Chapter 2
Low-Power Electron Devices

Nobuyuki Sugii

2.1 Introduction

Low-power electron devices are extremely important in terms of energy saving by taking advantage of the information and communication technology (ICT) as well as power reduction of electronic apparatuses themselves. Although the low-power approaches should be done in every technological aspect of the ICT technology, the low-power electron device (CMOS) technology is especially important because this technology is in a core part of many ICT apparatus technology.

In this chapter, the issues regarding the transistor miniaturization that has enabled both reducing power and enhancing functionality of CMOS large-scale integrations (LSIs) for about 40 years, and possible solutions regarding device structure and materials are reviewed.

2.2 CMOS Miniaturization and Issues for Low Power

2.2.1 CMOS Miniaturization and Scaling Rule

For about 40 years, the number of transistors has increased by twice per every two years as Gordon Moore predicted in 1965 [1]. The trend of transistor number in microprocessor chips is shown in Fig. 2.1 [2]. The state-of-the-art CMOS-LSI chip contains more than one billion transistors. This technological evolution has been done by a continuous miniaturization of transistor size. The paradigm of the CMOSFET miniaturization is known as the Dennard's scaling rule [3]. If the areal

N. Sugii (✉)
Low-Power Electronics Association and Project (LEAP), Ibaraki, Japan
e-mail: n-sugii@ieee.org

T. Kawahara and H. Mizuno (eds.), *Green Computing with Emerging Memory*,
DOI: 10.1007/978-1-4614-0812-3_2,

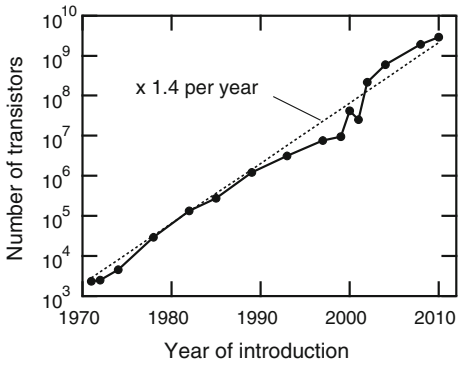

Fig. 2.1 Evolution of transistor number in microprocessors

dimensions are simply shrunk, the short-channel effect (SCE) occurs. The SCE degrades the cut-off characteristics of transistor and also decreases threshold voltage (V_{th}) that is the important transistor parameter to determine the gate voltage to turn on the transistor. The basic idea of the scaling rule is to shrink all the dimensional parameters including depth scale as well as areal scale. Table 2.1 shows the scaling parameters based on this rule. The parameter κ denotes scale ratio per generation and is $\sqrt{2}$ in general. The left-hand parameters in the table are the ideal scaling parameters. If we can follow these ideal parameters, we can increase the number of transistors by κ^2 in the same chip area and transistor speed by κ, and at the same time, we can reduce power consumption of single transistor by $1/\kappa^2$, that is, the power consumption of the LSI chip of the same area (power density) remains constant. The scaling obeying the ideal parameters has been very difficult or impossible any more, especially for supply-voltage (V_{dd}) reduction. As a result, the actual scaling parameters in recent years are as shown in the right-hand portion of the table by introducing the parameter α. This additional parameter automatically increases power consumption of CMOS LSIs significantly.

2.2.2 Power Consumption and Efficiency of CMOS Circuits

The simple expression of power consumption of CMOS circuits is,

$$P = n\left(CV_{dd}{}^2 f + I_{leak} V_{dd}\right), \tag{2.1}$$

where P, n, C, V_{dd}, f, and I_{leak} denote power, transistor number, load capacitance, supply voltage, operation frequency, and leakage current of a transistor, respectively. The first and second terms in RHS indicate ac (or active) and leakage (or static) power, respectively. Regarding the ac power, reducing V_{dd} is especially important because nC is nearly constant and f increases with scaling. Note that the circuit and architecture technologies that will be mentioned in chap. 7, such as power gating, clock gating, and multi-cores, can significantly reduce both n and f.

Table 2.1 Scaling parameters

Parameter	Ratio	Modified ratio
Physical dimension (t_{ox}, L, W)	$1/\kappa$	$1/\kappa$
Channel impurity density (N_A)	κ	$\alpha\kappa$
Voltage (V)	$1/\kappa$	α/κ
Current (I)	$1/\kappa$	α/κ
Electric field (E)	1	α
Depletion depth (X_d)	$1/\kappa$	$1/\kappa$
Gate capacitance ($\varepsilon LW/t_{ox}$)	$1/\kappa$	$1/\kappa$
Inversion carrier density (N_s)	1	α
Carrier velocity (v)	1	1
Delay ($\tau = C_{load}V/I$)	$1/\kappa$	$1/\kappa$
Power consumption ($P = IV$)	$1/\kappa^2$	α^2/κ^2
Power-delay product ($P\tau$)	$1/\kappa^3$	α^2/κ^3
Transistor density (n)	κ^2	κ^2
Power density ($P/LW = IV/LW$)	1	α^2

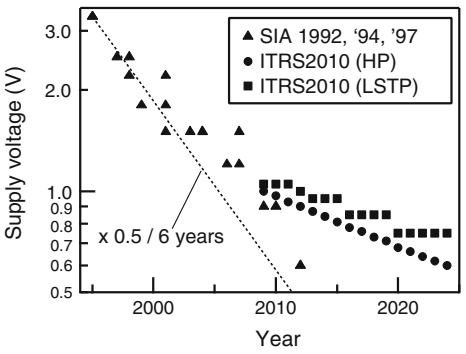

Fig. 2.2 Supply-voltage projections in roadmaps

The leakage power also decreases with decreasing V_{dd}. The reduction of V_{dd} is thus primary requirement for low power.

The past trend and future projection of V_{dd} are shown in Fig. 2.2 [4, 5]. Until beginning of 2000s, V_{dd} had decreased regularly, but the rate was slower than the requirement of the scaling rule. Furthermore, the most recent roadmap [5] suggests that the V_{dd} reduction is significantly retarded. This is mainly caused by difficulty in ensuring operation voltage margin in highly scaled and densely packed circuits. Actually, the minimum operation voltages for static random access memory (SRAM) and logic circuits, as shown in Fig. 2.3 [6], rather increase with miniaturization.

Another important aspect for low-power CMOS is power efficiency. The power efficiency can be defined as energy per instruction and, simply dividing (2.1) by f, this can be expressed as

$$E = n\,(CV_{dd}^2 + I_{leak}V_{dd}/af) \tag{2.2}$$

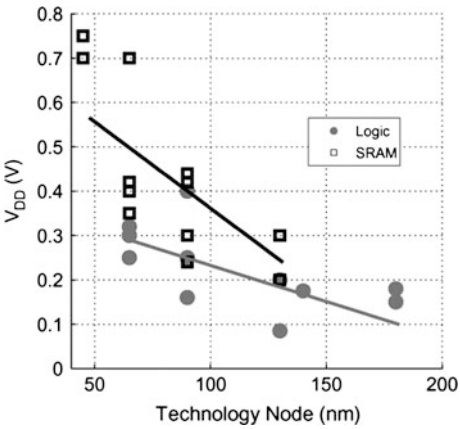

Fig. 2.3 Minimum operation voltages for logic and SRAM circuits [6] © IEEE

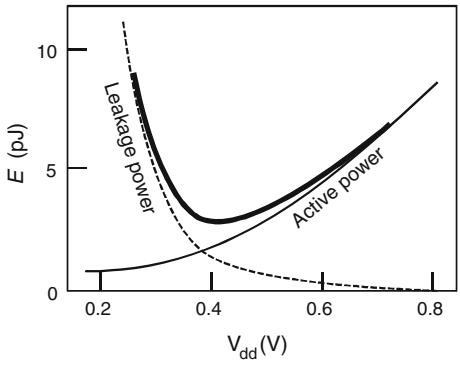

Fig. 2.4 Energy per operation as a function of supply voltage

where a denotes activity of the circuit to indicate the effect of leakage power during idle time.

The ac power can be decreased if we can reduce V_{dd} within the operation margin; however, operation speed also decreases. This means that it takes longer time to do the same task and increases the proportion of leakage power resulting to increase E. The best V_{dd} value thus exists compromising speed, ac power, and leakage power. Figure 2.4 schematically shows the relationship between E and V_{dd}. Generally speaking, in the CMOS circuits, the best V_{dd} value is at near 0.4 V [7]. Any transistor should be operated under the condition of minimum E in principle.

