JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

Despite the fact that a range of limitations are beginning to appear as CMOS technology is being raised to ever higher levels of perfection, it is anticipated that silicon will be the dominant material of the semiconductor industry for at least the first half of the 21st century. The forecast for microelectronics development published in 2006 by Semiconductor Industry Association (SIA) reaches ahead to the years 2014–2020. Moreover, a comparison with former SIA forecasts indicates that they become more aggressive (that is more optimistic) with time.

While the development of silicon microelectronics in the past could be attributed mostly to the reduction of the feature size (progress in lithography), today it relies more on new material solutions, such as SOI, SON, SiGe or SiC. The combination of this trend with continuous miniaturization provides the opportunity of improving IC functionality and speed of operation.

Telecommunications and information technology are arguably the most powerful drivers behind microelectronics product development nowadays. Plenty of new applications are being created for fast analog and rf circuits, as well as for information processing ones. It is clear that with the anticipated peak $f_{\max} = 425$ GHz and $f_T = 385$ GHz to be reached by rf SiGe-base bipolar transistors in 2011, according to the 2006 issue of ITRS, a lot of effort must be put into the development of appropriate material, processing, characterization and modeling. While progress in the bipolar technology is impressive, the increase of MOSFET speed is even more so. The same issue of ITRS predicts on-chip clock of 73 GHz for 2020, which will require MOSFET internal switching speed of 12 500 GHz.

High-speed isn't, however, everything. Portable wireless products push, for obvious reasons, for low-power solutions. This trend requires new architectural solutions (e.g., channel thinning), and in consequence, new material, such as SOI (or its possible successor SON), where current driveability is considerably higher than in conventional MOSFETs.

In this issue the Reader will find the invited lectures presented during the 7th Symposium Diagnostics & Yield: Advanced Silicon Devices and Technologies for ULSI Era, which took place at Warsaw University of Technology on June 26–28, 2006. A number of the papers are devoted to advanced materials, such as SOI, SiC and SiGe and the most important issues concerning semiconductor technology (difficulties faced by CMOS technology, gate-dielectric

fabrication, wafer-cleaning problems). Several papers address also semiconductor structure characterization (DC and noise analysis, wideband characterization) and modeling (negative bias temperature instability).

We hope the Readers will find this issue useful and interesting.

Organization of this Symposium would not be possible without the support of SINANO – a Network of Excellence funded by the European Commission under the 6th Framework Programme (EC contract no. 506844) and Polish Ministry of Science and Higher Education (research projects no. 3T11B 012 28, 3T11B 020 30). The organizers are also grateful to the National Institute of Telecommunications for making this *Journal* available for the publication of the Symposium papers.

Andrzej Jakubowski
Lidia Łukasiak
Guest Editors

Invited paper

SOI nanodevices and materials for CMOS ULSI

Francis Balestra

Abstract— A review of recently explored new effects in SOI nanodevices and materials is given. Recent advances in the understanding of the sensitivity of electron and hole transport to the tensile or compressive uniaxial and biaxial strains in thin film SOI are presented. The performance and physical mechanisms are also addressed in multi-gate Si, SiGe and Ge MOSFETs. The impact of gate misalignment or underlap, as well as the use of the back gate for charge storage in double-gate nanodevices and of capacitorless DRAM are also outlined.

Keywords— ballistic transport, gate misalignment, GIFBE, mobility enhancement, SOI, strain engineering, tunneling current.

1. Introduction

The silicon-on-insulator (SOI) devices are the best candidates for the ultimate integration of ICs on silicon. The flexibility of the SOI structure and the possibility to realize new device architectures allow optimum electrical properties to be obtained for low power and high performance circuits. These transistors are also very interesting for high frequency and memory applications [1–3]. In this paper, an overview of recently explored new effects in advanced SOI devices and material is given. The advantages and drawbacks of a number of new device architectures are also addressed.

2. Physical mechanisms in advanced SOI MOSFETs

Ultra-thin gate oxide (sub-2 nm) leads to direct gate tunneling currents [4] that consist of three main streams of carriers (Fig. 1). In partially-depleted (PD) SOI MOSFETs, the floating body of the device is isolated by the buried oxide (BOX) and charged by the direct tunneling currents, J_{EVB} and J_{HVB}. When a floating-body device is biased in inversion, the body is mainly charged by a hole current resulting from the tunneling of valence band electrons into the gate ($J_{HVB} \ll J_{EVB}$). When biased in accumulation, the body is charged with electrons coming from the gate conduction band. These currents strongly affect the body potential of the PD devices, giving rise to gate-induced floating body effect (GIFBE). The different gate current contributions are plotted in Fig. 1 to illustrate the body-charging mechanism.

