Eur. Phys. J. Appl. Phys. **36**, 197–214 (2007)
DOI: 10.1051/epjap:2006158

THE EUROPEAN PHYSICAL JOURNAL
APPLIED PHYSICS

Physical and technological limitations of NanoCMOS devices to the end of the roadmap and beyond

S. Deleonibus[a]

CEA-LETI/NANOTEC, CEA-Grenoble, 17 rue des Martyrs, 38054 Grenoble Cedex 09, France

Received: 10 July 2006 / Received in final form: 18 September 2006 / Accepted: 2 October 2006
Published online: 10 January 2007 –

Abstract. Since the end of the last millenium, the microelectronics industry has been facing new issues as far as CMOS devices scaling is concerned. Linear scaling will be possible in the future if new materials are introduced in CMOS device structures or if new device architectures are implemented. Innovations in the electronics history have been possible because of the strong association between devices and materials research. The demand for low voltage, low power and high performance are the great challenges for the engineering of sub 50 nm gate length CMOS devices. Functional CMOS devices in the range of 5 nm channel length have been demonstrated. The alternative architectures allowing to increase devices drivability and reduce power consumption are reviewed. The issues in the field of gate stack, channel, substrate, as well as source and drain engineering are addressed. HiK gate dielectric and metal gate are among the most strategic options to consider for power consumption and low supply voltage management. By introducing new materials (Ge, diamond/graphite carbon, HiK, ...), Si based CMOS will be scaled beyond the ITRS as the future System-on-Chip Platform integrating also new disruptive devices. For example, the association of C-diamond with HiK, as a combination for new functionalized Buried Insulators, will bring new ways of improving short channel effects and suppress self-heating. Because of the low parasitics required to obtain high performance circuits, alternative devices will hardly compete against logic CMOS.

PACS. 85.30.De Semiconductor-device characterization, design, and modeling – 85.35.-p Nanoelectronic devices – 85.40.-e Microelectronics: LSI, VLSI, ULSI; integrated circuit fabrication technology

1 International technology roadmap of semiconductors acceleration and issues

Since 1994, the International Technology Roadmap for Semiconductor (ITRS) [1] (Fig. 1) has accelerated the scaling of CMOS devices to lower dimensions continuously despite the difficulties that appear in device optimization.

However, uncertainties about lithography, economics and physical limitations will probably slow down the evolution. For the first time, since the introduction of poly gate in CMOS devices process, showstoppers other than lithography appear to be attracting special attention and could require some breakthrough or evolution if we want to continue scaling at the same rate. Design could also be affected by this evolution.

Which are the main showstoppers for CMOS scaling? In this paper, we focus on the possible solutions and guidelines for research in the next years in order to propose solutions to enhance CMOS performance before we need to skip to alternative devices. In other words, how can we offer a second life to CMOS?

To that respect, the roadmap distinguishes today three types of products: High Performance (HP) (Fig. 1), Low Operating Power (LOP) and Low Standby Power (LSTP) devices. In the HP case, a historical fact will happen by the 32 nm node: the contribution of static power dissipation will become higher than the dynamic power contribution to the total power consumption! This main fact could affect the MOSFET saturation current as can be observed

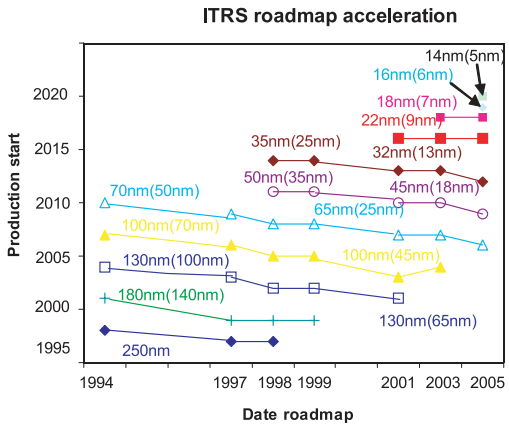


Fig. 1. ITRS roadmap acceleration since 1994 for MPU devices (HP devices) [1].

[a] e-mail: sdeleonibus@cea.fr

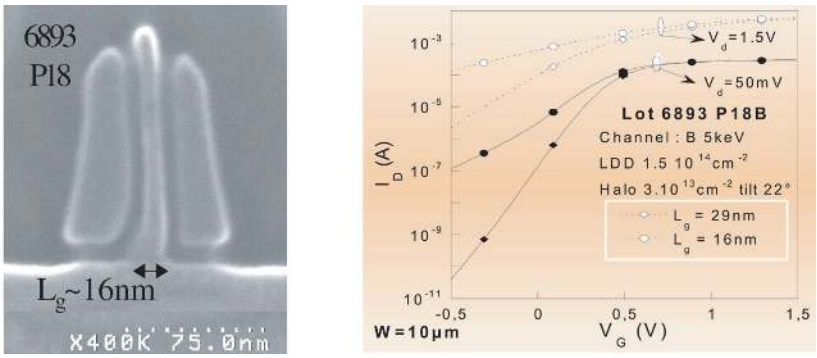


Fig. 2. Functional finished gate length 16 nm bulk n-MOSFET sub threshold characteristics. Gate oxide thickness is 1.2 nm [4]. Isat is 600 μA/μm.

on historical trends of smallest gate length devices [2]. Multigate devices could improve somewhat this evolution (see Sect. 4.2.2) by improving the ratio between saturation current and leakage current. In this paper, we will analyze the various mechanisms giving rise to leakage current in a MOS device and that can impact consumption of final devices. Gate leakage current is already a concern. In the case of LSTP devices, a High Dielectric Constant (HiK) gate insulator could be needed earlier than expected in order to limit static consumption (see Sect. 4.2).

In Section 2 of this review, we will first analyze the main limitations and showstoppers affecting bulk CMOS scaling. In Section 3, the issues in lowering supply voltage to reduce power dissipation are identified. In Section 4, the limitations to scaling must be taken into account in the device optimization in terms of gate stack, channel and source and drain engineering as well as new devices architectures (FDSOI or multigate devices). The alternative possibilities offered by new materials for enhancement of device transport properties or power dissipation are reviewed in Sections 5 and 6. Finally, in Section 7, we review the applications demonstrated by Single or Few Electronics in the field of memories or possible alternatives to CMOS.

2 Limitations and showstoppers coming from CMOS scaling

CMOS device engineering consists in minimizing leakage current together with the maximization of output current. In sub 100 nm CMOS devices, non stationary transport gains more importance as compared to diffusive transport.

2.1 Origin of leakage current in CMOS devices

Several mechanisms can generate devices leakage in ultra small MOSFETs, which can be sorted in two categories:

(a) Classical type
- Drain Induced Barrier Lowering (DIBL) is due to the capacitive coupling between source and drain.
- Short Channel Effect (SCE) due to the charge sharing in the channel in the short channel devices at low V_{ds}.
- Punch-Through between source and drain due to the extension of source space charge to the drain.

(b) Tunneling currents
- Direct tunneling through the gate dielectric.
- Field assisted tunneling at the drain to channel edge. This effect occurs if electric field is high and tunneling is enhanced through the thinnest part of the barrier.
- Direct tunneling from source to drain. This effect will occur in silicon for a thicker barrier than on SiO_2 because the maximum barrier height is lower (1.15 eV in Si versus 3.2 eV in SiO_2).

2.2 Issues related to non stationary transport

Velocity overshoot and ballistic transport are the mechanisms that will enhance drivability in sub 50 nm channel lengths devices. However, the impact of Coulomb scattering by dopants on transport is non negligible even in the 5 nm range channel lengths [3,4]. Superhalo doping is efficient to improve SCE and DIBL in 16 nm finished gate length (Fig. 2) [5] but will degrade the channel transport properties [5] by dopant Coulomb scattering (Fig. 3a) and high transverse electric field.

The degradation of transport properties can be observed on short channel mobility measurement by using a specific method with direct L_{eff} measurement ([6], Fig. 3b). A mobility degradation of a factor 2 to 3 or more can be measured on the most aggressive nano-scaled bulk technologies. The ITRS target of a transconductance increase by a factor 2 [1] is still very challenging on such gate length even if such an enhancement is reported on long channels. Furthermore, for such gate lengths access resistance due to extension scaling is an issue (Fig. 3a) [4].

3 Issues in supply voltage down scaling

In the future, the electronics market will require portable objects used in daily life and consequently low standby

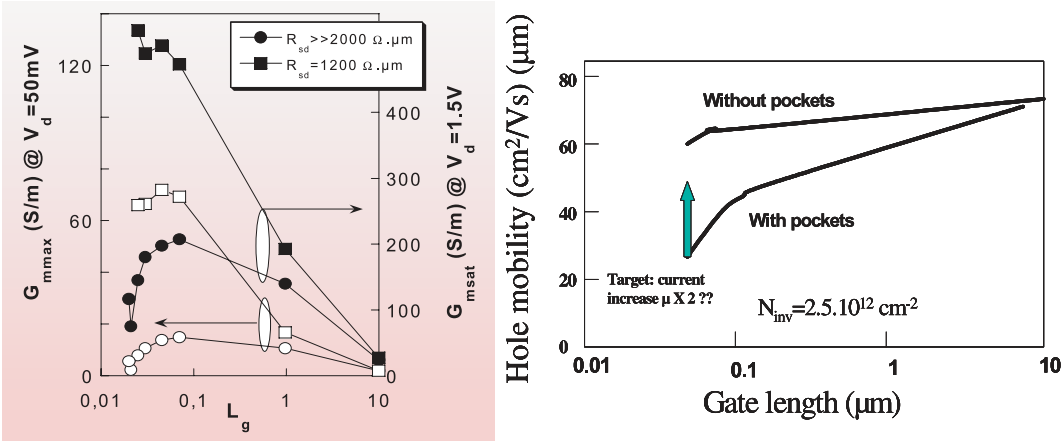


Fig. 3. (a) Effect of halo doping on nMOSFET short channel saturation and linear transconductance (Lg as low as 16 nm). The role of access resistance through extension doping is also investigated [4]; (b) typical measured p channel mobility loss when gate length is down-scaled due to halo/pockets doping [6].

power dissipation and low active power consumption. Scaling down of supply voltage is an essential leverage to decrease power dissipation. However, it raises several questions about the possible lower limits.

The power dissipation P of a MOSFET is due to static and dynamic contributions expressed by:

$$P = P_{\text{stat}} + P_{\text{dyn}} \tag{1}$$

$$P_{\text{stat}} = V_{dd}.I_{\text{off}} \tag{2.1}$$

and

$$P_{\text{dyn}} = CV_{dd}{}^2 f \tag{2.2}$$

P is the total power dissipation; P_{stat} and P_{dyn} are the static and the dynamic power dissipations respectively. The strong impact of supply voltage on power dissipation appearing in (1), (2.1) and (2.2), will also preclude a strategy of threshold voltage value adjustment depending on the application.

Information theory and statistical mechanics as well as the electrostatics of the device will set the limits of switching of binary devices. Moreover, dopant fluctuations will affect the control of device characteristics substantially: that is why low doping of CMOS channel will help in the down scaling of supply voltage.

3.1 Fundamental limits of binary devices switching

Quantum mechanics illustrates that switching involves non linear devices that would demonstrate a gain. That could occur with or without wavefunction phase changing. The Quantum limit on switching energy will be given by the Heisenberg's uncertainty principle:

$$E \geq \frac{\hbar}{\tau}$$

which gives a minimum switching energy of:

$$E_{\text{min}} = 10^{-5} aJ$$

considering $\tau = 10$ ps, $h = 2\pi\hbar$ is Planck's constant 6.34×10^{-34} J.s.