However, in many applications, the speed of logic circuits operated at $V_{dd} = 0.4$ V is insufficient. Many logic circuits in the current generation thus operate at 0.8–1.2 V at the expense of energy efficiency. The solution for this issue in the circuit technology of the recent generations is the dynamic voltage and frequency scaling (DVFS) [8] for example, and the ultimate solution might be a

parallelism (too many cores) with reduced V_{dd} down to 0.4 V and slower unit-circuit operation speed for the highest energy efficiency.

2.2.3 CMOS Transistor Scaling Issues

As indicated in the previous section, the past scaling rule has not been ideal. This is because there are many unscalable parameters and the influence of these parameters has become serious in recent years. Typical unscalable parameters are threshold voltage V_{th}, subthreshold swing S, gate capacitance C_{ox}, and average carrier velocity in a source side of transistor v_s.

In logic circuits, operation speed is basically determined by transistor delay ($\tau = C_{load}V_{dd}/I_{on}$). On-state current I_{on} is thus important parameter for CMOS. I_{on} is expressed as

$$I_{on} = WC_{ox}(V_{dd} - V_{th})v_s \tag{2.3}$$

where W is transistor width.

Again viewing the Table 2.1, the term $(V_{dd}-V_{th})v_s$ should be increased by a factor of κ. If we want to increase this term with decreasing V_{dd}, we should also decrease V_{th} generation by generation, moreover, v_s has a physical limit of saturation velocity v_{sat}. For Si CMOS, v_{sat} is about 1 x 10^7 m/s [9] and almost reaching this limit in the extremely scaled transistors. Gate capacitance C_{ox} has also increased generation by generation by decreasing the gate-oxide thickness t_{ox}. However, the t_{ox} value for the state-of-the-art CMOS is already below 1 nm. For the low-power CMOS, this value remains at around 2 nm. In such thickness ranges, tunnel leakage current through the oxide film significantly increases leakage power of the LSI chip, and thus the high-k technology has introduced into production to mitigate the gate leakage problem. Further scaling down the t_{ox} down to $\sim$0.5 nm is an important issue in order to continue the scaling.

On the other hand, we cannot decrease V_{th} without increasing standby leakage current. In the low-power CMOS design, off-state current I_{off} is an important parameter that should be kept as low as possible. I_{off} is usually defined as drain current I_d at $V_g = 0$ V. In MOSFETs, subthreshold current, that is, I_d at below V_{th}, exponentially changes. Subthreshold swing S is a parameter representing sharpness of I_d cut-off below V_{th} and defined as a gate voltage difference to increase I_d by one order of magnitude. Note that other leakage components such as gate leakage and drain-to-substrate leakage are omitted.

$$S = \log_{10}\frac{k_B T}{q}\left(1 + \frac{C_{dm}}{C_{ox}}\right) \approx 59.5 \times \left(1 + \frac{C_{dm}}{C_{ox}}\right)(\text{mV/decade at 300 K}) \tag{2.4}$$

where k_B, q, and C_{dm} denote Boltzmann constant, elementary charge, and maximum depletion capacitance below the channel, respectively.

The Dennald's scaling rule indicates that both C_{dm} and C_{ox} increases with miniaturization; however, S does not scale down because the ratio C_{dm}/C_{ox} remains non-zero positive value and is always higher than $\sim$60 mV/decade at room temperature. If we want to ensure the ratio between I_d at V_{th} and I_{off} at about five orders of magnitude, V_{th} should be higher than 0.3 V ($\sim$5 decades $\times$ 60 mV/decade if the C_{dm}/C_{ox} term is negligibly small). As just described, minimum V_{dd} should be at about 0.3 V or higher for the proper on–off ratio due to unscalable S and V_{th} values.

2.2.4 Issues for Operation-Voltage (V_{dd}) Reduction

As indicated in (2.2), decreasing V_{dd} down to the best value around 0.4 V is most effective way to decrease the power consumption of the LSIs. Before mentioning the lowest possible V_{dd} in the current CMOS technology, one important issue regarding circuit design is pointed out. There is a trade-off between performance and power. Although the best operating V_{dd} is around 0.4 V from the viewpoint of power efficiency, operation speed significantly decreases (MHz range or below) at V_{dd} around 0.4 V. Such a slow speed cannot match the requirements for most of applications and the technology can be used only in a niche market. In order to compromise the performance-power trade-off, various circuit technologies such as multi- or many-core technology and adaptive controlling including power-gating, clock-gating, DVFS, or body-bias technology are introduced. These technologies will be mentioned in Chap. 7.

Moving back to the lowest possible V_{dd} from the device viewpoint, there are obstacles in various aspects. In the integrated circuits, many transistors should operate synchronously. Circuit designers thus optimize timing margin in every block of circuits to ensure normal operation without any error. For the device operation, this timing margin reflects to voltage margin applied to transistors. Itoh analyzed the voltage margin for the scaled CMOS transistors that is used in logic circuits [10]. The main factors to determine minimum V_{dd} are; (i) V_{t0}; a margin to ensure proper I_{on}/I_{off} ratio that is mentioned above, (ii) ΔV; related to variability including both local and global factors, and (iii) ΔV_{ps}; a voltage drop through the power-distribution line and noise from the power supply and the power line. Among these factors, ΔV is most significant. These variability factors, especially for the local variability, inherently have a trend to increase with scaling down. As a result, this factor has become a majority term to determine minimum V_{dd} in the scaled devices.

2.2.5 Variability

Variability in the transistor characteristics is now a serious problem in CMOS LSIs. There are several variability factors such as V_{th}, on current, resistance, capacitance, and so on. Especially, the variability in V_{th} is very important because V_{th} is most influential to the transistor characteristics such as I_{on} and I_{off}. The V_{th} variation can be classified into four categories; global, local, systematic, and random. The former two are related to the spatial distribution of variability and the latter two are related to the regularity of the variation. Global variability generally comes from size variation such as gate length, channel width, film thickness, and/or impurity density. These factors can be decreased with technology development such as advanced process control and variability aware layout design. Generally speaking, the global variability is composed mostly of the systematic variation; thus, we can control these variability factors to a considerable extent. On the other hand, local variability is difficult to control because it is caused by the factor that is inherent to the device structure and also it has a random nature. The main cause for this local (and random) variation in the conventional bulk CMOS transistors is random-dopant fluctuation (RDF) [11]. This fluctuation is an electrical-potential fluctuation in the vicinity of the interface between the gate oxide and the silicon body where the inversion layer forms. As the potential is determined by the amount of charges from the impurities distributed in the silicon body, the potential is varied by the number and position of discrete impurity atoms. This fluctuation increases with decreasing the size of active area (product of channel length and width) in each transistor.

In the bulk MOS transistor with uniform impurity density in the silicon body, the standard deviation of V_{th} is written as:

$$\sigma V_{th} \propto \frac{t_{ox} N_A^{\frac{1}{4}}}{\sqrt{LW}} \tag{2.5}$$

where t_{ox}, N_A, L, and W are gate-oxide thickness, impurity density, channel length, and transistor width, respectively. The relationship between σV_{th} and transistor size is well known as shown in Fig. 2.5. This plot is known as the Pelgrom plot [12]. The slope of this plot is also known as the Pelgrom coefficient A_{Vt}. With scaling the minimum transistor area (LW) decreases, and the maximum value of $1/(LW)^{-1/2}$ (horizontal axis in Fig. 2.5) increases. In the 65- and 20-nm technology nodes, these maximum values are approximately 12.5 and 40 μm^{-1}, respectively, assuming $W_{minimum} = 1.5L$. Thus, A_{Vt} should be decreased with scaling in order not to increase maximum σV_{th}. As indicated in (2.5), A_{Vt} is proportional to t_{ox}. The A_{Vt} value thus decreases with scaling if t_{ox} is successfully scaled down according to the Dennald's scaling rule. However, more precisely, N_A also increases with scaling, and thus the effect of V_{th} variation has gradually increased so far. The past trend of A_{Vt} together with t_{ox} and V_{th} [13] is shown in Fig. 2.6. In the 65-nm bulk CMOS technology, the difference between highest and lowest V_{th} values of one