A direct consequence of the GIFBE is the sudden increase of the drain current characteristics for V_G close to 1.1 V.

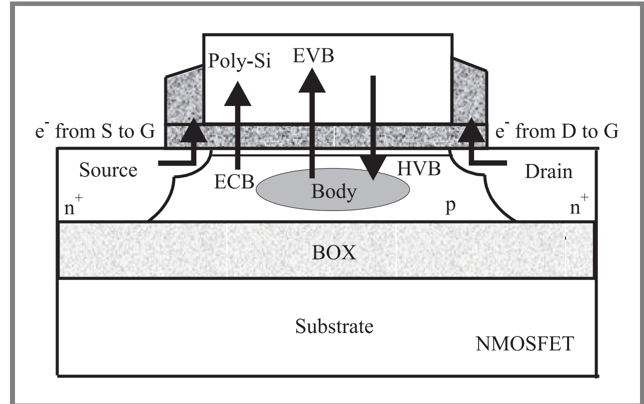


Fig. 1. Tunneling current components in a NMOSFET. Explanations: EVB – valence band electron tunneling, ECB – conduction band electron tunneling, HVB – valence band hole tunneling.

For this voltage, the gate-to-body current (I_{GB}) charges up the body and the drain current increases. This "kink-like" effect gives rise to a strong second peak in transconductance (up to 40% increase), which clearly appears in Fig. 2 for low drain bias. This figure illustrates the influence of the gate size on the GIFBE's amplitude and position. The voltage corresponding to the onset of the 2nd peak of transconductance (G_m) is nearly independent of the gate length (and width) whereas the amplitude of the peak depends on the device geometry. The 2nd peak is clearly reduced as the gate length (or width) are shrunk down. It is usually reported that floating body effects (FBEs) are reduced in short-channel devices by enhanced junction leakage or in narrow-channel devices by

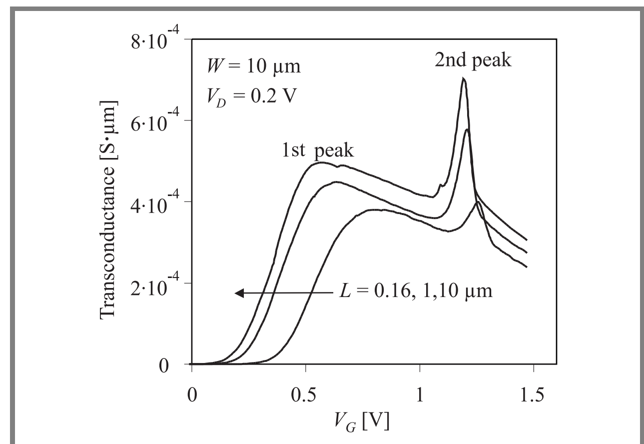


Fig. 2. Normalized transconductance of a 10 μm wide NMOSFET versus gate bias for various gate lengths.

increased recombination rate near the sidewalls. In both cases, the removal of majority carriers from the body is more efficient, hence the body charging by I_{GB} is less effective and the GIFBE is reduced. However, even in the smallest transistor, where both junction and sidewall contributions occur, the role of the gate tunneling current remains significant.

The drain power spectral density also presents a special behavior [4]. For V_G values inferior to the GIFBE onset gate voltage (around 1.1–1.2 V), conventional $1/f$ noise is observed, attributed to carrier fluctuations from the inversion layer due to carrier trapping/detrapping in the vicinity of the silicon/SiO_2 interface. Nevertheless, an excess noise occurs, characterized by the superposition of a Lorentzian-like component on the $1/f$ noise when the GIFBE is present. Similarly to FB PD SOI devices in saturation mode, a flat plateau is followed by a $1/f^2$ roll-off at a given corner frequency. In this case, the corner frequency shifts to higher frequencies as the drain bias increases: here, the front gate bias plays the role of the drain bias, and we have a similar behavior with frequency as the Kink-related excess noise.

From more than two decades for $L = 10$ μm, the excess noise decreases down to only one decade or less, and becomes almost insignificant for short devices ($L = 0.20$, 0.12 μm). Figure 3 represents the calculated ratio between the maximum drain current power spectral density ($S_{Id\,\max}$) and the minimum one ($S_{Id\,\min}$, value of the plateau at low V_G without GIFBE).