The second principle of thermodynamics imposes the maximization of entropy at temperature T. Applied to information theory this has a consequence on the minimal energy a system based on binary states of each bit of information will require to switch from one state to the other: $E \geq kTLn\,(2)$ with entropy $S = kLn\,(2)$ linked the quantity of information available in such a system. Thus: $E \geq 3 \times 10^{-3}$ aJ at $T = 300$ K.

If the system has a large number of gates N, with a response time τ, that could switch at an average rate time τ_{mbf}, then the mean time between failures ($MTBF$) is given by the expression: $\tau_{mbf} = \frac{\tau}{N}\frac{1}{P} = \frac{\tau}{N}e^{\frac{E}{kT}}$. $P = e^{-\left(\frac{E}{kT}\right)}$ is the switching probability of a single gate. We can demonstrate that the minimum switching energy is given by: $E \geq kTLn(\frac{N.\tau_{mbf}}{\tau})$. If we consider $N = 10^9$, $\tau = 10$ ps and $MTBF = 1000$ h (i.e. 3.6×10^6 s), then we get: $E \geq 0.25$ aJ.

Among the 3 limitations mentioned above, this latter is the largest one.

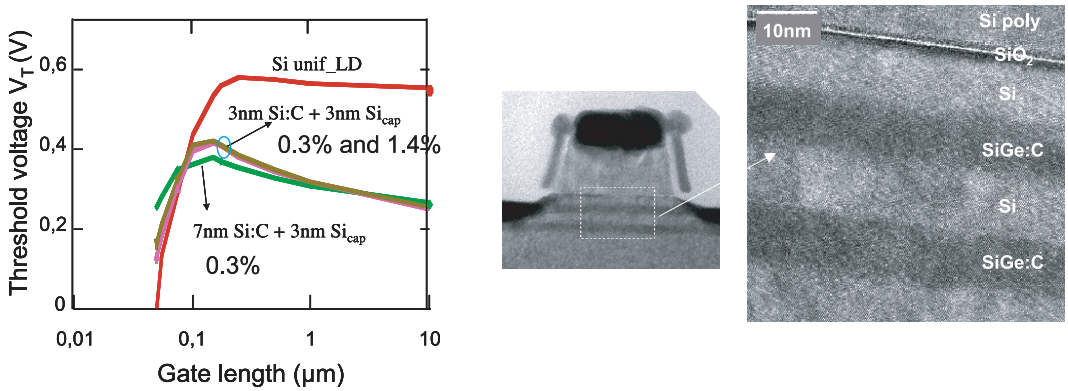
In order to estimate the associated minimal switching voltage V_{min} one must consider the capacitive load C_L associated to a switching gate. We will then extract V_{min} from the following relation:

$$kTLn\left(\frac{N.\tau_{mbf}}{\tau}\right) = C_L V_{\text{min}}^2$$

and get

$$V_{\text{min}} = \left(\frac{kTLn(\frac{N.\tau_{mbf}}{\tau})}{C_L}\right)^{1/2}.$$

At $T = 300$ K, $V_{min} = 10$ mV will be the limit if the load capacitance is in the range 0.4 fF (corresponding to 1 nm gate oxide thickness).



(a) (b)

Fig. 4. Introduction of Carbonated silicon in MOSFET channel: (a) influence on Short Channel effect. [17]; (b) optimization by a Multibarrier channel [18].

3.2 Issues related with decananometer gate length devices

In the decananometer range (less than 100 nm), besides classical 2 dimensional electrostatic effects, tunneling currents will contribute significantly to MOSFET leakage. In the following, we review the principal parasitic effects that could limit ultimate MOSFETs operation.

Direct tunneling through SiO_2 gate dielectric is significant for a thickness less than 2.5 nm. It contributes to the leakage component of power consumption. 1.4 nm thin SiO_2 is usable without affecting devices reliability [3,7–9].

High doping levels in the channel reaching more than 5×10^{18} cm^{-3} enhances Fowler-Nordheim field assisted tunneling reverse current in sources and drains up to values of 1 A/cm^2 (under 1 V) [10].

Direct tunneling from source to drain is easily measurable for very short channel lengths [4,5] lower than 10 nm. It will affect subthreshold leakage substantially at room temperature for channel lengths less than 5 nm.

Classical small dimension effects are more severe than the fundamental limits of switching (quantum fluctuations, energy equipartition, or thermal fluctuations). A minimum value is required for threshold voltage due to:

– *Subthreshold inversion.* For ideal fully-depleted SOI(FDSOI) 59.87 mV/dec subthreshold swing can be obtained at 300 K. The limit V_T value is 180 mV precluding a supply voltage V_S lower than 0.50 V. Impact Ionization MOS (I-MOS) would allow reducing subthreshold swing to less than 5 mV/dec. However, performance remains an issue [11].
– *Short channel effect* due to the charge sharing along the transistor channel following the relation:

$$\Delta V_T = -4\varphi_F \frac{C_w}{C_{ox}} \frac{x_j}{L} \left[\left(1 + 2\frac{W}{x_j} \right)^{1/2} - 1 \right]$$
$$= -4\varphi_F \frac{\varepsilon}{\varepsilon_{ox}} \frac{t_{ox}}{L} \frac{x_j}{W} \left[\left(1 + 2\frac{W}{x_j} \right)^{1/2} - 1 \right]. \quad (3)$$

Here V_T is expressed by:

$$V_T = V_{FB} + 2\varphi_F - \frac{Q_B}{C_{ox}} \quad (4)$$

where

$$V_{FB} = \varphi_{MS} - \frac{Q_{ox}}{C_{ox}} \quad (5.1)$$

and

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}; \quad \varphi_{MS} = \varphi_M - \varphi_s \quad (5.2)$$

ΔV_T is the threshold voltage decay; t_{ox} is the gate dielectric thickness; ε and ε_{ox} are the silicon and gate dielectric constant respectively; L is the channel length; X_j is the drain or source junction depth; W is the space charge region depth; V_T is the threshold voltage; V_{FB} the flatband voltage; φ_F the distance from Fermi level to the intrinsic Fermi level; Q_B the gate controlled charge; C_{ox} is the unit area capacitance of the gate insulator. φ_{MS} is the difference between the workfunctions of the gate and the semiconductor; Q_{ox} is the oxide charge density; φ_M and φ_S are the metal and the semiconductor workfunction.

Gate depletion and quantum confinement in the inversion layer will play an important role on short channel effect by adding their contribution to the gate to channel capacitance C_G. SCE is the main limitation to minimal design rule. For low V_T values it can be of the order of V_T. In order to maintain inverter delay degradation to less than 30%, we must observe the condition $V_T = \frac{V_{DD}}{3}$ [12]. V_{DD} is the supply voltage.

– *Drain Induced Barrier Lowering (DIBL)*

Classically, DIBL is due to the capacitive coupling between drain and source resulting in a barrier lowering on the source side. An eased charge injection from the source allows an increased control of the channel charge by the source and drain electrodes and reduces the threshold voltage. This effect (thus ΔV_T) increases with increasing Vds and decreasing L. A simple model shows that: $\Delta V_T = -\gamma \frac{Vds}{L^2}$ (γ is in the range of 0.01 μm^2).

3.3 Statistical dopant fluctuations

The effect of dopant fluctuations has already been considered by Shockley in 1961 [13]. Recently, special attention is being paid to this subject because the number of dopants in the channel of a MOSFET tends to decrease with scaling [14,15]. The random placement of dopants in the MOSFETs channel by ion implantation will affect devices characteristics for geometries lower than 50 nm. The discrete nature of dopant distribution can give rise to asymmetrical device characteristics [15].

Dopant fluctuations and Fowler Nordheim limitation at high electric field will encourage the use of low doped thin SOI.

4 Technological options to MOSFET optimization

In Sections 4.1, 4.3, the possible solutions to overcome the physical limitations encountered in classical scaling are reviewed through gate stack and channel/substrate engineering as well as source and drain engineering. Mastering and improvement of transport properties by strained channels and substrate engineering will be of primary importance in the future and not only limited to threshold voltage adjustment as it was the case in the past. The gate stack will also be reviewed on the electrical properties side as well as on the defect density view point. Source and drain engineering has to be addressed not only on the dopant activation side but also on the architecture side: access resistance to the channel can drastically reduce any advantage brought from channel transport properties optimization.

In Section 4.2, we review the alternative architecture candidates to replace bulk devices by leveraging the trade off between performance and power consumption. The power dissipation challenge will be the hardest challenge to face in the future whereas portable devices and systems will drive the market in the nanoelectronics era. That is why thin films and Multigate architectures are major alternative approaches to extend CMOS life to the end of the roadmap and possibly beyond.

4.1 Gate stack and channel/substrate engineering

Threshold voltage management issues in classical bulk MOSFET will guide its scaling.

Gate and channel engineering must be optimized together because both physical characteristics affect the nominal V_T value of expression (4) which can be written as:

$$V_T = V_{FB} + 2\varphi_F - Q_B/C_G \qquad (6)$$

(gate depletion and channel quantum effects are taken into account).

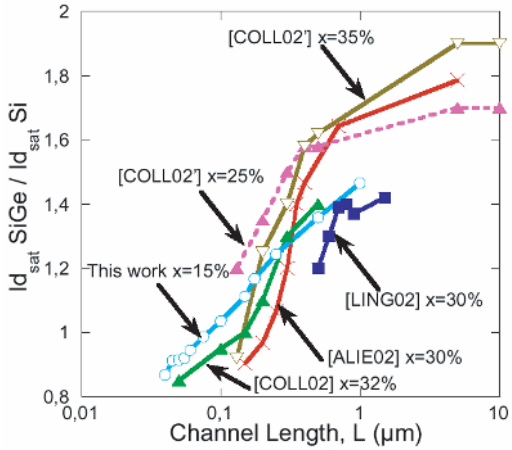


Fig. 5. Gain in drain current vs. gate lengths at $VGT = VDS = -1.3$ V for [ALIE98] = [25]; at $VGT = -0.5$ V $VDS = -2$ V for [LIND02] = [26] and at $VGT = -1$ V $VDS = -1.5$ V for [COLL02] = [27], [COLL02′] = [28] and [24] ($VGT = VG - VT$).

Low V_T values will result from:

- *tuning surface doping concentration (see Sect. 4.1.1)*,
- *strained channel engineering (see Sect. 4.1.2)*,
- *choosing the gate material (see Sect. 4.1.3)*,
- *adjusting gate insulator thickness (see Sect. 4.1.4)*.

4.1.1 Tuning surface doping concentration as low as possible

Excellent localization of the dopant profile is needed to minimize junction parasitic capacitance and body effect. Selective Si epitaxy of the channel has also been demonstrated to achieve almost ideal retrograde profiles [16]. Selective epitaxial Si:C acts as Boron diffusion barrier and thus help to improve drastically short channel effect [17] (Fig. 4a) as well as low field mobility. Multibarrier channels using an alternated Si/SiGeC epitaxial channel structure has been proven to be efficient in optimizing short channel effects immunity compatible with high devices drivability [18] (Fig. 4b). These solutions can give a longer breath to bulk CMOS devices scaling.

4.1.2 Strained channel engineering

4.1.2.1 Global strain

Strained SiGe [19], $SiGe_xC_y$ based alloys or strained Si epitaxy have been studied to increase the channel mobility [17,20] by introducing compressive or tensile strain to enhance hole or electron effective mass respectively. In order to achieve such channel architectures, bulk relaxed SiGe pseudo substrates obtained by graded SiGe buffer were intensively developed during the last decades [21, 22]. High-quality pseudomorphic silicon layer with very high biaxial-strain values (typically 1.2–1.5 MPa or more)

can be grown on those substrates. The resulting degeneracy leverage on the conduction bands leads to effective electron mass reduction and mobility increase up to around 80%.