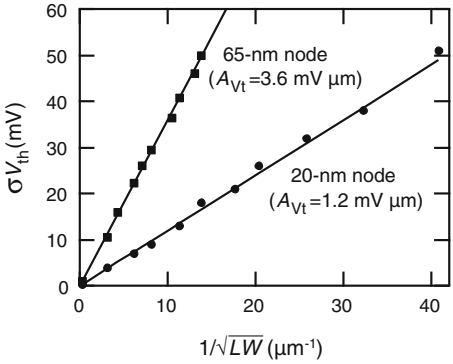

Fig. 2.5 Pelgrom plot (example)

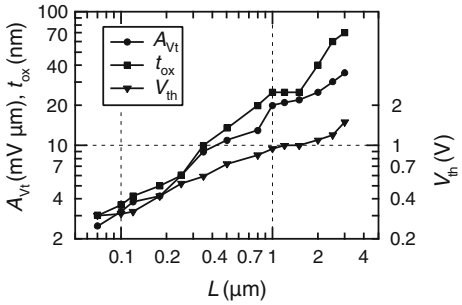

Fig. 2.6 Trend of A_{Vt}, t_{ox}, and V_{th}

million transistors (L and W is 60 and 140 nm, respectively) is about 0.45 and 0.37 V for NMOS and PMOS, respectively [14]. If we set the typical (median) value of V_{th} at 0.3 V, the highest V_{th} exceeds 0.5 V and the lowest is nearly 0 V. The local V_{th} variation is thus very serious especially when we want to decrease V_{dd} as low as possible.

2.3 Solutions for Low-Power CMOS Device

In this section, structural and material solutions for the low-power CMOS device are presented. High-k/metal-gate technology improves the trade-off between t_{ox} scaling and gate-leakage. Fully-depleted (FD) transistors such as FinFETs and planar FD silicon-on-insulator transistors (FDSOIs) improve the cut-off characteristics of extremely scaled transistors and also decrease V_{th} variability. And various mobility enhancement technologies that increase v_s in (2.3) are reviewed. These include strained silicon, high-mobility channel materials such as germanium, III–V compound semiconductor, and carbon. Finally, super-steep transistor technology to make a breakthrough against the unscalable S factor is shown.

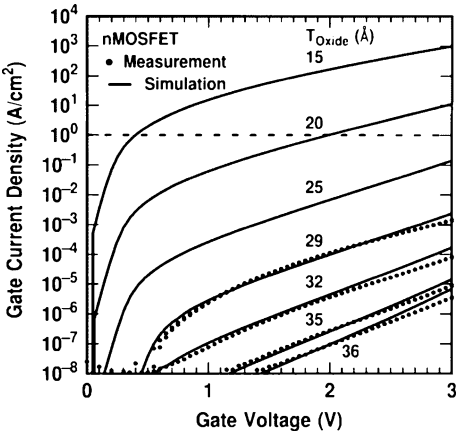

Fig. 2.7 Leakage current density of SiO_2 dielectric [15] © IEEE

2.3.1 High-k/Metal-Gate Technology

According to the scaling rule, gate-oxide thickness t_{ox} has gradually decreased as shown in Fig. 2.6. Already the t_{ox} value is less than 2 nm in the current technology. Note that the parameters corresponding to $L < 0.12$ μm in Fig. 2.6 are extrapolated ones. As shown in Fig. 2.7 [15], gate-leakage current density at $t_{ox} = 2$ nm is of the order of 1 A/μm^2. This level is high enough to increase the overall LSI-chip leakage current and intolerable for the low-power application. In such thicknesses less than 2 nm, the main leakage component is direct tunnel current. In order to reduce the direct tunnel leakage current, dielectric constant of the gate insulator should be increased because the thickness of the gate insulator can be increased at the same gate capacitance C_{ox}. Traditionally, SiO_2 film formed by thermal oxidation of silicon has been used for the gate insulator. Recently, the combination of SiO_2 and SiN or SiON has been used. From the 45-nm generation, "so-called" high-k materials have been introduced into production. "k" denotes relative dielectric constant.

2.3.1.1 High-k Material Selection

At the early phase of the high-k technology development, many oxide materials were investigated as a possible candidate, Al_2O_3, Ta_2O_3, ZrO_2, HfO_2, perovskite oxides such as $SrTiO_3$ and $LaAlO_3$, rare-earth oxides such as Y_2O_3 and La_2O_3, and various silicate materials combined with the above oxides and silicon. Important properties for the high-k gate insulator are large bandgap as well as large band discontinuity both from the conduction and valence band edges to decrease Fowler–Nordheim tunnel leakage, high dielectric breakdown strength, small interface-state density, small fixed and mobile charges inside and interface of the dielectric film, reactivity with silicon and with gate-electrode material, and so on [16]. After a lot

of efforts on the screening of dielectric materials, currently, hafnium silicon oxynitride (HfSiON) is widely used. Toward extremely thin equivalent oxide thickness (EOT) at around 0.5 nm, other candidate materials including rare earth oxides or silicates have been further investigated.

2.3.1.2 Issues and Solutions for High-k Technology

There are two major issues for the high-k technology. One is V_{th} controllability and the other is channel mobility degradation. In the conventional CMOS technology before the high-k era, V_{th} has been controlled only by the channel impurity density utilizing the conventional polysilicon gate electrodes. Simple replacement with the gate dielectric from SiO_2 to the high-k material resulted in significant increase in V_{th}. The main cause for this V_{th} shift is the Fermi-level pinning [17, 18] between the high-k material and the polysilicon. Mobility degradation is mainly caused by the increase in interface-state density, remote phonon scattering [19], and remote Coulomb scattering [20]. The degradation is more severe with thinner EOT because these scattering phenomena are strongly related to the distance between the inversion carrier (channel) and the scattering center (in dielectric or gate electrode).

To solve these problems, the metal gate technology has been developed. By a selection of gate materials both to have proper effective workfunction (EWF) and to reduce interface states and charges (including interface dipoles), V_{th}s both for NMOS and PMOS can be properly controlled and mobility can be significantly increased. However, these problems have not completely been solved yet by the current high-k/metal gate technology, the mobility degradation, and V_{th} controllability including thermal stability during the fabrication process further should be improved along with the EOT scaling.

2.3.1.3 Metal Gate Process

There are three types of metal gate (MG) processes; gate-first, gate-last, and fully silicided (FUSI) gate. Figure 2.8 shows schematic cross-sectional drawings of these processes. The gate-first process usually uses the metal inserted polysilicon stack (MIPS). Just upon the gate dielectric layer, a thin metal layer is stacked to control EWF, and the conventional polysilicon layer is formed on this layer. The gate-electrode etching process is the same as the conventional poly-silicon-gate process other than the additional metal-layer etching step. Thermal diffusion process to form source/drain region are done after the formation of the MIPS gate electrode. Thus, special cares to prevent metal contamination, metal diffusion, and de-tuning V_{th} due to EWF change by thermal treatment should be done during the process steps after the gate electrode formation. The gate-last process is also called as the damascene gate process. At first, the conventional polysilicon gate is formed, and the source and drain formation process is followed. After the

formation of an interlayer dielectric film covering gate, source, and drain electrodes, the surface of the gate electrodes are exposed usually by the chemical mechanical polishing (CMP) process. The polysilicon gate electrode and the dielectric layer are then removed to expose silicon surface. After that, the high-k layer and MG layer formation processes are followed. In some case, the high-k layer is formed before the poly-silicon layer formation, and only the gate electrode is replaced by the gate-last process. The advantage of this process over the gate-last process is stable EWF due to weaker thermal treatment for the gate electrode and dielectric. On the other hand, the process is more complex and it is more difficult to shorten the gate length than the gate-first process. The materials mainly used for the above processes are metallic compounds including titan or tantalum, such as TiN, TaN, TaSiN, or TaC. The fully silicided (FUSI) gate process converts the conventional polysilicon gate electrode to metal (usually nickel) silicide by a reaction between polysilicon and metal after the source/drain formation process. This process is very similar to the conventional self-aligned silicide (salicide) process. In the conventional process, only the top portion of the polysilicon gate electrode is silicided to reduce gate electrode resistance because metal silicides have small metallic resistivity. In the FUSI process, the poly-silicon gate electrode is thoroughly reacted to form metal silicide. This reaction stops at the interface between the polysilicon and the gate dielectric. This process is easier than the other two processes in terms of metal contamination; however, it is not widely used now because of the controllability of EWF due to the limitation in material selection and controllability of full silicidation. Currently, the gate-last process will be the first choice for the production fabrication.