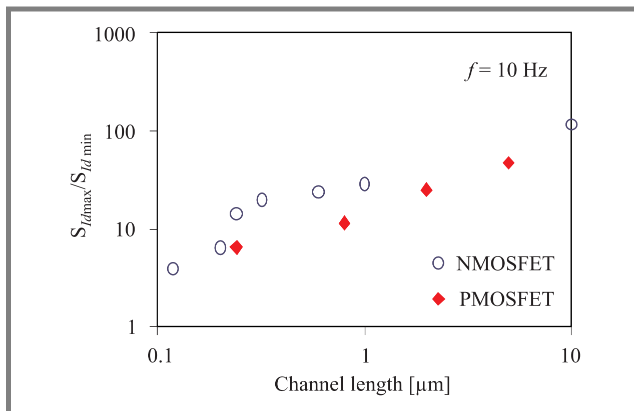


Fig. 3. Ratio between the maximum drain current noise ($S_{Id\,\max}$) and the minimum one ($S_{Id\,\min}$, value of the plateau at $V_G = 0.9$ V) for N- and PMOSFETs.

Two general features may explain the obtained results. On the one hand, the magnitude of the second transconductance peak is reduced as the channel length is shortened (FBEs are usually lowered by enhanced contributions from junctions), and the role of the gate current is partially offset, so that we notice a reduced contribution of the G_m 2nd peak on the noise overshoot. On the other hand, reducing the channel length causes an enhancement of the $1/f$ noise level, and this higher noise level probably masks the excess noise due to the GIFBE.

The GIFBE in a twin-gate (TG) structure (Fig. 4) is significantly reduced [5].

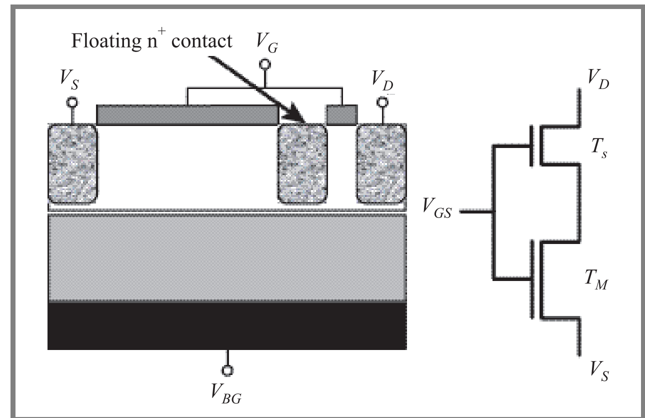


Fig. 4. Twin-gate NMOSFET.

In particular, the impact of the TG structure is pronounced on the Lorentzian noise overshoot (Fig. 5). Such a reduction results from a lowering of the part of the EVB current that reaches the source junction (the holes from the slave part (T_s) of the TG device are screened from reaching the source by recombination at the inner n^+ contact).

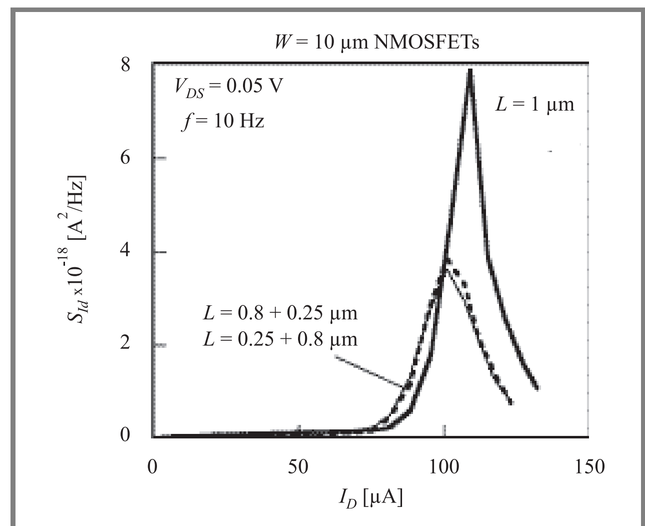


Fig. 5. Spectral density of drain current noise (S_{Id}) versus drain current (I_D) for NMOSFET (bold line) compared with the two TG combinations.

A GIFBE is also observed in fully depleted FinFET when a back gate bias is applied leading to an accumulation at the bottom of the fin (Fig. 6) [6].

