; and C_d is the depletion capacitance. The minimum value of the sub-threshold slope is $(\ln 10) * (kT/q)$, which is approximately 60 mv/decade. The sub-threshold slope can be made close to its minimum value by making $C_{ox} \gg (C_d + C_{it})$.

Now, C_d can be reduced by making the substrate doping very low. However, making substrate doping low causes 'punch through'. To avoid this, new technologies, like SOI and FINFET, have been invented.

The interface trap capacitance, C_{it}, can be reduced with a good insulator–substrate interface.

Junction Leakage Current (I_j)
The substrate current generated by impact ionization at the drain substrate junction triggers a parasitic bipolar action in the MOSFET. This onset of bipolar action further increases the leakage current. So, preventing this bipolar action is essential.

Gate Leakage Current (I_{gate})
Gate leakage current consists of band-to-band tunneling current, hot carrier injection current, and gate induced drain leakage current.

As the gate dielectric thickness is reduced, gate leakage current becomes a serious problem. One possible solution is to choose high dielectric constant materials instead of silicon dioxide (SiO_2). For the same capacitance, the insulator thickness can be increased for high k dielectrics compared to SiO_2. Therefore, the electric fields are reduced and hence the leakage currents are too. Some choices of high-k dielectrics are: Al_2O_3, HfO_2, ZrO_2, Y_2O_3, etc. [1].

Gate Induced Drain Leakage Current (I_{GIDL})
This phenomenon occurs due to an accumulation of the majority carriers in the substrate under the gate near the drain. In the case of NMOS, the accumulated holes under the gate, near the drain/substrate junction, causes pre-matured impact ionization. This leakage is called I_{GIDL} [5], which can be reduced by using high-k gate dielectrics.

5 New Technologies for Ultra Low Power and High Performance

As discussed in the previous section, lowering V_{DD} reduces power consumption and also suppresses reliability problems. In conventional CMOS technology, where $POCL_3$ degenerately doped N+ poly silicon is used as a gate material for both NMOS and PMOS transistors, the threshold voltage adjustment requires a separate boron implantation. In this technology, a PMOS transistor turns out to be a buried channel device, which is also known as a compensated MOSFET. Reducing the threshold voltage of these compensated MOSFETs is difficult. Therefore, to achieve low V_T for both NMOS and PMOS, dual-poly gate technology is adopted. Here, N+ poly is used for NMOS and P+ poly is used for PMOS. The poly doping is undertaken along with a self-aligned source/drain implant [6]. However, these conventional bulk CMOS technologies are not free from sub-threshold leakage, parasitic bipolar current, punch through, DIBL, high junction capacitances, and other short channel effects. There is one further problem with this technology—the poly depletion effect [7], which becomes more prominent as the gate oxide thins.

The performance and power goals for certain applications in the advanced nodes of 40, 32 nm, etc., could not be achieved with conventional bulk CMOS processes. This has led to alternative technologies.

These new technologies are SOI and FINFET. In both of these technologies the body is made very thin so that the gate has maximum control over the channel. The threshold voltage is adjusted using mid-gap materials for gate electrodes. These technologies postpone the short channel effects to future generations. These technologies ensure not only a reduced I_{off}, due to the decreased leakage currents, but also provide increased I_{ON}, due to an increase in carrier mobility owing to very lightly doped or undoped bodies.

5.1 SOI Technology

Silicon-on-insulator (SOI) technology was invented to reduce a device's leakage current, increase its 'ON' current, and reduce its capacitance [8–10].

In SOI fabrication technology, transistors are built on a silicon layer resting on an insulating layer of SiO_2 (as shown in Fig. 5). The insulating layer sits on top of the silicon substrate.

There are different ways of manufacturing SOI wafers: SOS (silicon-on-sapphire), SIMOX (Separation by IMplanted OXygen), BESOI (bond and etch-back SOI), Smart Cut, and ELTRAN (Epitaxial Layer TRANsfer). The description of all these processes is beyond the scope of this chapter.

If the body is not very thin, it will be partially depleted. The device is then called partially depleted SOI or PDSOI. These devices suffer history effect and kinks in output characteristics due to parasitic bipolar action. These problems are avoided by

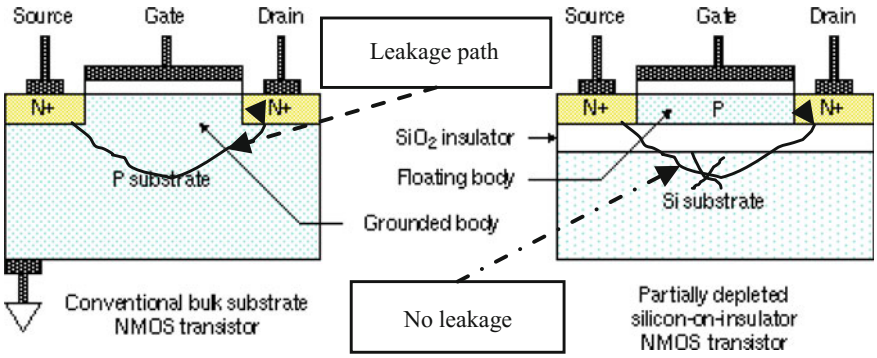

Fig. 5 A bulk NMOS transistor versus SOI NMOS

making the body and base oxide (BOX) thicknesses very thin. The body of these devices is fully depleted, and the technology is thus called fully depleted SOI or FDSOI [11]. If the body thickness is ultra-thin, the technology is called ultra-thin body SOI or UTBSOI.