2.3.2 Fully Depleted Transistors

Migration to the fully depleted (FD) transistors has been proceeding. Intel started to ship the 22-nm-generation chip in 2012 equipped with three-dimensional channel FD transistors, named tri-gate transistors [21]. Other CMOS device companies are also investigating variety of FD transistors in order to put them into production toward the 22–14-nm generations. This movement occurs mainly due to two reasons; the SCE immunity and local-variability problems in the conventional bulk transistors. In this section, various types of FD transistor structures are shown, and considerations on the SCE immunity, local variability, and back-gate-bias capability are mentioned.

2.3.2.1 FD Transistor Structures

In the conventional bulk transistors, part of the silicon body below the gate electrode is depleted under the subthreshold regime. Because this depletion layer lies between the source and drain electrodes, current is cut-off during the off state.

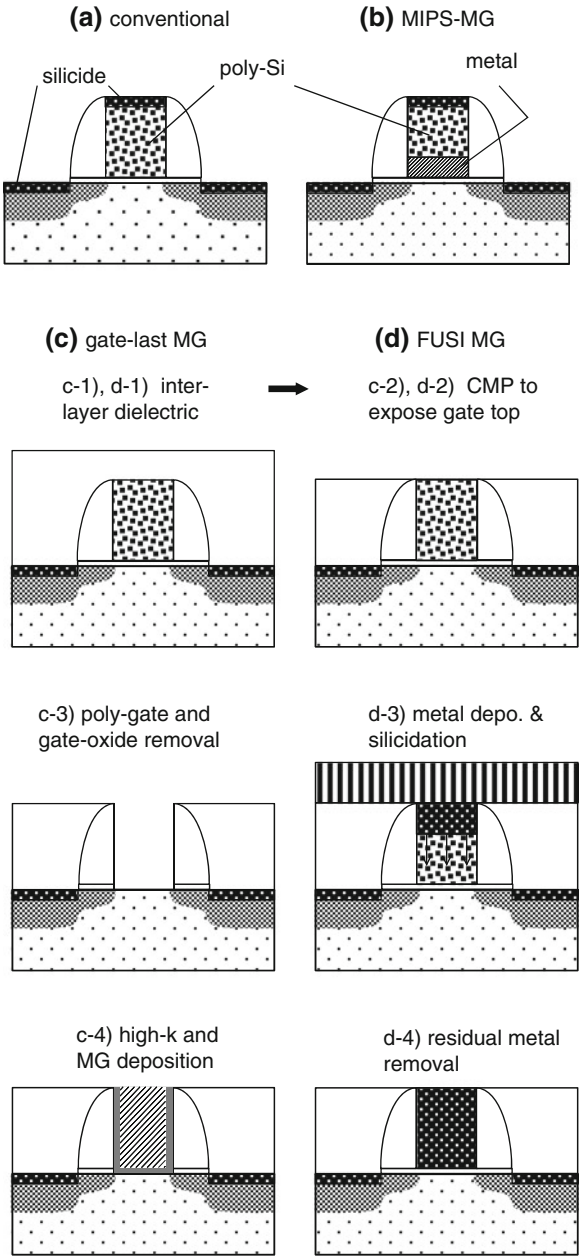

Fig. 2.8 Schematic cross-sectional drawings of metal-gate processes

In the FD transistors, a thin silicon body surrounded by insulator is used, and the silicon body is completely depleted under the subthreshold regime due to the electrostatic effect by the gate electrode(s). The SCE can be geometrically

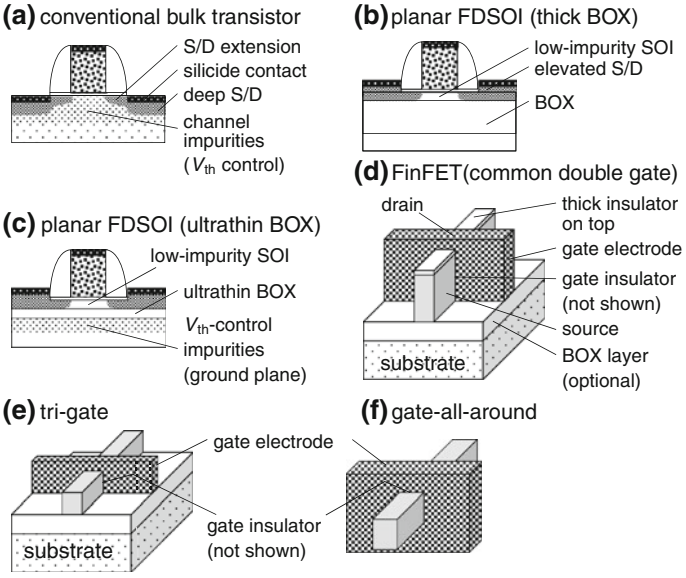

Fig. 2.9 Various FD transistor structures

controlled, that is, controlled by the size and shape of silicon body, gate electrode, and insulator.

There are several types of FD-transistor structures as shown in Fig. 2.9. Figure 2.9a shows the conventional bulk transistor for comparison. Figures 2.9b and c show the planar FDSOI transistors. Difference between these two structures is thickness of the buried oxide (BOX) layer. With the very thin BOX layer, the SCE immunity is improved, that is to be mentioned in the next section, and V_{th} can be controlled by the impurity density and bias voltage of the silicon substrate region just below the BOX layer. Figures 2.9d and e show FinFET and tri-gate transistors, respectively. In these structures, channel is formed on both sides of the surfaces of the silicon body, and in the tri-gate transistor channel is also formed on top surface. Note that the FinFET and the tri-gate transistors can be fabricated both from SOI and bulk wafer. In the latter case, bottom of the silicon body (fin) is not surrounded by insulator, thus impurities are doped at the bottom region to prevent leakage current from source to drain just like the conventional planar bulk transistors. Figure 2.9f shows the gate-all-around (GAA) structure. On all circumferences of the surface of the silicon body, channel is formed by the electrostatic action of the surrounded gate electrode. The silicon nanowire transistor is a type of the GAA transistor with extremely thin silicon body. From the planar Figure 2.9(a–c) structures to the three-dimensional Figure 2.9(d–f) structures approximately in this alphabetical order, the SCE can be improved due to stronger electrostatic controllability by the gate electrode and also areal density of transistors can be increased because effective channel width can be increased by increasing the fin height or by vertical stacking.

2.3.2.2 Short-Channel-Effect Immunities for Various Structures

The minimum possible gate length $L_{\min}$ for each transistor structure is an important feature for scaling. Many discussions for the $L_{\min}$ have been done from the old days [22].

As an indicator of $L_{\min}$, the parameter called as "natural length" λ is used. $L_{\min}$ is roughly 5–10 times larger than λ. Analytical expressions of λ for various structures are discussed in [23]. In the bulk transistor, λ is defined as,

$$\lambda = \sqrt{\frac{\kappa_s}{\kappa_{ox}} t_{ox} X_d} \tag{2.6}$$

where κ_s, κ_{ox}, and X_d denote relative dielectric constant for semiconductor (silicon) and gate insulator, and depletion depth of silicon body at the channel region averaged over from source side to drain side, respectively. In the planar FDSOI transistor with the thick BOX layer,

$$\lambda = \sqrt{\frac{\kappa_s}{\kappa_{ox}} t_{ox} t_s} \tag{2.7}$$

where t_s denotes thickness of silicon body. However, in the above expression, the electric field fringing effect from drain region through the BOX layer [24] is not taken into account; thus, in (2.7) λ can be underestimated especially for the structure with thick BOX. In order to include this effect, the same expression as the ground-plane transistor can be used that is also valid for the planar (independent) double-gate transistor.