In a double gate MOSFET, the application of a back gate voltage can lead to a volume inversion and to a screening reducing the number of trapped carriers in the gate oxides. This phenomenom induces a reduction of the low frequency noise (Fig. 7) [7].

The self-heating effect is also a harmful parasitic effect in SOI. The traditional buried silicon dioxide has a poor thermal conductivity that leads to an enhancement of the chan-

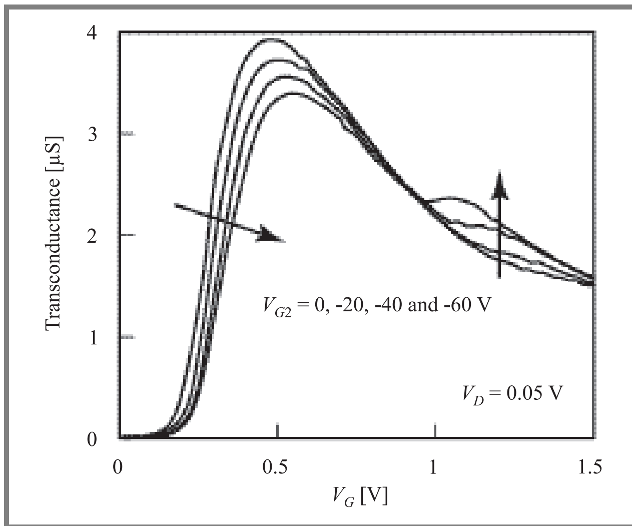


Fig. 6. Measured transconductance of a FD FinFET for different values of back gate bias ($L = 10\ \mu$m).

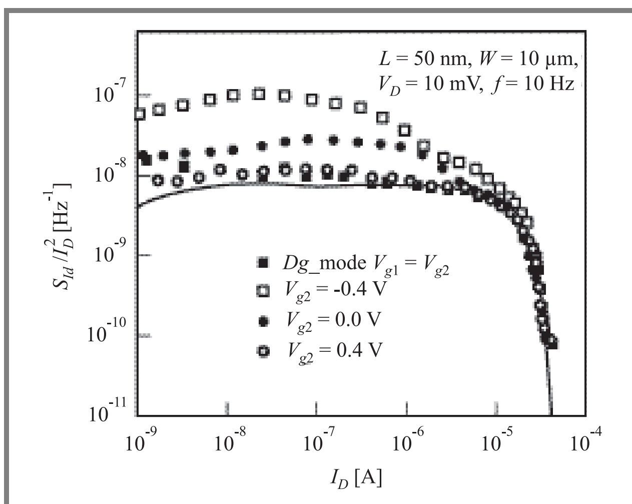


Fig. 7. Normalized drain current noise of a double gate NMOSFET for different back gate biases. Solid line: $S_{VG}\ (G_m/I_D)^2$ for double gate mode.

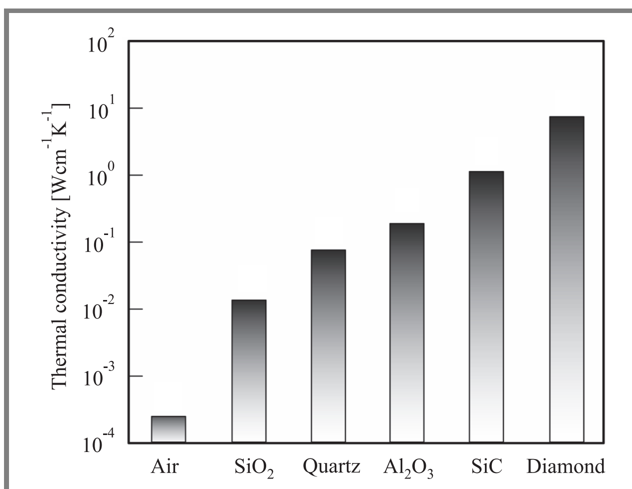


Fig. 8. Thermal conductivity of various buried oxide materials.

nel temperature and thus a reduction of carrier mobilities and drain current. The thermal conductance of various buried insulator materials is shown in Fig. 8 [8]. As it is shown in this figure, many insulators have a better thermal conductivity compared to SiO_2. In addition, diamond and quartz are also best suited dielectrics for controlling short channel effects and therefore to replace SiO_2. SiC and Al_2O_3 needs the use of thin buried insulator together with a ground plane architecture.