The quality of those substrates has been spectacularly improved. Independently of possible remaining defects (dislocation pile ups, stacking faults, etch pits [23]) a major limitation remains: the reported gain in current enhancement decreases with gate length reduction [24] (Fig. 5). This I_{ON} gain decrease with L was attributed to self heating (monitored pulse drain current measurement) due to low thermal conductivity of SiGe [29]. But some authors have pointed out than even at low drain voltage (insensitive to self heating) the current gain loss is still relevant. Both possible S/D implantation damages [30] and lateral strain S/D relaxations [31] may explain the loss on mobility increase on those short channel strained devices.

However, high quality gate insulator and subthreshold characteristics optimization require a Si cap layer on top of the channel and low thermal budget [15]. Ultimately, a HiK gate insulator is needed in these architectures [32,33].

In parallel high quality strained silicon on insulator substrate, with or without SiGe for dual channel operation has been developed [34,35]. SiGe condensation technique can lead to high quality SiGe on Insulator (SGOI) whereas high quality SGOI and sSOI substrated by Smartcut ® were reported.

4.1.2.2 Process induced strain

Process induced strain is the most mature option for today's IC and is proposed in the 65 nm and 45 nm platforms [36]. In those technologies external strain mostly uni-axial is applied by various means. The most currently used approach is compressive or tensile contact etch stop layer to obtain respectively tensile channel nMOS or compressive channel pMOS. The I_{ON} gain by using those technique is still moderate (–15–20% typically but its low cost is very attractive. Recent studies quantify by direct measurements the mobility enhancement on short channels with process induced strain [37] showing a direct correlation between low and high Vd regime.

4.1.2.3 Other substrate solutions

Unstrained solutions may use the chemical composition of the substrate or the crystalline surface or transport orientation.

Changing surface silicon orientation or transport orientation can lead to mobility improvement by a factor 2 or more [38]. The (110) surface orientation lead to an improvement for hole. Dual channel with (100) orientation for electrons and (110) orientation for holes was reported [39]. Germanium and Germanium-on-insulator were proposed as unstrained substrates. One of the higher channel mobility improvement by using column IV elements is compressive Germanium with more than a factor 10 of hole inversion charge mobility improvement [40] which could bring a solution for dual channel optimization.

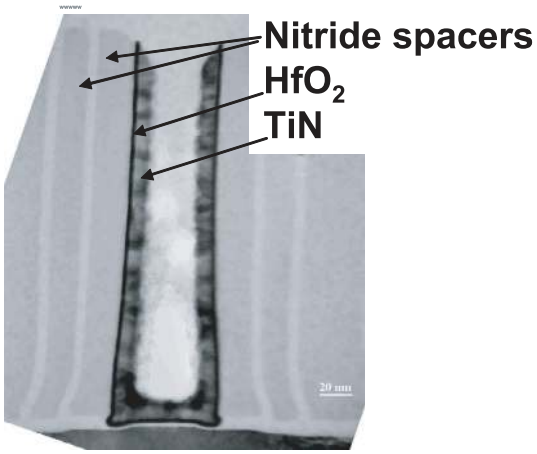


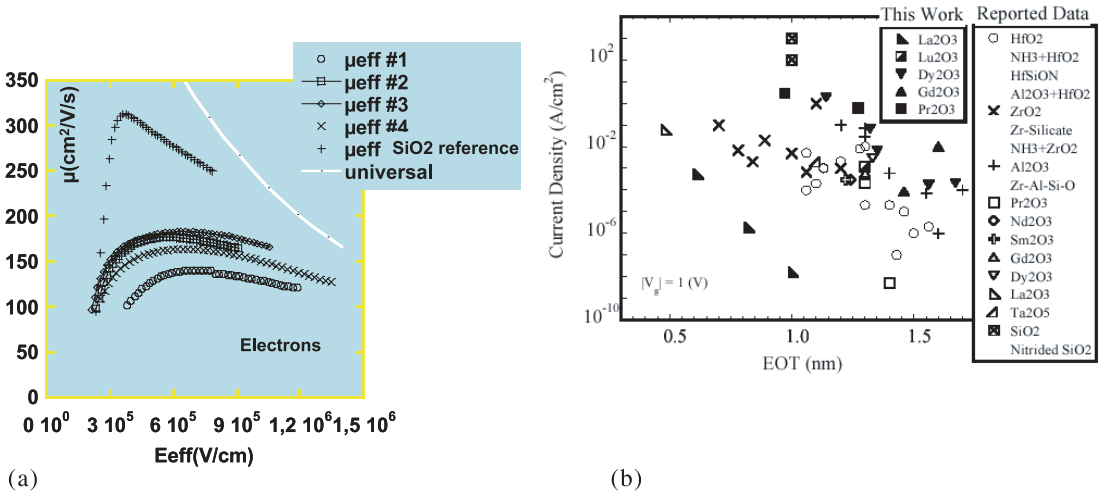
Fig. 6. TEM cross section of TiN/HfO2 Damascene gate stacks [43].

4.1.3 Choosing the gate material

Ideal transfer CMOS inverters characteristics requires symmetry of threshold voltage for n and p channel devices (i.e. $V_{TP} = -V_{TN}$). Several alternatives have been envisaged:

- *The use of n+ poly gate for nMOSFET and p+ poly gate for pMOSFET.* This solution suffers from Boron penetration into SiO_2 coming from the $p+$ doped gate. Nitrided SiO_2 *limits* this effect *without avoiding it*: trapping centers are created near or at the SiO_2/Si interface decreasing carrier mobility.
- *The use of metal gate material.* No gate depletion is observed in this case. The use of midgap gate (TiN for example) on bulk silicon or partially depleted SOI will be dedicated to supply voltages higher than 1 V. Workfunction engineering for Dual metal gates is challenging: the highest CMOS performance/lowest leakage current trade off can be obtained. It is mandatory on low doped FDSOI.

Several approaches have been proposed for metal gate integration. The classical process integration, so called direct gate, requires the protection of the metal gate material from ion implantation as well as from oxidation during the dopant activation anneal. TiN has often been chosen as a gate material [41] because it is available as a standard in the industry. Alternatives such as the damascene gate (Fig. 6) [42,43] have been achieved in order to avoid the issue of source and drain activation temperature. It is noteworthy that, thanks to the damascene architecture, High Frequency and Multi threshold devices could be embedded in Systems On Chip. Complete silicidation of polysilicon gate has been demonstrated to lead to metallic behavior of both n and p gates [44–46]. However, integration with HiK dielectrics gives rise to the so called Fermi level pinning similar to what is obtained with polysilicon gates [47].



(a) (b)

Fig. 7. (a) Degradation of electron mobility with HfO2/Si [43]; (b) leakage current as a function of EOT for various HiK materials reported from [52].

4.1.4 Gate dielectric engineering

The gate leakage due to direct tunneling in standard SiO_2 or SiO_xN_y is one major show stopper [1]. It will impact directly the static power dissipation P_{stat} according to relation (2.1) Let us consider a circuit with active area of the order of 1 cm^2 and gate oxide SiO_2 t_{ox} = 1.2 nm. Considering the contribution of gate leakage to Ioff under the condition V_{dd} = 0.5 V, then $Pstat$ (0.5 V) = 5 W. We would get P_{stat} (1.5 V) = 750 W if Vdd = 1.5 V! This results as a major show stopper for scaling of CMOS technology. That is why High K will be urgently needed in the near future. Besides affecting static power, gate leakage also impacts negatively delay time [48] and affects the functionality of logic circuits.

4.1.4.1 From SiO_2 to High K gate dielectrics

A decrease of devices performance has been reported if SiO_2 thickness is lower than 1.3 nm [49] suggesting a surface roughness limited mobility process due to the proximity of sub-oxide. The strong band bending due to quantum mechanical corrections affects the lower limit of supply voltage in the constant field scaling approach [50]. Solutions compatible with silicon gate are also investigated to keep compatibility with a standard CMOS process flow: $HfSiO_x$, $ZrSiO_x$ are given much attention as good candidates [51]. These solutions are *dielectric thickness budget* consuming (SiO_x interface) and Fermi level pinning occurs at the HiK/poly gate interface [47].

Very low leakage current has been reported by using HfO2 of 1.3 nm Equivalent Oxide Thickness (EOT) combined with a TiN gate integrated on 45 nm CMOS by a damascene process [43] (Fig. 6). Electron mobility degradation is reported compared to SiO_2 gate dielectric [43] attributed to stress induced phonon scattering (Fig. 7a). These materials have a smaller bandgap than SiO_2: thus trapping is a strong reliability issue [5]. That is why a SiON interface could be helpful to reduce the leakage current thanks to the higher bandgap of SiON.

La_2O_3 films with EOT as thin as = 0.61 nm have been proven to demonstrate very low leakage current as low as $J = 5.5\times10^{-4}$ A.cm^{-2} [52] compatible with high interface quality and acceptable mobility values (Fig. 7b). These results are obtained on low temperature end of process and aluminum gate. Integration into a direct gate process is still an issue.

4.1.4.2 Combining gate stack and channel workfunction engineering

Specific technological optimization may be necessary to maximize the transport gain in short channels. In particular, maintaining the high stress of 1.2 or more GPa in a nanoscaled device and reduce ion implantation damages are among the main challenges. Meanwhile, the combination of strained Si and SiGe channel can be a promising solution for future applications. For instance, it was shown that both surface conduction and hole mobility enhancement (65% at high transverse electric field) could be achieved by using selective SiGe for PMOS coupled with high-k and metal gate [33,53] (Fig. 8).

Even in the case of low gain in short channel I_{ON} values [33], it is possible to adjust V_T by locally strained layers by using a mid gap metal gate.

4.2 Architecture alternatives to improve CMOS performance and integration

4.2.1 Fully depleted SOI devices

In order to obtain the lowest subthreshold slope (60 mv/dec) and acceptable DIBL on FDSOI a practical rule is used: $T_{Si} \leq L_{gate}/4$ [54]. The spreading of potential into the buried oxide, due to the coupling with the top

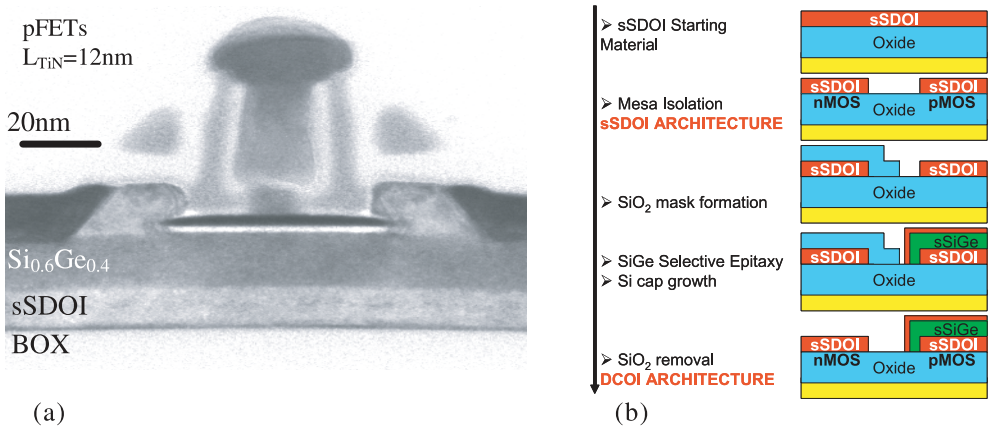


Fig. 9. (a) Cross sectional TEM pictures of the co-integrated dual channels MOSFETs with a HfO_2/TiN/Poly/NiSi gate stack [34,37]; (b) Strained Dual channels CMOS Process Flow [34].