$$\lambda = \sqrt{\frac{\kappa_s}{2\kappa_{ox}} \frac{t_{ox} t_s^{eq}}{1 + \frac{\kappa_s t_{ox}}{\kappa_{ox} t_s^{eq}}}} \tag{2.8}$$

where t_s^{eq} denotes equivalent silicon thickness to include the field fringing effect and t_{ox} denotes front gate oxide thickness. For the thin-BOX structure and the independent double-gate structure, t_s^{eq} is the thickness that represents series capacitance of the silicon body and back side insulator. The independent double gate means that the voltages of the gate electrodes can be independently (differently) controlled. In the common (tied) double-gate transistor like the FinFET, gate voltages of both sides are the same. In this structure λ can be expressed as

$$\lambda = \sqrt{\frac{\kappa_s}{2\kappa_{ox}} t_{ox} t_s} \tag{2.9}$$

In the GAA transistors with a square cross-section, in other words, the quadruple-gate transistors,

$$\lambda = \sqrt{\frac{\kappa_s}{4\kappa_{ox}} t_{ox} t_s} \tag{2.10}$$

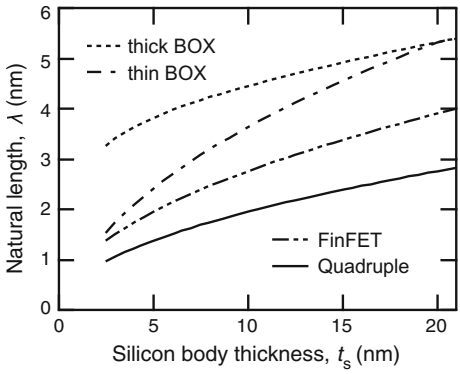

Fig. 2.10 Natural length λ of FD structures as a function of silicon-body thickness

Figure 2.10 compares λ of these typical structures as a function of silicon-body thickness t_s. In this example calculation, t_{ox} is 0.5 nm as an ultimate value. In order to include the electric field fringing effect, t_s^{eq} for the planar FDSOI transistor is set at the value 20 nm larger than the SOI thickness (t_{SOI}) when the BOX thickness t_{BOX} is 20 nm or thicker. In the case for the thin BOX structure, this additional thickness is limited to t_{BOX} that is the same as t_{SOI} considering scaling the both SOI and BOX layers. The calculation clearly indicates that the SCE can be improved by thinning the silicon body, decreasing the distance between the silicon body and the gate electrode, and increasing the number of gate electrodes or increasing the coverage surrounding the silicon body by the gate electrode. Even for the GAA structure (quadruple gate), L_{min} is not less than 5–10 nm ($\lambda = 1$ nm) for $t_s = 4$ nm.

If t_s is thinned down below $\sim$5 nm, quantum mechanical effect (QME) and mobility degradation become evident. QME induces increase in V_{th} due to discretization of energy levels [25] and V_{th} variation sensitivity over t_s fluctuation. Mobility significantly decreases mainly due to increasing interface roughness scattering, but a modest increase in mobility occurs due to decrease in phonon scattering caused by the subband structure change at t_s around 3 nm [26]. By considering the above thinnest t_s, the minimum gate length among all the above transistor structures is estimated to be about 10 nm or a little shorter. The ITRS roadmap 2010 revised version [5] predicts the gate length for logic transistors to be 7.4–7.9 nm in 2024 that is thought to be the ultimate small size.

2.3.2.3 Local Variability Reduction with FD Structures

As indicated previously, the local variability in the conventional bulk transistor is mainly determined by RDF. The analytical expression of the Pelgrom coefficient A_{Vt} determined by RDF for the bulk transistor with uniform impurity profile is [27],

$$A_{Vt} = \frac{q}{C_{ox}} \sqrt{\frac{N_A X_d}{3}} \tag{2.11}$$

Note that X_d is determined by N_A and the above formula is consistent with (2.5).

In the planar FD transistors, the expression of the RDF-related variability is similar to (2.11), simply replacing X_d to t_s. In the FD structures, it is no more needed to increase N_A to suppress SCE because of strong electrostatic controllability from the gate electrode(s). Thus A_{Vt} can decrease by decreasing both N_A and t_s. In the FD transistors other than the thin-BOX type, V_{th} is controlled by EWF of the gate electrode taking advantage of the high-k/metal-gate technology and not by impurities. Thus, contributions from other variation sources such as line-edge roughness and polycrystalline granularity of the gate electrode can increase [11]. In the thin-BOX-type planar FD transistor in which V_{th} can be controlled by impurities below the BOX layer; however, RDF from these impurities is negligibly small because impurity atoms are distant from the gate electrode [28, 29].

From the viewpoint of the circuit operation, reducing the V_{th} variation is not only the requirement that should be satisfied for low-voltage and low-power. On-state current, drain-induced barrier lowering (DIBL), and even random telegraph noise (RTN) will be significant and the FD structures can also improve these factors [30].

2.3.2.4 Considerations on V_{th} Controllability

V_{th} in the FD transistors other than the thin-BOX type is controlled by EWF of the gate electrode as mentioned above. In other words, V_{th} controllability for these structures is not so flexible compared to the conventional bulk transistors. For the low-power circuit design, multiple V_{th} is mandatory in order to satisfy the trade-off between power and performance. The operation energy per instruction E can be decreased by decreasing V_{th} when the activity a is high and the circuit should be operated faster, and vice versa. Thus, transistors with different V_{th} levels should be used in each circuit, that is, low V_{th} transistors are used only in the critical path that determines the overall circuit speed. Usually, two or three types of V_{th} levels are employed in the current low-power CMOS LSI. The ion implantation of impurities is an easy way to control V_{th} and widely used in the industry; however, this convenient technique cannot be used for V_{th} control of the FD transistors except for the case that EWF is controlled by implanting some elements into the gate electrode or the gate dielectric film. Developing a cost-effective technique to control V_{th} is thus an important issue for these transistors.

A time-to-time and/or area-to-area adaptive control of V_{th} is another important feature in the low-power circuit technology to decrease E. Adaptive body-bias (ABB) technique can be used for this purpose, that is, V_{th} of transistors is changed by applying body-bias voltage below the channel region. In the bulk CMOS technology, ABB was partly used [31]; however, it is difficult to utilize in the current scaled bulk CMOS technology because of significant increase in

drain-substrate leakage. The thin-BOX type is the only structure that ABB can be used among the transistor structures for the scaled CMOS [32] because it has proper ABB controllability and its drain-substrate leakage is prevented by the BOX layer. For the FinFET structure, ABB also can be possible but the controllability is rather small [33].

In the independent double-gate transistor like the thin-BOX type, there is an important trade-off relationship between ABB controllability and S that should be considered in the transistor and circuit design. Body-bias coefficient γ is a parameter to determine the sensitivity of V_{th} per body-bias voltage V_{b}, that is $\partial V_{\mathrm{th}}/\partial V_{\mathrm{b}}$ (at $V_{\mathrm{b}} = 0$). γ for the bulk and independent double-gate transistors are expressed as

$$\gamma = \frac{3t_{\mathrm{ox}}}{X_{\mathrm{d}}} = \frac{C_{\mathrm{dm}}}{C_{\mathrm{ox}}}\,(\mathrm{bulk})\ \text{and} \tag{2.12}$$

$$\gamma = \frac{1}{C_{\mathrm{ox}}^{\mathrm{front}}}\left(\frac{C_{\mathrm{ox}}^{\mathrm{back}} C_{\mathrm{s}}}{C_{\mathrm{ox}}^{\mathrm{back}} + C_{\mathrm{s}}}\right)(\mathrm{double\ gate}), \tag{2.13}$$

where $C_{\mathrm{ox}}^{\mathrm{front}}$, $C_{\mathrm{ox}}^{\mathrm{back}}$, and C_{s} are front-, back-gate oxide capacitances, and silicon-body capacitance, respectively. By comparing (2.4) and (2.12), it is clear that S increases with increasing γ. This relationship is the same for the double gate simply replacing the respective term in (2.4) by that in (2.13). If we want to largely change V_{th} with smaller V_{b}, that is, larger γ with larger $C_{\mathrm{ox}}^{\mathrm{back}}$, we must accept the increase in S. These parameters thus should be designed taking the circuit operation mode, such as ABB control range, I_{off} at both standby and active state and so on, into account. In the common double-gate transistor, S is around 60 mV/decade at room temperature whereas γ is very small.

2.3.3 Mobility Enhancement Technologies

In order to increase the circuit performance, transistors should be operated faster, that is, higher I_{on} in the logic circuits as a primary requirement. When V_{dd} and V_{th} are optimized, the other key parameter to increase I_{on} is the carrier velocity v_{s} in (2.3). In the scaled MOS transistors, v_{s} is strongly depended on the injection velocity of carriers from the source electrode. This means that effective mass of carriers should be lighter and higher mobility is preferred because of the relationship,

$$\mu = \frac{q\tau}{m^*} \tag{2.14}$$

where μ, τ, and m^* denote mobility, average scattering period (relaxation time), and effective mass in crystal, respectively.