On the other hand, it is worth noting that the thermal conductivity of Ge films is lower than that of Si films for bulk materials (Fig. 9) [9]. However, for ultra-thin films,

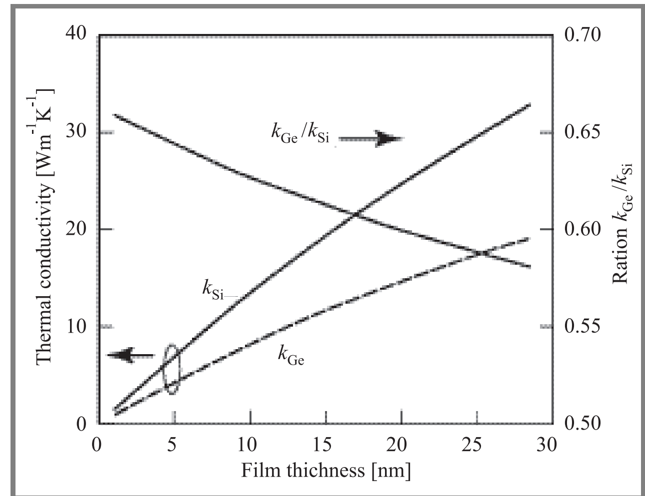


Fig. 9. Estimated thermal conductivity of thin Si and Ge layers.

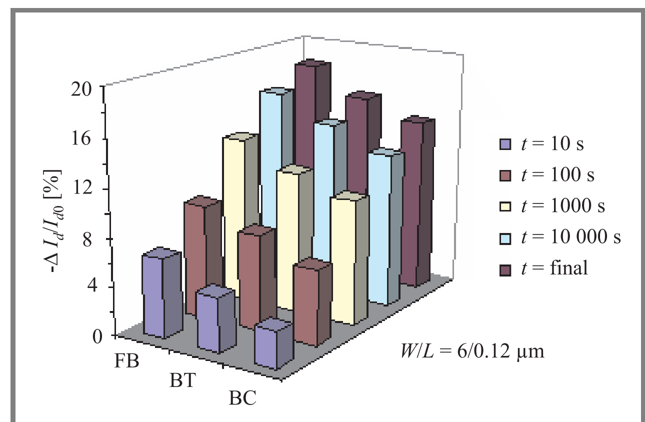


Fig. 10. Drain current degradation measured at $V_{GT} = 0.5$ V and $V_D = 50$ mV for various NMOSFET architectures in the worst-case aging scenario (aging conditions: $V_G = V_D = 2.2$ V).

these values are very close and therefore Ge films will present similar self-heating (SH) effects as Si films for deep sub-0.1 μm devices realized on nanometric layers.

Hot carrier effects are limiting long term device reliability. In SOI structures, special hot carrier regimes exist. Figure 10 shows the relative degradation of the drain current for various PD device architectures: floating body (FB), body connected (BC) and body tied (BT) [4]. This figure

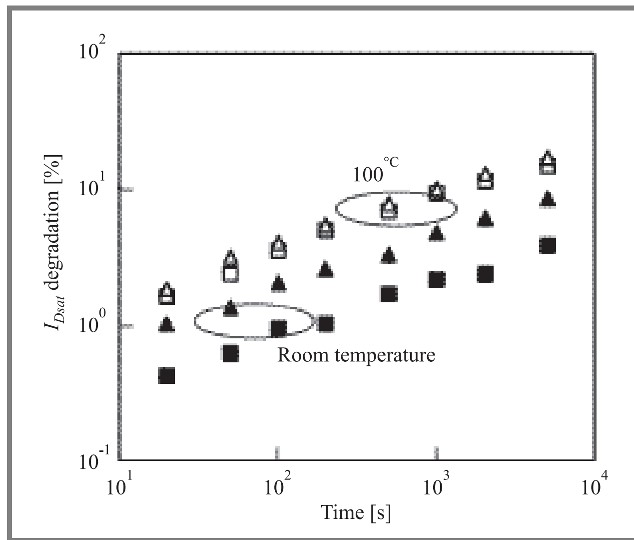


Fig. 11. Channel-width dependence of hot-carrier-induced degradation at room and high temperatures (squares – narrow, triangles – wide).

is plotted for the worst case aging in advanced SOI devices (maximum gate current, $V_G = V_D$). Body connected devices exhibit enhanced hot carrier immunity because of the collected hole coming from the impact ionization at the drain edge. Device degradation is also lowered for narrow channels due to reduced floating body effects (Fig. 11) [10].