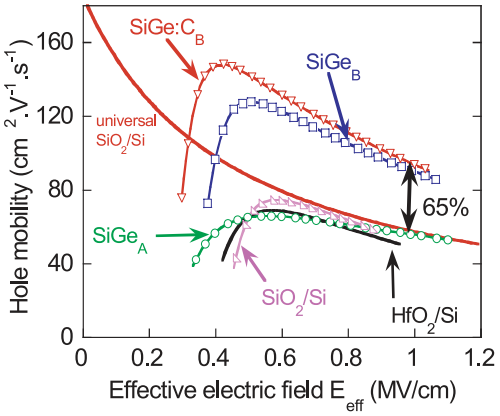


Fig. 8. Effective hole mobility versus effective field for the various channel-gate dielectric stacks [53].

gate, increases the coupling between source and drain and thus DIBL. Ultra-low SOI films thickness is difficult to control. That is why partially depleted SOI has been proposed [54,55]. Because of complete isolation of the SOI devices as well as lower junction capacitance, improved figures of merit are obtained as compared to bulk [54]. The threshold voltage is dependent on Si film thickness whenever the film thickness becomes lower than the space charge region. V_T is expressed as [54]:

$$V_T = V_{FB} + 2\varphi_F + \frac{qN_A T_{Si}}{2C_{ox}}. \quad (7.1)$$

In the case of a low doped channel, expression (7.1) can be simplified as the well known relation:

$$V_T = (\varphi_M - \frac{E_i}{q}) + \frac{kT}{q} \ln\left(\frac{2.C_{ox}.kT}{q^2 n_i T_{Si}}\right) \quad (7.2)$$

N_A is the acceptor concentration; T_{Si} is the silicon thickness; C_{ox} is the gate insulator capacitance; E_i is the semiconductor intrinsic Fermi level energy; n_i is the intrinsic carrier concentration.

Scaling of FD devices encounters some limitations due to the quantum confinement in ultra thin films and its incidence on the threshold voltage value [56]: the increase of the fundamental level of the conduction band will increase flat band voltage and V_T consequently.

The functionality of ultra small 6 nm gate length devices on 7 nm thin Si film was demonstrated [57]. However, the electrical performances of these devices are extremely sensitive to the SOI film thickness variations due to the fact that a compromise must be found between series resistance minimization and DIBL [58].

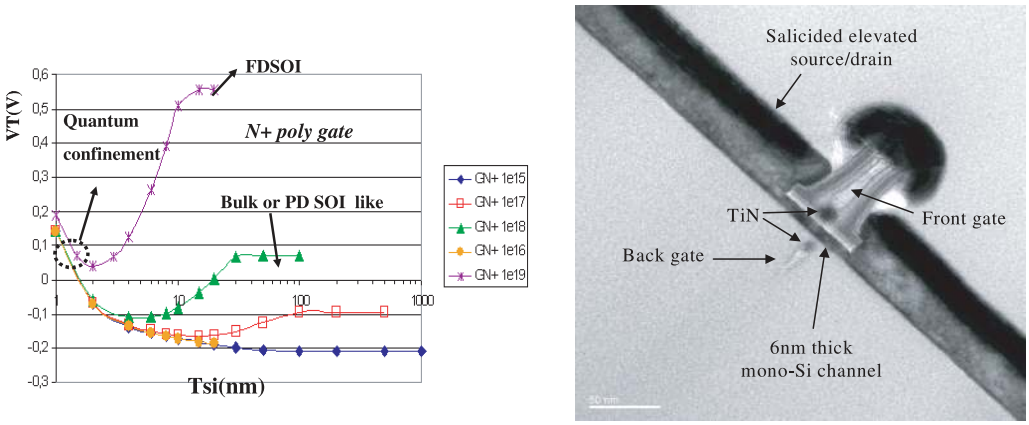
Combination of strained channels and SOI could result in optimized trade off between short channel effects reduction and enhanced transport properties. Si and SiGe Dual strained channels operation on insulator has been demonstrated functional down to gate lengths of 15 nm (Fig. 9) [34,37].

Self-heating is an issue on fully isolated devices because of the low thermal conductivity of SiO_2. Replacing SiO_2 by Al_2O_3 has been proposed as a solution because the thermal conductivity of Al_2O_3 is ten times larger than for SiO_2 [59,60].

4.2.2 Multigate devices

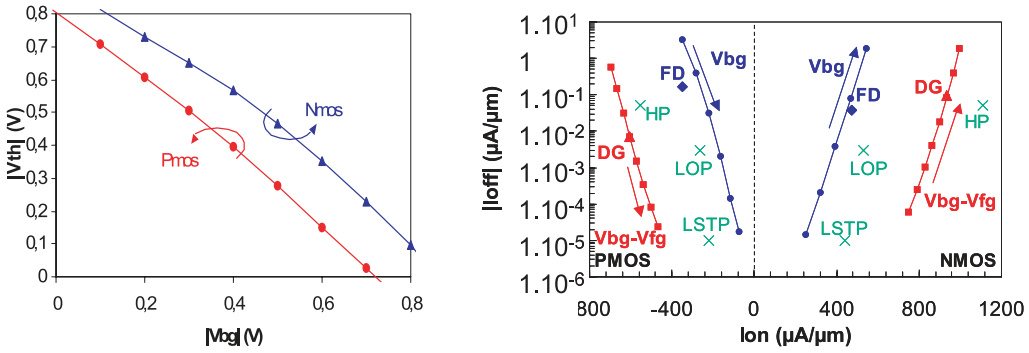
SOI material should allow to realize attractive devices like multi gated MOSFETs [61] that will extend further scaling of FD devices which are limited by the quantum confinement issue as well as DIBL via the coupling of the gate with buried oxide [56] (Fig. 10a). With multi gate devices, short channel effects and leakage current can be drastically reduced because 60 mV/dec subthreshold swing and high drivability can be obtained (Fig. 1b). In the saturation regime, transport occurs by volume inversion due to the coupling of both gates. The conditions for controlling short channel can be relaxed compared to single gate FD devices [56,62–66]. Nevertheless, the control of thin SOI and design of high density circuits with these devices have to be demonstrated.

Another main feature of these devices is to bring a solution to the channel dopant fluctuation issue in small volume. Reducing the film thickness to the minimum, allows using nearly intrinsic Si films because bulk punch-through



(a) (b)

Fig. 10. (a) Threshold voltage dependence of SOI devices as a function of SOI thickness for different values of channel doping [56]; (b) TEM cross-section of a 10 nm planar bonded double gate transistor with TiN metal gate [70].



(a) (b)

Fig. 11. (a) Tunable threshold voltage of the devices as a function of back gate voltage; (b) I_{off} vs. I_{on} of tunable DG MOS (adjustable $V_{bg} - V_{fg}$) and tunable DG MOS operating in FD mode (adjustable V_{bg}) between Low-stand-by-power (LSTP) and High-performance (HP) – 90 nm node [70].

is no more a problem. Adjusting V_T to match the overdrive defined by $(V_s - V_T)$ with a low supply voltage V_S will require adjusting the gate workfunction φ_M according to relation (5.1). That is why, workfunction engineering on metal gate and HiK stacks is mandatory for low V_S applications.

Among the various studies published on multi-gate devices [67–69], many architectures have been proposed in which the channel is controlled by two or more gates.

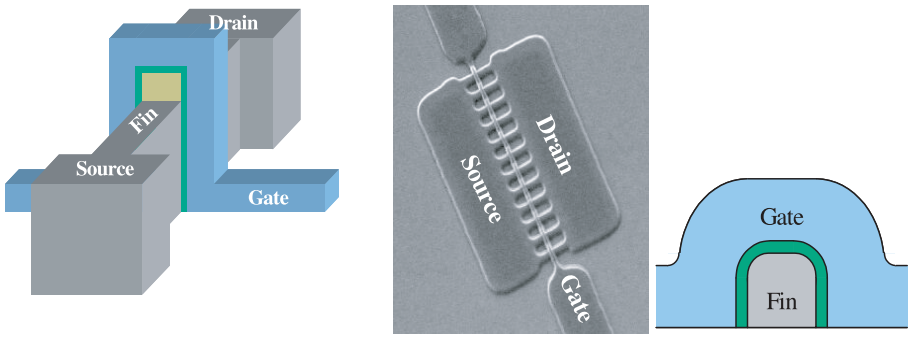
In planar architectures, the structure can be non self-aligned, i.e. fabricated with one photo-lithography step for each gate, or self-aligned, using only one lithography step to define both gates. The non self-aligned architecture by wafer bonding is the most straightforward approach to fabricate planar double gate. The success of this approach depends on the lithography capability to align very short gates one to the other. Figure 10b shows a 10 nm non self-aligned planar double gate transistor, fabricated thanks to the use of wafer bonding and e-beam lithography [70–73]. Notice that a quasi-perfect gate alignment, with an accuracy of a few nanometers, could be achieved thanks to the self-aligned regeneration of the alignment marks after the bonding step [74].

Several approaches have been proposed to fabricate self-aligned planar double gate MOSFETs. The first one consisted in patterning a narrow silicon active area on a SOI substrate, etching a localized cavity under this active area into the buried oxide, and filling it by the gate material [75]. After gate patterning, the silicon active area is surrounded by the gate.

Another gate-all-around (GAA) architecture, based on the silicon-on-nothing (SON) process, has been proposed more recently [76] and demonstrated down to very short gate lengths. This approach relies on successive epitaxial growth of crystalline SiGe and Si layers. The SiGe layer is then selectively etched to form a tunnel below the silicon film, and this tunnel is filled by the gate material.

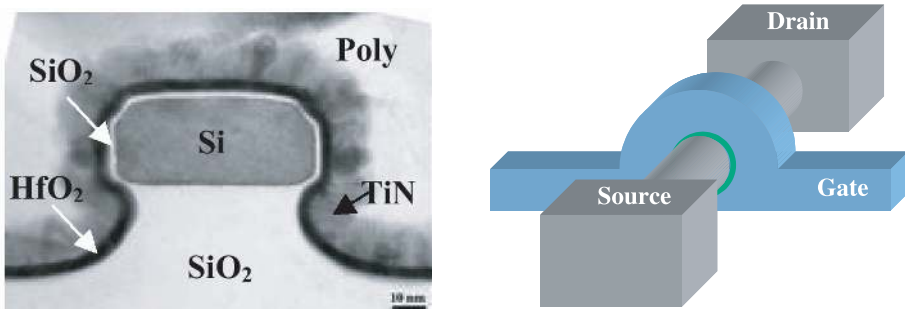
In the PAGODA architecture [77], the unpatterned back gate stack is deposited and encapsulated before wafer bonding. After initial substrate removal, the front gate is patterned and silicon spacers recrystallized from the channel are formed and silicided. These silicided spacers are used as a hard-mask for back gate etching and undercut.

The process flow proposed in [78] starts also from back gate stack deposition and wafer bonding. The whole stack, comprising the front gate, the channel and the back gate



(a) (b)

Fig. 12. (a) Schematic of a FinFET device. (b) Left: SEM top-view of a 20 nm gate length multifinger Trigate device. Right: schematic cross-section of one Trigate fin.



(a) (b)

Fig. 13. (a) Ω-shaped FET. Functional devices with gate length as low as 10 nm are obtained [86]. (b) Schematic of a cylindrical surrounding-gate device [84].

is then patterned. Insulated layers are formed besides the gates by use of oxidation rate difference between the gate and the channel materials. Source/drain regions are then regenerated by lateral epitaxial regrowth from the channel edges.