In this section, various mobility enhancement techniques regarding the structure and material are reviewed.

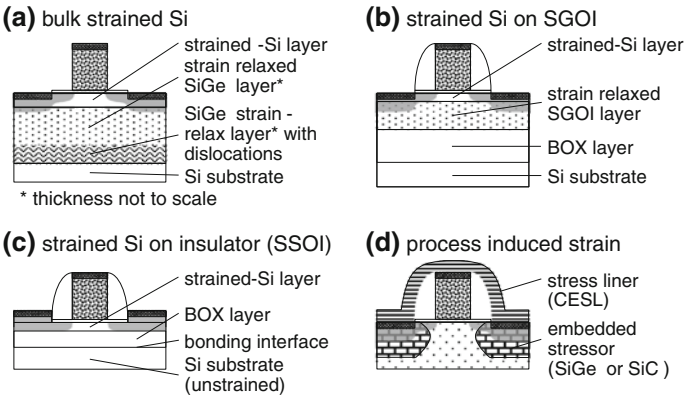

Fig. 2.11 Typical strained-silicon structures

2.3.3.1 Strained Silicon

Carrier mobility is varied by applying stress. This phenomenon is known as the piezoresistance effect [34]. There are a variety of techniques applying stress to the silicon channel. There are roughly two categories on these techniques; a substrate technique and a process technique. Typical structures are schematically shown in Fig. 2.11.

The former has been studied from the early stage of the strained-silicon research [35]. Using the Si/SiGe-heterostructure technology, a strained-Si top layer as a channel layer is formed on a SiGe pseudo substrate. The pseudo substrate can be fabricated by growing a thick SiGe layer on a Si substrate. There is lattice mismatch between SiGe and Si and this lattice mismatch is the cause of strain. When a SiGe layer is grown on a Si substrate, the SiGe layer is compressively strained. With increasing the SiGe layer thickness, the strain becomes to be gradually relaxed with introducing dislocations at near the interface between the SiGe layer and the Si substrate. On the contrary, a Si layer is tensilely strained when it is grown on the relaxed SiGe layer. The pseudo substrate also can be fabricated from a SOI substrate named the SiGe-on-insulator (SGOI) substrate [36]. Further, by using the wafer bonding technique, the strained-Si layer directly bonded on the BOX layer (SSOI) can be fabricated [37]. These substrates can apply biaxially isotropic strain to the topmost Si layer. The intensity of strain is about 1 % for the Ge content of 24 % in the SiGe pseudo substrate because the lattice mismatch between SiGe and Si is 4.2 %. Electron mobility can increase to about twice at 1 % strain; however, hole mobility cannot increase at this strain level and slightly increase at the strain intensity over 1 % [38]. Because of this imbalance in the mobility enhancement and of substrate quality (dislocation density problem) and cost, the substrate technique is not used now for the production process of CMOS.

The latter, the process technique, is currently widely used in the production process. The typical technique, the easiest one, is a stress-liner technique [39, 40]. In the conventional CMOS process, a contact-etch-stop layer (CESL) is deposited on the active area including gate, source, and drain electrodes of transistors before the contact-hole etching process connecting to the first metal wiring layer. A silicon nitride film can be used as the CESL layer, and the stress of this film can be widely controlled by changing the deposition conditions and thicknesses. When the CESL layer is deposited, because of the height difference between the gate and the source/drain electrodes, considerable amount of strain can be induced near the channel region of transistors. In order to increase mobility, tensile and compressive strain is preferred for NMOS and PMOS, respectively. Thus, different CESL films with proper strain are independently formed on NMOS and PMOS active regions. Unlike the substrate technique, the strain induced by the process technique tends to be anisotropic. Both electron and hole mobilities can be enhanced by a proper control of strain anisotropy. Another typical method to induce strain is the embedded SiGe technique. During the source/drain formation process, the source/drain region is selectively etched and the SiGe layer is then selectively deposited by an epitaxial growth technique. This embedded SiGe layer induces uniaxial compressive strain to the channel region and increases hole mobility. This technique is first introduced into production by Intel from the 90-nm generation [41]. In order to induce tensile strain to increase electron mobility, embedded SiC technique is also used [42]. Note that SiC is not silicon carbide but carbon-doped silicon with carbon content of less than a few percent. Other process techniques such as shallow-trench-isolation (STI) induced strain and gate-electrode induced strain are partly used. Combinations of these techniques are also carried out to maximize the strain effect. By these process techniques, electron mobility can increase by about 50 % and hole mobility can increase to about two or three times [38]. Due to this huge increase in hole mobility, I_{on} of the PMOS transistor significantly increases and reaches to almost the same value as NMOS in the state-of-the-art CMOS technology. Note that I_{on} ratio between NMOS and PMOS had been approximately two in the classical CMOS technologies.

Mechanisms of mobility enhancement by strain have been thoroughly studied [38, 43]. In short, mobility increases by two major causes; decrease in effective mass along the conduction direction and decrease in band-to-band scattering, both due to sub-band-structure change induced by strain.

2.3.3.2 Ge and III–V FETs

Novel channel materials with higher mobility beyond the strained silicon have been extensively studied aiming to make a debut in the extremely scaled era of the CMOS technology. Germanium and III–V compound semiconductors are candidates for the novel channel material. Important requirements for the channel material are small effective mass m^* that is related to mobility as shown in (2.14), higher density of states (DOS) to increase carrier density injected from the source

Table 2.2 Properties of group IV and III–V semiconductors

Material	E_g (eV)	μ_n (cm^2/Vs)	μ_p (cm^2/Vs)	m_n^*/m_0	m_p^*/m_0
Si	1.12	1450	500	0.98[a], 0.19[b]	0.16[c], 0.49[d]
Ge	0.66	3900	1900	1.64[a], 0.082[b]	0.04[c], 0.28[d]
GaAs	1.42	8000	400	0.063	0.076[c], 0.5[d]
GaN	3.44	400	10	0.27	0.80
GaP	2.26	110	75	0.82	0.60
GaSb	0.72	5000	850	0.042	0.40
InAs	0.36	33000	460	0.023	0.40
InP	1.35	4600	150	0.077	0.64
InSb	0.17	80000	1250	0.0145	0.40

[a] long-axis direction, [b] short-axis direction, [c] light hole, [d] heavy hole

region, and larger band gap E_g for a better cut-off characteristic. These channel materials can be used in the extremely scaled transistors in which the carrier transport is quasi ballistic or ballistic that carriers injected from the source region run through the channel without any scattering into the drain region. In such a transport mode, both higher injection velocity and carrier density are essential for higher on-state current. Small m^* increases the injection velocity; however, there is a trade-off relationship between m^* and DOS [44]. Table 2.2 summarizes the important parameters of the candidate materials. Germanium has the highest hole mobility among these materials; however, small E_g may increase off-state current especially at elevated temperatures. By contrast, III–V compounds have very high electron mobility but modest hole mobility. To make a compromise, the hybrid structure employing III–V (InGaAs) and germanium for NMOS and PMOS, respectively, is proposed [45]. The important issue for the transistor fabrication is quality of the gate insulator/semiconductor interface. Many efforts have been done to improve the interface quality. It is very difficult to improve the interface quality because the insulator thickness should be extremely thin to match the requirement for the ultimately scaled era. Recently, the gate-oxide technology for Ge transistors has been rapidly improving and very high mobilities both for electron and hole have been observed even at EOT around 1 nm [46]. The goal will be ultrahigh mobility at EOT $\sim$0.5 nm. A combination of selecting a proper material, applying the straining technique like silicon, low-resistivity electrode including metal source/drain, and three-dimensional structure like GAA or nanowire to improve the cut-off characteristic will be a key technology to implement these new channel materials. As for the lower E_g materials, high-frequency transistors for milliwave or higher also will be a promising application field.