3. Influence of strain and surface orientation on the electrical properties of thin layers on insulators

Compressive and tensile biaxial and uniaxial stress silicon technologies are promising for enhancing CMOS performance in bulk and SOI MOSFETs. The combination of strained layers and ultra thin films SOI structures is one of the best candidates for decananometer MOSFETs.

Figure 12 is a plot of the dependence of electron and hole mobilities as a function of the charge density [11]. The strained Si layer is fabricated with sacrificial thin relaxed SiGe and smart cut. In the SSOI devices, substantial enhancements of both electron (about 100%) and hole (about 50%) mobilities are obtained compared with the control SOI device at intermediate charge densities for long channel transistors.

An enhancement of the electron mobility of about 15–20% has been obtained for short channels (70 nm technology) SGOI MOSFETs (strained Si-on-SiGe-on-insulator) together with superior short channel control [12, 13]. Figure 13 shows the enhancement of the drain current for sub-0.1 μm devices.

In Fig. 14, the electron mobilities are represented for various Ge content of the SiGe layer and different Si film thicknesses. The electron mobility enhancement is max-

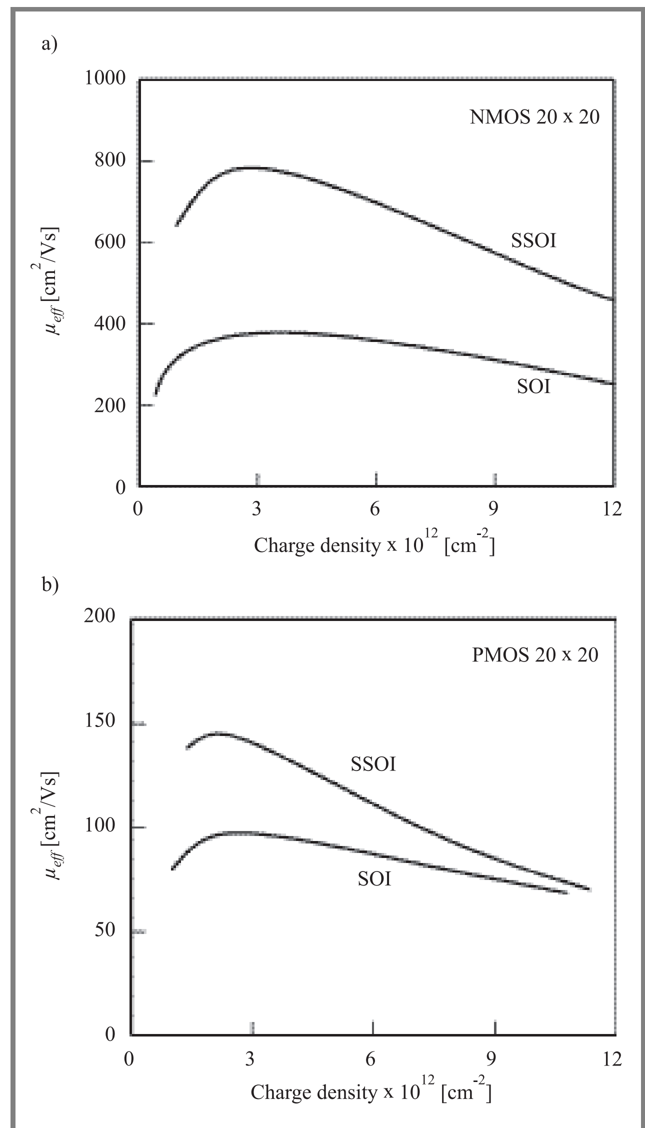


Fig. 12. Effective mobility comparison between SSOI and SOI MOSFETs: (a) electron mobility; (b) hole mobility.