The key technological issues of the planar architectures are the precise controls of the very thin film thickness and of the back gate dimension, since the back gate is not directly accessible from the top of the wafer.

However, with the planar bonded architectures it is possible to bias the front and back gate independently [74] (Figs. 11a, b). That allows the use of different transistors families with several threshold voltages values available on the same chip by using one single type of device. The electrical characteristics of the devices can fulfill the specifications of the 3 families of devices proposed in the ITRS [1], so-called High Performance (HP), Low Operating Power (LOP) and Low Standby Power (LSTP) [74] (Fig. 11b). Moreover, the planar bonded Double Gate devices are co integratable with single gate FDSOI and allow a metallic Ground plane by using the backside gate. The planar bonded architecture approach brings a unique innovative option to future Systems On Chip [79].

On the other hand, structures with fingered vertical channel, such as FinFET [80] (Fig. 12a), Trigate [81] (Fig. 12b), Ω-FET [82] (Fig. 13a), Π-Gate [83] and nanowire-FET [84] have been extensively studied. Fabrication of FinFETs relies on high aspect ratio fin definition and short gate patterning on this topography (Fig. 12a). Contrary to planar devices, the conduction takes place on the vertical sidewalls of the fin. The conduction width is thus twice the fin height (h_{fin}). As the fin height is limited to typically 50 to 100 nm, FinFETs are usually designed as multifinger transistors, with a conduction width quantified by $2.h_{fin}$. In order to obtain the same drive current per silicon area as planar double gate transistors, the spacing between the fingers has to be lower than the fin height.

Thus, one key technological issue lies in the multi-fin definition. Dense array of narrow fins have to be patterned, with a good control of the fin width and shape. The use of spacers as hard-mask for fin patterning seems unavoidable, as it allows to double the fin density and to design sub-10 nm wide fins [85].

Another approach consists in designing the fin with roughly a square cross-section (Fig. 12b). In that case, the channel is controlled by the gate on three sides. This device, so called Trigate [81], has a conduction width given by twice the fin height plus the fin width. Trigate is still a multifinger device, and the pitch between fins has to be lower than $h_{fin} + w_{fin}/2$ to obtain higher drive currents per silicon area than with planar devices. This limit is far more strict for Trigate than for FinFET, since the fin height must be as low as the fin width in order to operate

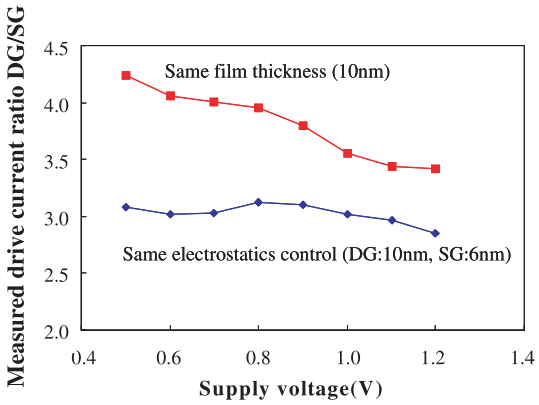


Fig. 14. Experimental drive current ratio between a 20 nm double gate and two 20 nm single gate devices as a function of the supply voltage [73].

in trigate mode, and comparable to the gate length to benefit from a good electrostatic channel control.

The Ω-FET [86] and Π-Gate architectures are basically similar to Trigate, but their channel control is close to that of a quadruple-gate device, thanks to the extension of the gate below the fin into the buried oxide [87]. The best electrostatic control can be achieved theoretically in a cylindrical channel completely surrounded by the gate (Fig. 13b). The most advanced practical realization of such a device is the 5 nm gate length nanowire-FET [84].

Thanks to their better electrostatics control, multiple gate transistors are likely to allow a triple drive current with respect to single gate transistors at a given off-state current [73,88].

To illustrate this, we have plotted in Figure 14 the ratio of the drive currents obtained experimentally on 20 nm co-integrated single gate and double gate devices. The drive current of the double gate transistor is 1230 μA/μm for an off-state current of 1 μA/μm at $V_{dd} = 1.2$ V, which can be considered as a high performance device.

Two cases can be considered:

1- Both devices have the same film thickness of 10 nm. The single gate transistor suffers from much more electrostatic control loss and the drive current ratio at $I_{off} = 1$ μA/μm is between 3.4 and 4.0.
2- Both devices exhibit roughly the same electrostatic control (subthreshold swing and DIBL respectively lower than 100 mV/dec and 250 mV/V). The film thickness is reduced to 6 nm for the single gate transistor. The current ratio is still around 3, because of the increased access resistances due to a thinner film for the single gate device.
Furthermore, if we consider loading capacitances (for example wires and junctions) in addition to intrinsic gate capacitance in the previous discussion, the multiple gate device advantage over single gate is further increased, because of the higher drive currents delivered by the multiple gate architectures.
Finally, since each added gate allows a better device scalability [87,89,79], the advantage of multiple gate devices is more and more evident as the gate length is reduced.

Several critical issues are associated with the use of thin film or narrow fin devices. An intrinsic limitation is the mobility reduction observed for film thickness below 5 to 7 nm [90]. This effect is partly due to an increased phonon scattering mechanisms on thin films [91] and can be further accentuated by a more pronounced impact of the surface roughness.

In addition, devices with ultra-thin films are sensitive to thickness fluctuations through short channel effects variations. The scaling length λ derived in [92] for low-doped double gate transistors is given by the expression:

$$\lambda = \frac{t_{\mathrm{Si}}}{2}\sqrt{\frac{1}{2} + \frac{2.C_{\mathrm{Si}}}{C_{ox}}}. \quad (8)$$

For an EOT of 1 nm, $\delta\lambda/\lambda$ is about 70% of $\delta t_{\mathrm{Si}}/t_{\mathrm{Si}}$. As short channel effects depend on L/λ, a fluctuation of 1 nm on a film thickness of 7 nm is equivalent to a gate length variation of 10%.

4.3 Source and drain engineering

Low energy implant (<1 keV) [49] and heavy molecules (BF3 [93], B10H14 [94],...) have been extensively studied to replace Boron to achieve $p+$ shallow junctions. Plasma doping is investigated as an alternative to obtain as implanted $p+$ junction depths lower than 10 nm [95, 96]. Transient Enhanced Diffusion (TED) is still the limiting process to reach the specified final junction depths (Fig. 15). Fast ramp up and down – so called spike or Flash annealing [96] – must be combined with Low Energy Ion Implantation [96] to reduce TED as much as possible, by reducing the role played by extended and dopant defects. Excimer Laser Anneal [97,98] (Fig. 15) has demonstrated the best trade off between low sheet resistance and junction depth shallowness: highest solid solubility combined with fast processing can be achieved. Low sheet resistance combined with low silicon consumption can be obtained with monosilicides (NiSi, PtSi) instead of disilicides (TiSi2, CoSi2) [99].

The same behavior will apply to SOI as well as bulk substrates (Fig. 15). However, on SOI films, several issues are linked with the access resistance optimization. As the film thickness decreases, achieving silicon doping becomes more and more challenging, because on one hand the square resistance of the silicon film increases in $1/t_{\mathrm{Si}}$ as shown in Figure 15. On the other hand, increasing dose and/or energy leads to surface silicon amorphization [73]: as long as the whole layer is not damaged, activation annealing allows the recrystallization of the film giving thus an active doping process window which is very narrow for a 5 nm thick silicon film. The surface species diffusion velocity during high thermal processes being strongly dependent on temperature and silicon thickness, the film becomes very sensitive to high temperature treatments [73,100] as silicon thickness decreases.

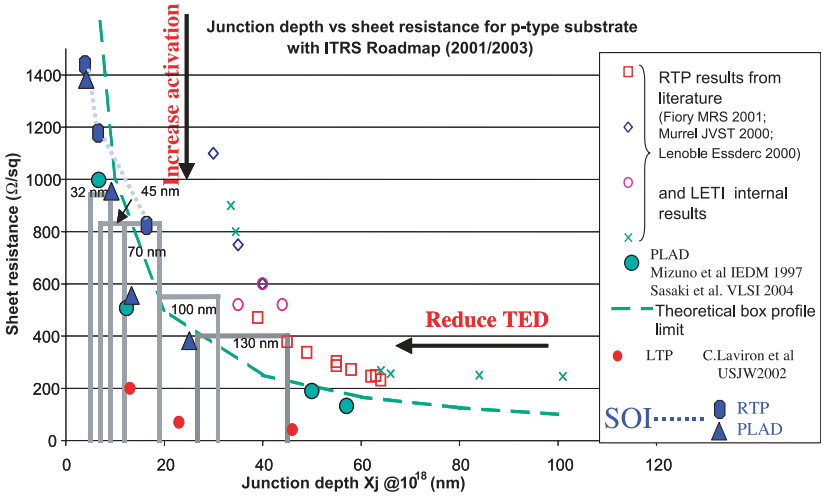


Fig. 15. P+ Sheet resistance as a function of junction depth or Si thickness for SOI [95–98].

Devices on thin SOI will require raised sources and drains by epitaxial growth to facilitate further silicidation: pre-anneal before epitaxial growth can lead to a destabilization which dramatically transforms the continuous silicon film into silicon solid droplets on the buried oxide as shown in Figure 16a. Therefore selective epitaxy of raised source/drain requires technological developments such as temperature optimization, modulation of the interface energy between silicon and buried oxide to ensure that the silicon film will keep its integrity during the whole fabrication process. Figure 16b illustrates results obtained when the temperature of the pre-anneal is lowered (down to 650 °C).

Silicidation process also requires technological optimization. Indeed diffusive metals have been introduced to suppress the voiding that occurs in the silicon films when silicon diffuses into the silicide. One way to overcome these technological difficulties could be to design MOS transistors with metallic source and drain either based on Schottky barriers [101] or modified Schottky barrier [102]. In both cases, selective epitaxy can be suppressed as source and drain are made out of metal. The key issue in this option is to find metals for N and PMOS with adjusted work function to design either adequate Schottky barrier or low specific resistance ohmic contacts.

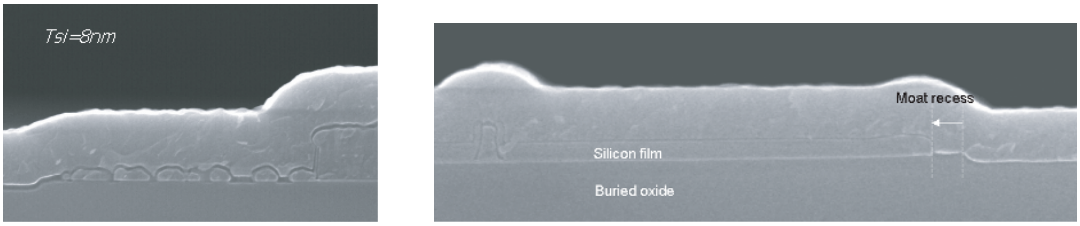
5 Exploiting non stationary transport or CMOS on semiconductors other than silicon?

The introduction of strained channels is limited by saturation velocity values at high electric fields. Under these conditions, non stationary transport can occur for very short channels and devices performances can benefit from velocity overshoot. Unless transport is limited by surface roughness or impurity scattering [4,103,104] ballistic transport can offer a new degree of freedom to the increase of devices performance in sub 100 nm Si channel length devices. If the low field mobility is high, then the mean free path of carriers becomes comparable to or higher than the channel length: ballistic transport is likely to be taken into account [49,105–107]. These transport properties can be enhanced whenever undoped or nearly undoped channels can be used. Architectures based on ultra thin bodies like Fully Depleted SOI or Multigate devices can ease the exploitation of these phenomena due to the fact that short channel doping can be minimized while keeping low short channel leakage. Reduction of channel length and supply voltage poses the issue of new scaling paradigms through the exploitation of non stationary effects. Germanium and GaAs for example have low field carrier drift velocities higher than in Silicon. However, at high electric fields the reverse situation occurs. Still the energy relaxation time is higher in Germanium than it is in Silicon thus velocity overshoot may occur for less aggressive channel lengths. Limitations will however come from integration of the new materials which could request new gate dielectrics. Typically, High K materials are needed to fabricate Ge based CMOS devices due to the Ge oxides instabilities. In these devices, hole mobility has been reported to be improved whereas electron mobility enhancement is still an issue [108]. Germanium offers the unique possibility for low temperature dopant activation [109].