2.3.3.3 CNT and Graphene FETs

Carbon is also a promising channel material like Ge and III-V compound semiconductor. Carbon nanotube (CNT) and graphene are known to have extremely high mobility and inherent ballistic transport nature. Unlike most of III-V

materials, carbon has high availability (high Clarke number) and is less toxic; however, nanocarbon materials lsuch as CNT should be cautious to handle because they may have a cancer-causing property. CNT has a very thin hollow cylindrical structure of about 1-nm diameter and this nanowire like structure itself is suitable for the extremely scaled transistor. There are two main issues for CNT as a channel material. CNT exhibits semiconducting and metallic conduction depending on the arrangement of six-membered ring structures (chirality) and in general a mixture of both phases is obtained as synthesized. Thus, controlling proper E_g for the logic transistor is difficult. The other issue is a controlled arrangement of nanotubes during the transistor fabrication process. Graphene is a single-sheet state of graphite with sp^2-bonding configuration. Like CNT, graphene also has extremely high mobility [47]. Because it has a sheet-type structure unlike CNT, it might be better for the planar integrated circuit fabrication than CNT. Graphene is known as a zero-gap semiconductor, and thus it is not suitable for logic devices that require high cut-off characteristic. However, with decreasing the width of the graphene sheet down to less than a few nanometers, called graphene nano ribbon, E_g increases to a just enough level that can be used as a logic transistor. The issues for graphene are similar to those for CNT. Electrical property control for graphene is related to the fabrication process, size, edge shape, and so on. In the basic research of graphene, the "scotch-tape" method is often used to peel off a graphene monolayer from graphite. Processes that can be implemented in the device fabrication, such as reduction of SiC layer or chemical vapor deposition, are still in an early stage of development. Although CNT and graphene tend to have smaller E_g, their high mobility might be useful for rf applications [48]. Another promising application for CNT and graphene is wiring in the integrated circuit because of their high current density and thermal conductivity.

2.3.4 Another Important Approach for Low Power: Super-Steep Transistor

As discussed in the previous section, reduction of V_{dd} is essential for low power. However, the subthreshold cut-off characteristic represented by S limits the minimum V_{th} in order to ensure a proper on–off ratio. The minimum value of S is limited to be 60 mV/decade at room temperature in the conventional MOS transistors including both planar and three-dimensional structures. Super-steep cut-off transistors that exhibit S less than the above value might break down the above limitation in V_{dd}. There are several proposals for the super-steep cut-off transistors. I-MOS [49] utilizes impact ionization at the drain side. This transistor has a p-i-n structure from source to drain region and the gate electrode is placed on a part of the i-region. The I-MOS turns on current by triggering with the avalanche multiplication process caused by the impact ionization. This multiplication process induces very sharp switching of the transistor current and realizes very small

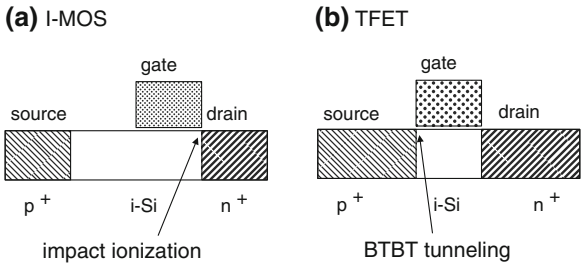

Fig. 2.12 Schematic cross-section of I-MOS and TFET

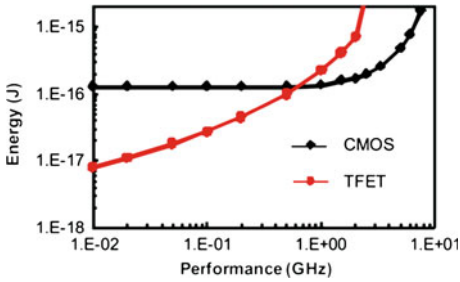

Fig. 2.13 Energy per operation E for conventional FET and TFET [51] © IEEE

S value less than 10 mV/decade at 300 K. Tunnel field-effect transistor (TFET) is another promising candidate for the super-steep transistor. The schematic cross-section of TFET is shown together with I-MOS in Fig. 2.12. TFET also has a p-i-n structure and band-to-band tunneling (BTBT) current near the source side can be controlled by the gate voltage [50]. This tunnel current can be modulated more abruptly than thermionic emission current of the conventional MOS transistor. The S value can decrease to about 50 mV/decade or less at room temperature. The impact of smaller S in terms of low power is decreasing E less than the conventional MOS transistors as shown in Fig. 2.13 [51, 52]. Decrease in E about one order of magnitude is very promising; however, there are still many issues in order to implement these transistors for integration. There are other candidate structures such as ferroelectric-gate transistors [53] and the conventional planar FD transistors but utilizing single latch behavior [54]. At present, it is unclear which type of transistor can be used. Further research on various types of super-steep transistors is strongly needed.

References

1. Moore GE (1965) Cramming more components onto integrated circuits. Electronics 38(8):114–117
2. http://www.intel.com/about/companyinfo/museum/exhibits/moore.htm?wapkw=moore
3. Dennard RH, Gaensslen FH, Yu HN, Rideout VL, Bassous E, LeBlanc AR (1974) Design of ion-implanted MOSFETs with very small physical dimensions. IEEE J Solid State Circuits 9(5):256–268

4. http://www.sia-online.org/ (roadmap currently unavailable)
5. http://www.itrs.net/
6. Chandrakasan AP, Daly DC, Finchelstein DF, Kwong J, Ramadass YK, Sinangil ME, Sze V, Verma N (2010) Technologies for ultradynamic voltage scaling. Proc IEEE 98(2):191–214
7. Zhai B, Nazhandali L, Olson J, Reeves A, Minuth M, Helfand R, Pant S, Blaauw D, Austin T (2006) A 2.60pJ/Inst subthreshold sensor processor for optimal energy efficiency. In: IEEE VLSI 2006 circuit, pp 154–155
8. Nakai M, Akui S, Seno K, Meguro T, Seki T, Kondo T, Hashiguchi A, Kawahara H, Kumano K, Shimura M (2005) Dynamic voltage and frequency management for a low-power embedded microprocessor. IEEE J Solid State Circuits 40(1):28–35
9. Sze SM, Ng KK (2007) Physics of semiconductor devices, 3rd edn. Wiley, New York, pp 35–39
10. Itoh K (2009) Adaptive circuits for the 0.5-V nanoscale CMOS era. In: IEEE international solid-state circuits conference, pp 14–20
11. Asenov A (2010) Statistical nano CMOS variability and its impact on SRAM. In: Singhee A, Rutenbar RA (eds) extreme statistics in nanoscale memory design, integrated circuits and systems, chapter 3. Springer Science + Business Media, LLC
12. Pelgrom MJM, Duinmaijer ACJ, Welbers APG (1989) Matching properties of MOS transistors. IEEE J Solid State Circuits 24(5):1433–1439
13. Bult K (2000) Analog design in deep sub-micron CMOS. In Proceedings of 26th ESSCIRC, pp 126–132
14. Tsunomura T, Nishida A, Yano F, Putra AT, Takeuchi K, Inaba S, Kamohara S, Terada K, Hiramoto T, Mogami T (2008) Analyses of 5σ Vth fluctuation in 65 nm-MOSFETs using Takeuchi plot. In: IEEE VLSI 2008 technology, pp 156–157
15. Lo S-H, Buchanan DA, Taur Y, Wang W (1997) Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's. IEEE Electron Device Lett 18(5):209–211
16. Robertson J (2005) Interfaces and defects of high-K oxides on silicon. Solid State Electron 49(3):283–293
17. Hobbs CC, Fonseca LRC, Knizhnik A, Dhandapani V, Samavedam SB, Taylor WJ, Grant JM, Dip LG, Triyoso DH, Hegde RI, Gilmer DC, Garcia R, Roan D, Lovejoy ML, Rai RS, Hebert EA, Tseng H-H, Anderson SGH, White BE, Tobin PJ (2004) Fermi-level pinning at the polysilicon/metal oxide interface—part I. IEEE Trans Electron Devices 51(6):971–977
18. Shiraishi K, Yamada K, Torii K, Akasaka Y, Nakajima K, Konno M, Chikyow T, Kitajima H, Arikado T (2004) Oxygen vacancy induced substantial threshold voltage shifts in the Hf-based high-K MISFET with p^+ poly-Si Gates—a theoretical approach. Jpn J Appl Phys 43(11A):L 1413–L 1415
19. Chau R, Datta S, Doczy M, Doyle B, Kavalieros J, Metz M (2004) High-k/metal-gate stack and its MOSFET characteristics. IEEE Electron Device Lett 25(6):408–410
20. Saito S, Torii K, Hiratani M, Onai T (2002) Improved theory for remote-charge-scattering-limited mobility in metal—oxide—semiconductor transistors. Appl Phys Lett 81(13):2391–2393
21. http://www.intel.com/content/www/us/en/silicon-innovations/intel-22nm-technology.html
22. Brews JR, Fichtner W, Nicollian EH, Sze SM (1980) Generalized guide for MOSFET miniaturization. IEEE Electron Device Lett 1(1):2–4
23. Yan R-H, Ourmazd A, Lee KF (1992) Scaling the Si MOSFET: from bulk to SOI to bulk. IEEE Trans Electron Devices 39(7):1704–1710
24. Yeh PC, Fossum JG (1995) Physical subthreshold MOSFET modeling applied to viable design of deep-submicrometer fully depleted SO1 low-voltage CMOS technology. IEEE Trans Electron Devices 42(9):1605–1613
25. Colinge JP (ed) (2008) FinFETs and other multi-gate transistors. Springer Science + Business Media, LLC, pp 28–36
26. Uchida K, Watanabe H, Kinoshita A, Koga J, Numata T, Takagi S (2002) Experimental study on carrier transport mechanism in ultrathin-body SOI n-and p-MOSFETs with SOI thickness less than 5 nm. In: IEDM technical digest, pp 47–50