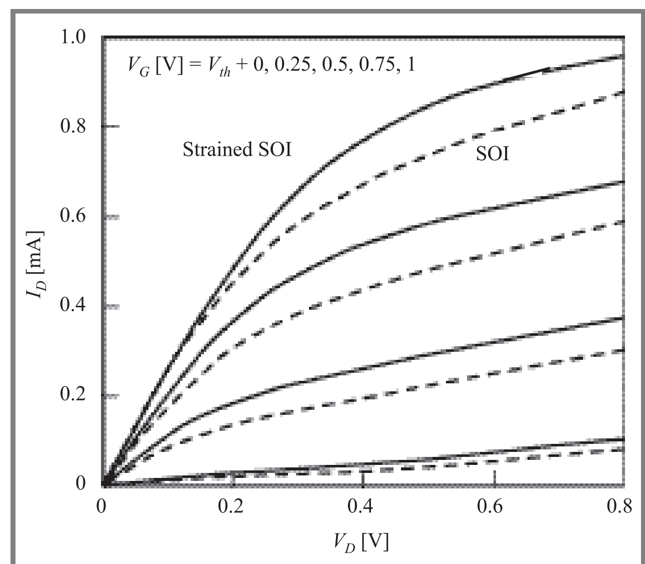


Fig. 13. I_D–V_D characteristics of 70 nm MOSFETs ($W = 1$ μm).

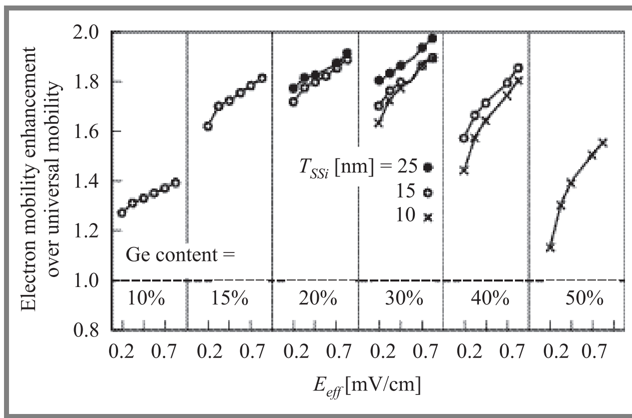


Fig. 14. Effective-field (E_{eff}) dependence of electron mobility enhancement as a function of Ge content and film thickness.

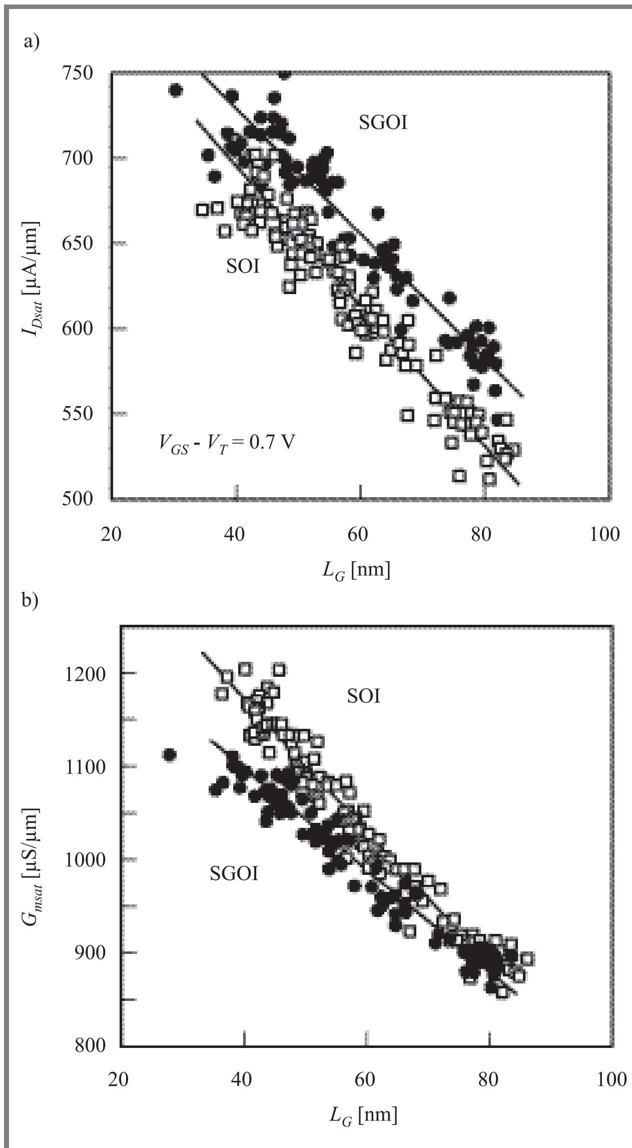


Fig. 15. Comparison of I_{Dsat} (a) and G_{msat} (b) at a constant gate overdrive.

imum for 30% of Ge due to the increase in alloy scattering and/or surface roughness for higher Ge concentrations and the hole mobility continuously increases with Ge up to 50% [13]. It is also worth noting that the enhancement of carrier mobility is reduced in thinner strained Si films due to interface states and fixed charges induced by the diffusion of Ge atoms to the interfaces.