The advantages of SOI over Bulk CMOS are: higher performance due to reduced capacitances and increased I_{ON}; latch-up elimination due to vertical and lateral isolation of the transistors; reduced temperature sensitivity; a removal of the need for body or well taps; reduced antenna effects; small transistor sizes saving areal extent; low-power consumption due to ultra-low-leakage currents; and reduced parasitic capacitances.

The disadvantages are: high wafer costs and self-heating.

Heating may not be a problem in ultra-low-power devices. We consider SOI technology is the best choice for RF applications.

5.2 FINFET Technology

As mentioned in the previous section, the solution to leakage problems is to make the body ultra-thin in order to have better gate control. In SOI technologies, the body is thin and planar. However, the body can also be made thin and vertical, where the gate will be on both sides, or on all sides (as shown in Fig. 6). This technology is called FINFET technology [12, 13].

The thin silicon fin can be made on a bulk silicon substrate or on a SOI substrate. As shown in Table 2, both these technologies have their own advantages and disadvantages.

INTEL uses FinFET technology down to 10-nm nodes.

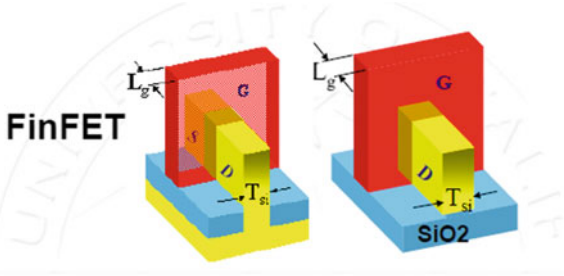

Fig. 6 Fin on bulk and fin on SOI

Table 2 Fin on bulk versus fin on SOI

Fin on bulk substrate	Fin on SOI substrate
Δ HFin (variation in Fin height) is greater (−)	Δ HFin is less (+)
Substrate is cheaper (+)	SOI substrate is costly (−)
Heat is easily dissipated (+)	Heat dissipation is a problem (−)
Parasitic BJT will exist (−)	No parasitic BJT (+)
Source/drain region epitaxy is possible (+)	Source/drain epitaxy is not possible (−)

6 Conclusion

Demand for high performance and low-power consumption has led CMOS scaling from micrometer ranges to nanometer ranges. These demands could not be met with conventional CMOS technologies and thus led to the invention of new technologies. Various components of power consumption in CMOS are discussed in detail in this chapter. The drive for low-power consumption and the methodologies adopted to reduce power consumption in CMOS are also discussed in detail. Finally, the technologies developed to achieve these goals have been reviewed.

References

1. Sze SM Physics of semiconductor devices, 3rd edn. Wiley India publications
2. Moore GE (1965) Cramming more components on integrated circuits. Electronics 38(8):19
3. Haron NZ, Hamdioui S, Why is CMOS scaling coming to an end? http://ce-publications.et.tudelft.nl/publications/510_why_is_cmos_scaling_coming_to_an_end.pdf
4. Kim K, Agrawal VD (2011) Ultra low energy CMOS using below-threshold dual-voltage supply. J Low Power Electron 7:1–11
5. Chen J et al (1987) Sub-break down leakage current in MOSFETS. IEEE Electron Device Lett EDL-8:515

6. Pandey SM, Kumar NR, Sankar GK, Pal DK, Roy JN (1999) Optimization of dual poly gate low voltage CMOS process. Proceedings of the Tenth International Work shop on the Physics of Semiconductor Devices,Vol part2, pp 871–874 ISSN: 0277786X, ISBN: 9780819436016 [0819436011]
7. Sallase JM (2000) Improved analytical modeling of polysilicon depletion in MOSFETS for circuit simulation. Solid-State Electron 44:905–912
8. Bhat KN, Silicon on insulator (SOI) devices. Lecture notes. http://ece.iisc.ernet.in/~navakant/nano/2007/Lecture23.pdf
9. Singh RK, Saxena A, Rastogi M (2011) Silicon on insulator technology review. Int J Eng Sci Emerg Tech 1(1):1–16 ©IJESET. ISSN: 2231-6604
10. Crisoloveanu S (2000) Architecture of SOI transistors: what's next? In: 2000 IEEE international SOI conference, pp 1–2
11. Sakurai T, Matsuzawa A, Douseki T (2006) Fully-depleted SOI CMOS circuits and technology for ultralow-power applications. Springer, New Jersey
12. Liu T-JK (2012) FinFET history, fundamentals and future. In: 2012 symposium on VLSI technology short course
13. Hu C, FinFET and other new transistor technologies. Lecture notes. http://microlab.berkeley.edu/text/seminars/slides/2011-8_FinFET_and_the_Concept_Behind_It.pdf

http://www.springer.com/978-981-10-3811-2

Innovations in Electronics and Communication Engineering
Proceedings of the Fifth ICIECE 2016
Saini, H.S.; Singh, R.K.; Reddy, K.S. (Eds.)
2018, XXII, 596 p. 320 illus., Hardcover
ISBN: 978-981-10-3811-2