27. Taur Y, Ning TH (1998) Fundamentals of modern VLSI devices. Cambridge University Press, Cambridge
28. Takeuchi K, Tatsumi T, Furukawa A (1997) Channel engineering for the reduction of random-dopant-placement-induced threshold voltage fluctuation. In: IEDM technical digest, pp 841–844
29. Ohtou T, Sugii N, Hiramoto T (2007) Impact of parameter variations and random dopant fluctuations on short-channel fully depleted SOI MOSFETs with extremely thin BOX. IEEE Electron Device Lett 28(8):740–742
30. Hiramoto T, Mizutani T, Kumar A, Nishida A, Tsunomura T, Inaba S, Takeuchi K, Kamohara S, Mogami T (2010) Suppression of DIBL and current-onset voltage variability in intrinsic channel fully depleted SOI MOSFETs. In: IEEE 2010 SOI Conference, pp 170–171
31. Miyazaki M, Ono G, Ishibashi K (2002) A 1.2-GIPS/W microprocessor using speed-adaptive threshold-voltage CMOS with forward bias. IEEE J Solid State Circuits 37(2):210–217
32. Tsuchiya R, Horiuchi M, Kimura S, Yamaoka M, Kawahara T, Maegawa S, Ipposhi T, Ohji Y, Matsuoka H (2004) Silicon on thin BOX: a new paradigm of the CMOSFET for low-power and high-performance application featuring wide-range back-bias control. In: IEDM technical digest, pp 631–634
33. Nagumo T, Hiramoto T (2006) Design Guideline of Multi-Gate MOSFETs With Substrate-Bias Control. IEEE Trans Electron Devices 53(12):3025–3031
34. Smith CS (1954) Piezoresistance effect in germanium and silicon. Phys Rev 94(1):42–49
35. Welser J, Hoyt JL, Gibbons JF (1992) NMOS and PMOS transistors fabricated in strained silicon/relaxed silicon-germanium structures. In: IEDM technical digest, pp 1000–1002
36. Tezuka T, Sugiyama N, Takagi S, Kawakubo S (2002) Dislocation-free formation of relaxed SiGe-on-insulator layers. Appl Phys Lett 80(19):3560–3562
37. Rim K, Chan K, Shi L, Boyd D, Ott J, Klymko N, Cardone F, Tai L, Koester S, Cobb M, Canaperi D, To B, Duch E, Babich I, Carruthers R, Saunders P, Walker G, Zhang Y, Steen M, Ieong M (2003) Fabrication and mobility characteristics of ultra-thin strained Si directly on insulator (SSDOI) MOSFETs. In: IEDM technical digest, pp 47–52
38. Thompson SE, Suthram S, Sun Y, Sun G, Parthasarathy S, Chu M, Nishida T (2006) Future of strained si/semiconductors in nanoscale MOSFETs. In: IEDM technical digest, pp 681–684
39. Ootsuka F, Wakahara S, Ichinose K, Honzawa A, Wada S, Sato H, Ando T, Ohta H, Watanabe K, Onai T (2000) A highly dense, high-performance 130 nm node CMOS technology for large scale system-on-a-chip applications. In: IEDM technical digest, pp 575–578
40. Ito S, Namba H, Yamaguchi K, Hirata T, Ando K, Koyama S, Kuroki S, Ikezawa N, Suzuki T, Saitoh T, Horiuchi T (2000) Mechanical stress effect of etch-stop Nitride and its impact on deep submicron transistor design. In: IEDM technical digest, pp 247–250
41. Ghani T, Armstrong M, Auth C, Bost M, Charvat P, Glass G, Hoffmann T, Johnson K, Kenyon C, Klaus J, McIntyre B, Mistry K, Murthy A, Sandford J, Silberstein M, Sivakumar S, Smith P, Zawadzki K, Thompson S, Bohr M (2003) A 90 nm high volume manufacturing logic technology featuring novel 45 nm gate length strained silicon CMOS transistors. In: IEDM technical digest, pp 978–980
42. Ang KW, Chui KJ, Bliznetsov V, Du A, Balasubramanian N, Li MF, Samudra G, Yeo Y (2004) Enhanced performance in 50 nm N-MOSFETs with silicon-carbon source/drain regions. In: IEDM technical digest, pp 1069–1071
43. Takagi S, Hoyt J, Welser J, Gibbons J (1996) Comparative study of phonon-limited mobility of two-dimensional electrons in strained and unstrained-Si metal-oxide-semiconductor field-effect transistors. J Appl Phys 80(3):1567–1577
44. Fischetti MV, Laux SE (1991) Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. II. Submicrometer MOSFET's. IEEE Trans Electron Devices 38(3):650–660

45. Heyns M, Alian A, Brammertz G, Caymax M, Chang YC, Chu LK, De Jaeger B, Eneman G, Gencarelli F, Groeseneken G, Hellings G, Hikavyy A, Hoffmann TY, Houssa M, Huyghebaert C, Leonelli D, Lin D, Loo R, Magnus W, Merckling C, Meuris M, Mitard J, Nyns L, Orzali T, Rooyackers R, Sioncke S, Soree B, Sun X, Vandooren A, Verhulst AS, Vincent B, Waldron N, Wang G, Wang WE, Witters L (2011) Advancing CMOS beyond the Si roadmap with Ge and III/V devices. In: IEDM technical digest, pp 299–302
46. Zhang R, Taoka N, Huang P-C, Takenaka M, Takagi S (2011) 1-nm-thick EOT high mobility Ge n- and p-MOSFETs with ultrathin GeO_x/Ge MOS interfaces fabricated by plasma post oxidation. In: IEDM technical digest, pp 642–645
47. Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA (2004) Electric field effect in atomically thin carbon films. Science 306(5696):666–669
48. Han S-J, Valdes-Garcia A, Bol AA, Franklin AD, Farmer D, Kratschmer E, Jenkins KA, Haensch W (2011) Graphene technology with inverted-T gate and RF passives on 200 mm platform. In: IEDM technical digest, pp 19–22
49. Gopalakrishnan K, Griffin PB, Plummer JD (2005) Impact ionization MOS (I-MOS)-Part I: device and circuit simulations. IEEE Trans Electron Devices 52(1):69–76
50. Aydin C, Zaslavsky A, Luryi S, Cristoloveanu S, Mariolle D, Fraboulet D, Deleonibus S (2004) Lateral interband tunneling transistor in silicon-on-insulator. Appl Phys Lett 84(10):1780–1782
51. Kam H, King-Liu T-J, Alon E, Horowitz M (2008) Circuit-level requirements for MOSFET-replacement devices. In: IEDM technical digest, p 427
52. Kam H, Liu T-JK, Alon E (2012) Design requirements for steeply switching logic devices. IEEE Trans Electron Devices 59(2):326–334
53. Salvatore GA, Bouvet D, Ionescu AM (2011) Demonstration of subthrehold swing smaller than 60 mV/decade in Fe-FET with P(VDF-TrFE)/SiO_2 gate stack. In: IEDM technical digest, pp 167–170
54. Fossum JG, Sundaresan R, Matloubian M (1987) Anomalous subthreshold current-voltage characteristics of n-channel SOI MOSFET's. IEEE Electron Device Lett 8(11):544–546