6 Optimization of carrier transport and power dissipation

6.1 Electrostatics, transport and self heating issues

The best choice to maximize the CMOS integration density is obtained under the condition $\mu_n = \mu_p$ (μ_n and μ_p are respectively the n-channel and p-channel mobilities). Dual channels obtained from strained epitaxial layers could be a possible approach [40] (see Sect. 4.1.3). As far as a monolithic solution can be found, this unique



(a) (b)

Fig. 16. (a) SEM cross-section – After H_2 anneal, silicon agglomeration is observed for thin films. (b) Lowering the anneal temperature leads to less dramatic consequences of silicon agglomeration as in this case, only moat recess is observed [73].

Table 1. Electrons, holes bulk mobilities and saturation velocities (@300 K) of mostly used semiconductor materials.

Material	μ_n ($cm^2V^{-1}s^{-1}$)	μ_p ($cm^2V^{-1}s^{-1}$)	v_{sat} (10^7 cm/s)
Si	1400	500	0.86
Ge	3900	1900	0.60
GaAs	8900	400	0.72
C Diamond	1800	1800	2.7
4HSiC	900	120	2.0
InSb	78000	750	5.0

condition occurs in the case of C-diamond (Tab. 1). However, n dopant activation in this material is still limited [110] whereas, recently progress has been made for p doping [111]. However, ohmic contacts of metal to diamond need to be optimized. Moreover, C-diamond is far the highest thermal conducting material (10 times the thermal conductivity of silicon or 50 times the thermal conductivity of Al_2O_3) and could be integrated as a buried layer to limit self heating in future Semiconductor On Insulator substrates. The dielectric constant of C-diamond ($K_C = 5.7$) offers the best compromise between HiK and SiO_2 to control short channel effect according to relation (3).

However, the isolation on the valence band side is difficult (Tab. 2): the C/Si barrier height is far less than the SiO_2/Si barrier height (0.30 eV for C/Si instead of 4.93 eV for SiO_2/Si!). That is why a HiK insulator is needed. Among the best candidates, BeO or AlN offer a good compromise in terms of short channel effect ($K_{Beo} = 6.7$ or $K_{AlN} = 8.9$) and thermal conductivity (Tab. 2). Furthermore, their valence band is at least at –6.2 or –10.6 eV from vacuum. Thus a good isolation is obtained for holes whereas for C-diamond by itself would not be a good insulator on the valence band side.

Thus the integration of C-diamond has to be combined with HiK buried insulators if we wish to integrate it on silicon as a possible solution to limit power dissipation and suppress self-heating of CMOS devices (Fig. 17)! [112]

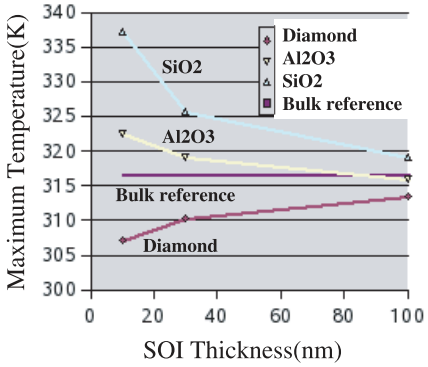


Fig. 17. Maximum channel temperature in $Lg = 50$ nm FD-SOI transistors with different Buried Insulators as a function of SOI thickness. $VDD = 1.2$ V [112].

6.2 Germanium on Insulator: a second life for Germanium?

Germanium was initially used to fabricate microelectronics through the realization of the first transistor. Many interesting properties can be accounted to Ge: larger low electric field mobility values than in Si as well as smaller μ_n/μ_p ratio (see Tab. 2), despite lower saturation velocity at high fields. However, Ge has a higher energy relaxation time which potentially relaxes linear gate length scaling constraint to gain performance as compared to Si.

Due to its compatibility with silicon processing and its availability in many fabs, Ge has recently been given much interest again as a promising candidate for high performance MOSFETs. Thanks to High-K materials, the non stable native Ge oxide is not a limitation anymore to the use of Ge in the CMOS technology. Low band gap materials show high diode leakage current. The impact of this leakage on MOS characteristics (I_{OFF}, bulk leakage) is a severe limitation for the use of bulk Ge for CMOS devices. Thus, a more realistic use of Ge for CMOS is Germanium On Insulator (GeOI) Fully Depleted MOSFETs since the bulk leakage is suppressed by the BOX and S/D leakage can be reduced by using ultra thin Germanium in a device operating in the Fully Depleted regime. We have realized Fully Depleted deep sub-micron (gate length down to 0.25 μm) Ge p-MOSFETs on Ultra Thin Germanium-On-Insulator (GeOI) wafers [113]. The Ge layer obtained by hetero-epitaxy on Si wafers is transferred using the Smart-Cut^{TM} process to fabricate 200 mm GeOI wafers with Ge thickness down to 60 nm (Fig. 18).

A full CMOS compatible p-MOSFET process was implemented with HfO2/TiN gate stack. An I_{ON}/I_{OFF} ratio higher than 10^3 and a 300 mV/decade sub-threshold slope

Table 2. Electrons affinity, bandgap, maximum valence band level, thermal conductivity and dielectric constant for various pertinent mostly used semiconductors and High K materials.

Material	Electron Affinity (V)	Gap (V)	Ev (V)	Thermal Conductivity σ_{th}(W/m/K)	Dielectric constant κ
Si	4.05	1.12	5.17	**141**	**11.9**
Ge	4.13	0.66	4.79	59.9	16
GaAs	4.07	1.42	5.49	46	12.5
C diamond	0	5.47	5.47	**>2000**	**5.7**
4HSiC	3.55	3.00	6.55	500	6.52
InSb	4.59	0.16	4.75		16.0
SiO_2	1.10	9.00	10.1	1.38	3.9
Si_3N_4	2.00	5.00	7.00	30.1	7.5
Al_2O_3	1.92	6.2	8.12	25.1	10
HfO_2	2.07	5.6	7.67	11.4	24
ZrO_2	2.07	5.5	7.57	1.30	24
AlN	2.00	6.2	8.20	**175**	**8.9**
BeO	2.00	10.6	12.6	**260**	**6.7**

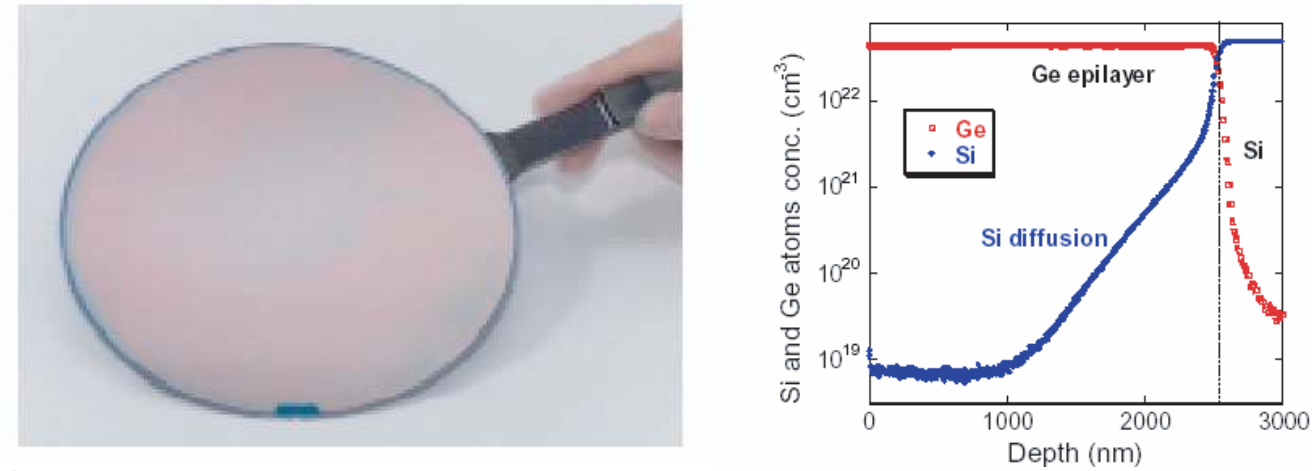


Fig. 18. Features of GeOI using epitaxial Ge on Si. (a) Top view photograph of a final GeOI wafer 200 mm in diameter ($TGe = 60$ nm, $TBOx = 400$ nm). The donor wafer is a 200 mm epiwafer [113]. (b) SIMS depth profile of the Si and Ge atoms inside a 2.5 μm thick Ge layer grown on Si(001) that has subsequently submitted to in situ anneals.

are measured. These results suggest that both the quality of the Ge layer and the gate stack have to be improved. Nevertheless I_{ON} vs. L_G state-of the-art values reported in Figure 19 for Ge and GeOI devices illustrate the excellent performances of our devices [114–117]. We have also performed TCAD simulations of GeOI MOSFET structures using a Ge CVT mobility model. The CVT parameters were theoretically calculated or adapted by calibration. From these simulations the I_{ON} current values for L_G down to 0.25 μm have been extracted, and show a good agreement with our electrical results and also with literature data [114–117].

7 Alternative CMOS or alternative to CMOS on silicon?

Many research teams are making efforts on Single Electron Transistors (SET) operation based on the Coulomb blockade principle. Demonstration of CMOS inverter operation at 27 K has been achieved by using a Vertical Pattern Dependent Oxidation (V-PADOX) process [118]. No solution has been found that could compete with CMOS devices. Some possibilities to achieve memory functional devices by using single electron trapping by a Coulomb blockade effect for DRAM [119], or Non Volatile applications [120–122] have been pointed out. This effect supposes that the Coulomb energy:

$$e^2/2C \tag{9}$$

is larger than the thermal energy of electrons kT (e is the electron charge; C is the capacitance of the quantum box). This energy is necessary to localize the electrons in a Coulomb box provided that tunneling is the limiting process: implicitly, one has to use very low capacitance and sufficiently high tunneling resistance. However, the Coulomb blockade process will be self limiting due to charge repulsion which reduces the speed of the charge transfer. Non Volatile Memory (NVM) applications can be envisaged by using trapping in nanometer size Si quantum

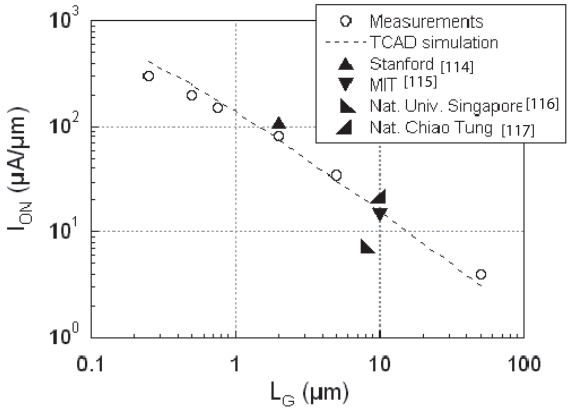


Fig. 19. Comparison of the ION performance of our GeOI P-MOSFETs ($LGmin = 0.25\ \mu$m) with literature. The ON current is measured for $VDS = -1.5$ V, $VGS - Vt = -2$ V. TCAD simulations of our GeOI devices show good agreement with the electrical results [113].

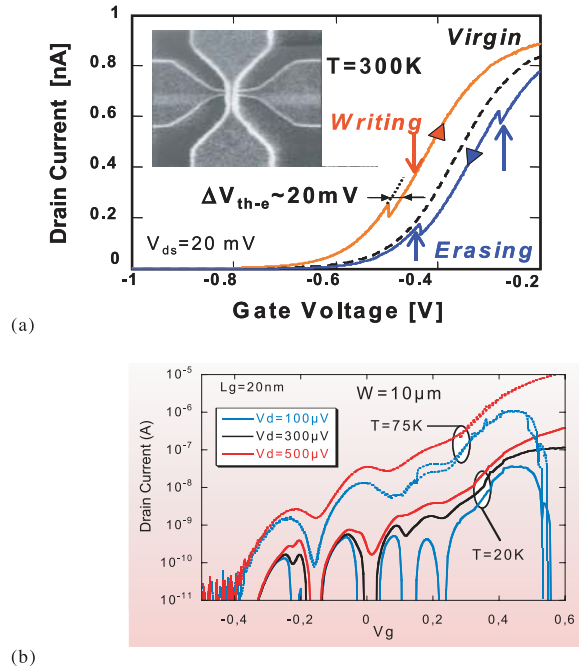


Fig. 20. Devices characteristics evidencing Single Electron phenomena (a) writing and erase characteristics of 20 nm × 20 nm ($W.L$) devices at room temperature. Top view of 20 nm × 20 nm nanowire [123] inserted. (b) Drain current oscillations in a $Lg = 20$ nm MOSFET at 75 and 20 K, demonstrating that Coulomb blockade is possible in such devices [5].

dot [121]: Al_2O_3 has been chosen as the tunnel insulator due to the increased dot density as compared to other materials (in the range of 10^{12} cm^{-2}), with reasonable interface states density (less than 10^{11} cm^{-2}). Recently, large capacity NVM's have been demonstrated [60]. Whether the involved writing or erase mechanisms are due or not to single electron transfer has been a controversial debate. In large area devices, with a large amount of randomly distributed Si-dots, it is very difficult to identify whether the single electron transfer is occurring or not, due to the large distribution of dot sizes and consequently of Coulomb energies. It is thus very important to use a device of the smallest size possible, containing only one dot or a low number of dots, to get a high sensitivity to single electron transfer. Such a result has been obtained at room temperature on 20 nm × 20 nm Non Volatile Memory Silicon wire based on Silicon quantum dots (Fig. 20a) [123]: current spikes on the writing or erasing characteristics have been identified as single electron trapping or detrapping respectively. Coulomb blockade oscillations can be observed if the series access resistance with the quantum well is high enough compared to the resistance quantum

$$(e^2/h)^{-1}. \tag{10}$$

This effect has already been reported on 50 nm gate length N channel MOS transistors at 4.2 K [125] making CMOS transistors attractive as single electron devices candidates. As gate length is scaled down to 20 nm, access resistance becomes larger and channel conductance oscillations appear at higher temperatures (here 75 K) (Fig. 20b) [4].

The Silicon Nanocrystals (Si-nc) technology (Fig. 21a) offers new scaling possibilities to Flash memories in the sub-90 nm nodes (Fig. 21b) [122] because of superior Stress Induced Leakage (SILC) immunity of the tunnel oxide. Thus NOR type architectures show a larger tolerance to threshold voltage fluctuations than NAND type devices [122]: if one considers a Si-nc density of 10^{12} cm^{-2}, NOR type can be scaled down to the 35 nm node whereas NAND type would reach the 65 nm node (Fig. 21b). In few electron memories, the stored charge discreteness makes these devices much sensitive to stochastic fluctuations of writing and retention times [126]: however the use of few electrons makes the Si-nc devices more attractive for low voltage, low power operation (Fig. 22) [126]. Double bit operation has also been demonstrated [122,127]. This solution is compatible with high standard retention times and endurance cycles [122], down to gate lengths of 35 nm [127].

8 Conclusions

By the end and beyond the end of the roadmap, power consumption will be the greatest issue whatever the application. We reviewed the physical limitations of MOSFET that will be encountered in the optimization of the performance versus leakage trade off and screened the different possibilities on the architecture or material sides. Multigate devices using strained channels will be widely used for high performance CMOS. Si based alloys or compatible semiconductors will be introduced to enhance the possibilities of future Systems on Chip. New materials including HiK dielectrics, Ge and C-diamond could be integrated to optimize integration density of logic circuits as well as

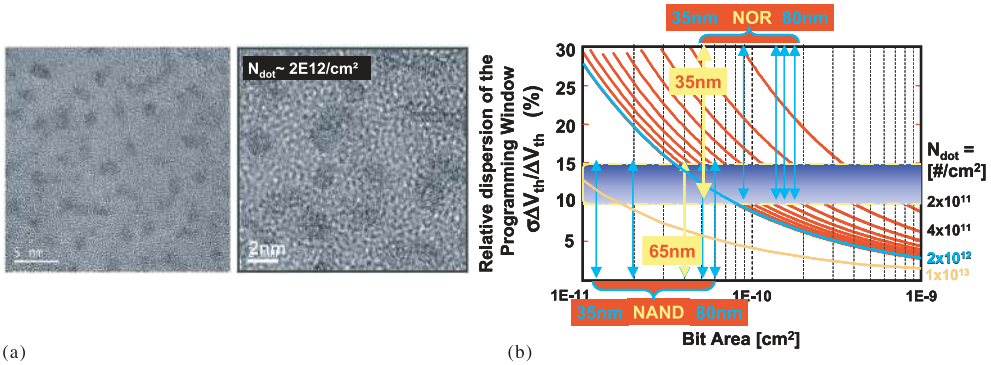


Fig. 21. Si-nc based Flash memories use (a) 2×10^{12} cm^{-2} CVD density of nanometers size Si dots; (b) the scaling of the devices will depend on their architecture and thus on their programming scheme [122].

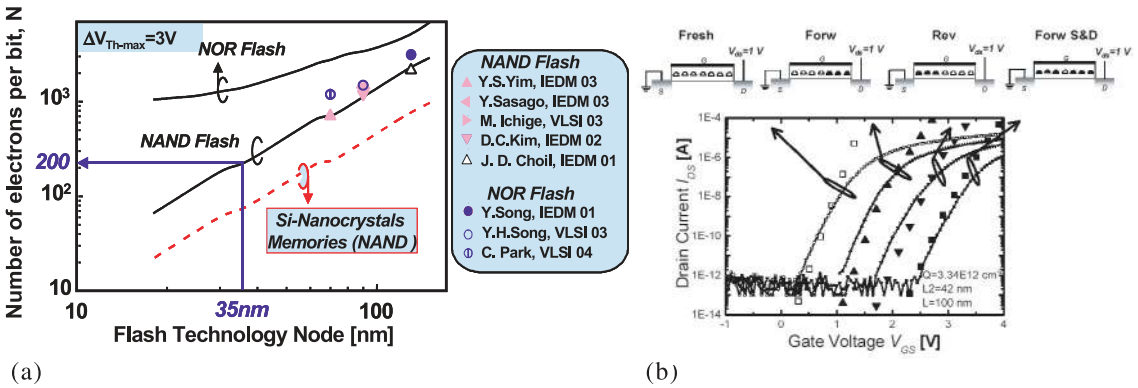


Fig. 22. Si-nc allow: (a) lower number of electrons per bit for programming: that reduces the programming voltages and power consumption [126]. (b) Double bit operation: transfer characteristics of a scaled SOI device charged consecutively on drain, source and on both sides with the same stressing conditions. Four clear states are apparent also if the two pockets of charge are very close to one another [127].

for limitation of short channel effects and power dissipation. New devices architectures requiring a low number of electrons for operation have good potentials in low power, low voltage Flash memories applications by the use of silicon nanocrystals. Single electronics will be a major study subject to optimize the use of ultra small devices.

I wish to warmly thank J. Gautier, B. de Salvo, L. Clavelier, T. Ernst, O. Faynot, T. Poiroux and M. Vinet from LETI – Electronics Nanodevices Laboratory for very fruitful discussions.

References

1. *The International Technology Roadmap for Semiconductors*, edn. 2005
2. H. Iwai, *IEDM Tech Digest 2004, San Francisco, CA, Dec. 2004*, pp. 11–14
3. S. Deleonibus et al., IEEE Electr. Device L. 173 (2000)
4. G. Bertrand et al., *Silicon Nanoelectronics Workshop 2000, Honolulu, HI, June 2000*, pp. 10, 11
5. G. Bertrand et al., Solid State Electron. **48**, 505 (2004)
6. K. Romanjek et al., IEEE Electr. Device L. **25**, 583 (2004)
7. S. Deleonibus et al., *ESSDERC Tech. Digest 1999, Leuven, Sept. 1999*, pp. 119–126
8. H. Iwai et al., *IEDM Tech. Digest 1998, San Francisco, CA, Dec. 1998*, pp. 163–166
9. C. Caillat et al., *VLSI Tech. Symp. Tech. Digest 1999, June 1999, Kyoto, Japan*, pp. 89, 90
10. Y. Taur et al., IBM J. Res. Dev. **39**, 245 (1995)
11. K. Gopalakrishnan et al., *IEDM2002 Tech. Digest, San Francisco, CA, Dec. 2002*, pp. 289–291
12. T. Oyamatsu et al., *VLSI Tech. Symp. Tech. Digest, Kyoto, Japan, June 1993*, pp. 89, 90
13. W. Schockley, Solid State Electron. **2**, 35 (1961)
14. V. De et al., *VLSI Tech. Symp. Tech. Digest 1998, Honolulu, HI, June 1998*, pp. 95, 96
15. W. Wong et al., *IEDM Tech. Digest 1993, Washington, DC, Dec. 1993*, pp. 705–708
16. T. Ohguro et al., IEEE T. Electr. Device L. **45**, 710 (1998)
17. T. Ernst et al., *VLSI Technology Symposium 2003 Tech. Digest, Kyoto, Japan*, pp. 51, 52
18. F. Ducroquet et al., *IEDM Tech. Digest 2004, San Francisco, CA, Dec. 2004*, pp. 437–440
19. M. Carroll et al., *IEDM Tech. Digest 2000, San Francisco, CA, Dec. 2000*, pp. 145–148
20. K. Rim et al., *IEDM Tech. Digest 1998, San Francisco, CA, Dec. 1998*, pp. 707–710
21. J.M. Hartmann et al., Semicond. Sci. Tech. **19**, 311 (2004)
22. E.A. Fitzgerald et al., Phys. Status Solidi A 171 (1999)
23. S. Bedell et al., *MRS Spring Meeting Proc., April 2004*

24. F. Andrieu et al., *ESSDERC 2003 Proc.*, *Estoril, Portugal, Sept. 2003*, pp. 267–270
25. J. Alieu et al., *ESSDERC 98*, *Bordeaux, France, Sept. 1998*, p. 144
26. A.-C. Lindgren et al., *ESSDERC 02*, *Bologna, Italy, Sept. 2002,* p. 175
27. N. Collaert et al., *Silicon Nanoelectronics Workshop Digest 2002*, *Honolulu, HI, June 2002*, pp. 15, 16
28. N. Collaert et al., *ESSDERC 02*, *Bologna, Italy, Sept. 2002*, p. 263
29. K.A. Jenkins, K. Rim, IEEE Electron. Device L. **23**, (2002)
30. G. Xia et al., IEEE T. Electron Dev. **51**, (2004)
31. H. Kawasaki et al., *IEDM 2004 Tech. Digest*, *San Francisco, CA, Dec 2004*, pp. 169–172
32. K. Rim et al., *VLSI Tech. Symp Digest 2002*, *Honolulu, HI, June 2002*, pp. 12, 13
33. O. Weber et al., *VLSI Tech. Symp. Digest 2004*, *Honolulu, HI, June 2004*, pp. 42, 43
34. F. Andrieu et al., *2006 VLSI Tech. Symp. Digest*, *Honolulu, HI, June 2006*, pp. 168-169 and F. Andrieu et al., *IEEE Intern. SOI Conf. Digest*, *Honolulu, HI, Oct. 2005*, p. 223
35. T. Tezukae et al., 2002 *VLSl Tech. Symp. Digest*, *Honolulu, HI, June 2002*, pp. 96, 97
36. F. Boeuf et al., *IEDM 2004 Digest*, *San Francisco, CA, Dec 2004*, pp. 425–428
37. F. Andrieu et al., *2005 VLSI Tech. Symp. Digest*, *Kyoto, Japan, June 2005*, pp. 176, 177
38. T. Mizuno et al., *2003 VLSI Tech. Symp. Digest*, *Kyoto, Japan, June 2003*, pp. 97, 98
39. M. Yang et al., *2004 VLSI Tech. Symp. Honolulu, HI, June 2004*, pp. 160, 161
40. M.L. Lee, E.A. Fitzgerald, *IEDM 2003* Dig, *Washington, DC, Dec. 2003,* pp. 131–429
41. A. Chatterjee et al., *IEDM Tech. Digest 1998*, *San Francisco, CA, Dec. 1998*, pp. 777–780
42. A. Yagashita et al., *IEDM Tech. Digest 1998*, *San Francisco, CA, Dec. 1998*, pp. 785–788
43. B. Guillaumot et al., *IEDM 2002 Tech. Digest*, *San Francisco, CA, Dec 2002*, pp. 335–338
44. B. Tavel et al., *IEDM 2001 Digest*, *Washington, DC, Dec 2001*, pp. 825–828
45. J. Kedzierski et al., *IEDM 2002 Digest*, *San Francisco, CA, Dec. 2002*, pp. 247–250
46. W.P. Maszara et al., *IEDM 2002 Digest*, *San Francisco, CA, Dec. 2002*, pp. 367–370
47. C. Hobbs et al., *VLSI Tech. Symp. 2003 Tech. Digest*, *Kyoto, Japan*, pp. 9, 10
48. D. Souil et al., *3rd ULIS Workshop 2002*, *Munich, FRG, March 2002*, pp. 139–142
49. G. Timp et al., *IEDM Tech. Digest 1998*, *San Francisco, CA, Dec. 1998*, pp. 615–618
50. S. Takagi et al., *IEDM Tech. Digest 1998*, *San Francisco, CA, Dec. 1998,* pp. 619–622
51. J. Lee et al., *IEDM Tech. Digest 1999*, *Washington, DC, Dec. 1999*, pp. 133–136
52. H. Iwai et al., *IEDM Tech. Digest 2002*, *San Francisco, CA, Dec 2002*, pp. 625–627
53. O. Weber et al., *IEDM 2004, San Francisco, CA, Dec. 2004*, pp. 867–670
54. J.L. Pelloie, *ISSCC Tech. Digest 1999*, *San Francisco, CA, Feb. 1999*, p. 428
55. L. Leobandung et al., *IEDM Tech. Digest 1998*, *San Francisco, CA, Dec. 1998*, pp. 403–407
56. J. Lolivier et al., *ECS Spring 2003 Proc.*, *Paris, France, April 2003*, pp. 379
57. B. Doris et al., *IEDM 2002 Tech. Digest*, *San Francisco, CA, Dec 2002*, pp. 267–270
58. J. Lolivier et al., *SOI Conference 2004*, *Charleston, SC, October 2004*, pp. 17, 18
59. H. Nakayama et al., *IEEE SOI Conf. 2000 Proc.*, *Wakefield, Mass, Oct 2000*, pp. 128, 129
60. K. Oshima et al., *SOI Conference 2002 Tech. Digest*, *Oct 2002*, pp. 95, 96
61. H.S.P. Wong et al., *IEDM Tech. Digest 1997*, *Washington, DC, Dec. 1997*, pp. 427–430
62. F. Allibert et al., *ESSDERC 2001*, *Nurnberg, FRG, Sept 2001*, pp. 267–270
63. B. Yu et al., *IEDM 2002 Tech. Digest*, *San Francisco, CA, Dec 2002*, pp. 251–253
64. J. Kedzierski et al., *IEDM 2002 Tech. Digest*, *San Francisco, CA, Dec 2002*, pp. 247–250
65. F.L. Yang et al., *IEDM 2002 Tech. Digest*, *San Francisco, CA, Dec 2002*, pp. 255–258
66. E.C. Jones et al., *Dev. Res. Conf. 2001 Digest, June 2001*
67. T. Sekigawa, Solid State Electron **27**, 827 (1984)
68. D. Hisamoto, in *Tech. Digest of IEDM (1998)*, pp. 833–836
69. F. Balestra et al., IEEE Electr. Device L. **8**, 410 (1987)
70. M. Vinet et al., IEEE Electr. Device L. **26**, 317 (2005)
71. M. Vinet et al., *ECS Spring meeting 2005 Proc.*, *Québec, CA, May 2005*, pp. 285–296
72. T. Poiroux et al., *ULIS2005 Proc*, *Bologna, Italy, April 2005*, pp. 71–74
73. T. Poiroux et al., Microelectron. Eng. **80**, 378 (2005)
74. M. Vinet et al., *Int. Conf. SSDM 2004 Proc*, *Tokyo, Japan, Sept 2004*, pp. 768, 769
75. J.P. Colinge et al., *IEDM 1990 Digest*, *San Francisco, CA, Dec 1990*, pp. 595–598
76. S. Harrison et al., *IEDM 2003 Digest*, *Washington, DC, Dec 2003*, pp. 449–452
77. K.W. Guarini et al., *IEDM 2001 Digest*, *Washington, DC, Dec 2001*, pp. 425–428
78. J.H. Lee et al., *IEDM 1999 Digest*, *Washington, DC, Dec 1999*, pp.71–74
79. J. Lolivier et al., *ESSDERC 2004 Proc.*, *Leuven, Belgium, Sept 2004*, pp. 177–180
80. X. Huang et al., *IEDM 1999 Digest*, *Washington, DC, Dec. 1999*, pp. 67–70
81. B. Doyle et al., *VLSI Tech. Symp. 2003 Digest*, *Kyoto, Japan, June 2003*, pp. 133, 134
82. F.L. Yang et al., *IEDM 2002 Digest*, *San Francisco, CA, Dec 2002*, pp. 255–258
83. J.T. Park et al., IEEE Electr. Device L. **22**, 405 (2001)
84. F.L. Yang et al., *VLSI Tech. Symp. 2004 Digest*, *Honolulu, HI, June 2004*, pp. 196, 197
85. Y.K. Choi et al., Solid State Electron **46**, 1595 (2002)
86. C. Jahan et al., *IEEE VLSI Tech. Symp. 2005*, *Kyoto, Japan, June, 2005*, pp. 112, 113
87. J.P. Colinge, Solid State Electron **48**, 897 (2004)
88. J.G. Fossum et al., IEEE T. Electron Dev. **49**, 808 (2002)
89. H.S.P. Wong et al., *IEDM 1998 Digest*, *San Francisco, CA, Dec 1998*, pp. 407–410
90. T. Ernst et al., in IEEE T. Electron Dev. **50**, 830 (2003)
91. F. Gamiz et al., J. Appl. Phys. **94**, 5732 (2003)

92. K. Suzuki et al., IEEE T. Electron Dev. **40**, 2326 (1993)
93. J.M. Ha et al., *IEDM Tech. Digest 1998*, *San Francisco, CA, Dec. 1998*, pp. 639–642
94. K. Goto et al., *IEDM Tech. Digest 1997*, *Washington, DC, Dec. 1997*, pp. 471–474
95. M. Takase et al., *IEDM Tech. Digest 1997*, *Washington, DC, Dec. 1997*, pp. 475–478
96. Y. Sasaki et al., *VLSI Techn. Symp. 2004 Tech. Digest*, *Honolulu, HI, June 2004*, pp. 180, 181
97. T. Noguchi et al., Proc. Mat. Res. Soc. **146**, 35 (1985)
98. C. Laviron et al., *2nd IWJT, IEEE-Cat. No. 01EX541C, 2001, Tokyo, Japan, Nov. 2001*, pp. 91–94.
99. T. Ohguro, *ECS Symp on ULSI 1997*, *Montreal, CA, Oct. 1997*, p. 275
100. R. Nuryadi et al., J. Vac. Sci. Tech. B **20**, 167 (2002)
101. E. Dubois, G. Larrieu, Solid State Electron 997 (2002)
102. B.Y. Tsui, C.P. Lin, IEEE Electron Device L. 430 (2004)
103. G. Niu et al., IEEE T. Electron Dev. **46**, 1912 (1999)
104. K. Chen et al., Solid State Electron **39**, 1515 (1996)
105. S. Datta et al., Superlattice. Microst. **23**, 771 (1998)
106. F. Assad et al., IEEE T. **47**, 232 (2000)
107. F. Assad et al., *IEDM Tech. Digest*, *Washington, DC, Dec. 1999*, pp. 5547–5550
108. A. Rinetour et al., *IEDM Tech. Digest*, *Washington, DC, Dec. 2003*, pp. 433–436
109. Chi On Chui et al., *IEDM Tech. Digest*, *San Francisco, CA, Dec. 2002*, pp. 437–440
110. Nishitami-Gamo et al., Diam. Relat. Mater. **9**, 941 (2000)
111. Lagrange et al., Carbon **37**, 807 (1999)
112. S. Deleonibus et al., Int. J. High Speed Electron. Syst. **16**, 193 (2006)
113. L. Clavelier et al., *2005 Silicon Nanolectronics Workshop*, *Kyoto, Japan, June 2005*, pp. 18, 19
114. C. On Chui et al., *IEDM 2002*, pp. 437–440
115. A. Ritenour et al., *IEDM 2003*, pp. 433–436
116. S. Zhu et al., IEEE Electron Device L. **26**, 81 (2005)
117. D.S. Yu et al., IEEE Electron Device L. **25**, 138 (2004)
118. Y. Ono et al., *IEDM 2001 Tech. Digest*, *Washington, DC, Dec 2001*, pp. 367–370
119. S. Tiwari et al., Appl. Phys. Lett. **68**, 1377 (1996)
120. K. Yano, *IEDM Tech. Digest 1998*, *San Francisco, CA, Dec. 1998*, pp. 107–110
121. A. Fernandes et al., *IEDM 2001 Tech. Digest*, *Washington, DC, Dec 2001*, pp. 155–158
122. B. de Salvo et al., *IEDM Tech. Digest 2003*, *Washington, DC, Dec. 2003*, pp. 597–600
123. G. Molas et al., *WODIM 2002 Proc.*, *Grenoble, France, Nov. 2002*, pp. 175–178
124. M. Sanquer et al., *SNW 2003*, *Kyoto, Japan*, pp. 70–71
125. M. Specht et al., *IEDM Tech. Digest 1999*, *Washington, DC, Dec. 1999*, pp. 383–341
126. G. Molas et al., *IEDM Tech. Digest 2004*, *San Francisco, CA, Dec. 2004*, pp. 877–880
127. L. Perniola et al., *IEDM Tech. Digest 2005*,*Washington, DC, Dec. 2005*, pp. 877–880