Figure 15 shows I_{Dsat} and G_{msat} as a function of channel length for SGOI and SOI MOSFETs. An enhancement of I_D is outlined down to sub-50 nm transistors for SGOI, but the difference diminishes at smaller channel lengths due in particular to larger SH in SiGe than in Si. This SH effect in SGOI degrades G_{msat}, which is more sensitive to SH than I_D. Therefore the transconductance appears degraded in SGOI as compared to SOI but after correction of the self-heating a similar increase is obtained for I_D and G_m in the SGOI structure [14].

The HOI structure (strained Si/strained SiGe/ strained Si heterostructure-on-insulator) presents also substantial electron and hole mobility enhancements [15]. In particular, hole mobilities are very high for thin Si cap layer (enhancement of about 100%) compared with the universal SOI mobility and are also significantly higher that the best SSDOI mobility (strained Si directly-on-insulator) due to the compressively strained buried SiGe channel (Fig. 16).

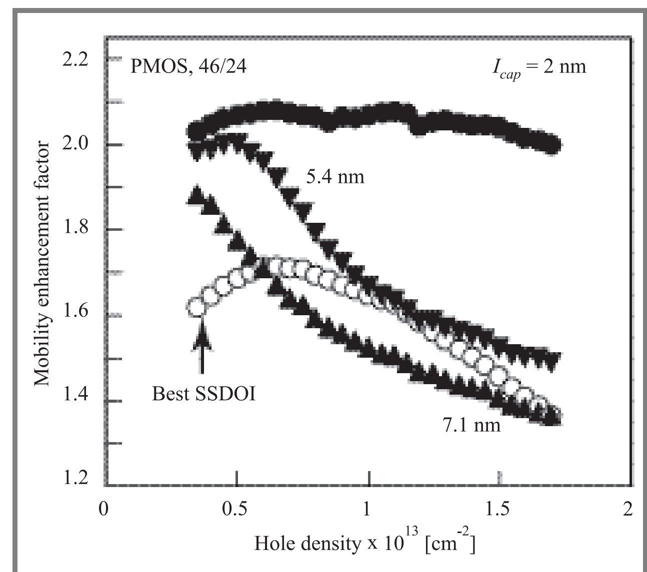


Fig. 16. Mobility enhancement in HOI compared with the best SSDOI curve relative to the universal SOI mobility.

Uniaxial strain engineering is also useful for mobility enhancement for Si film thickness in the sub-10 nm range [16]. A similar enhancement of electron mobility in 3.5 nm SOI devices under biaxial and uniaxial tensile strain has been obtained. The electron mobility is also enhanced in 2.3 nm Si layer under uniaxial tensile strain (Fig. 17), and the hole mobility increases in 2.5 nm film under uniaxial compressive strain.

It has recently been shown that the use of a metal gate (TiN) can induce significant compressive stress along the channel direction. This stress is increased as the gate length decreases. This phenomenon progressively degrades electron

mobility while hole transport is improved. Similar behaviors are obtained in single and double gate SOI devices, and the use of ⟨110⟩ channel orientation is the most favorable in terms of electrical performance [17].

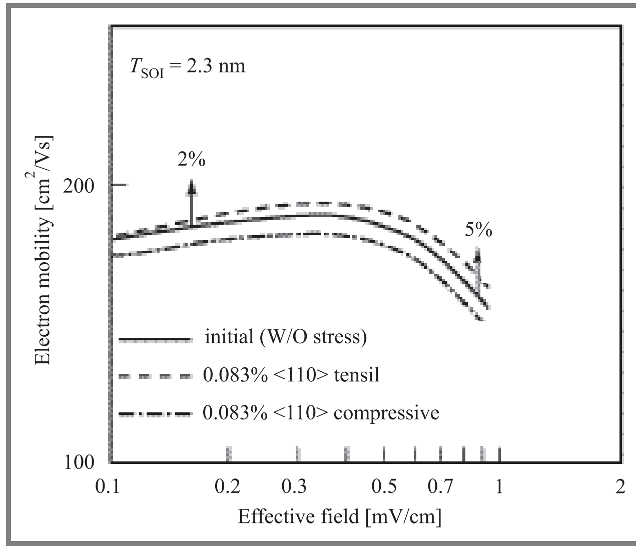


Fig. 17. Electron mobility in 2.3 nm ultra-thin-body MOSFET under ⟨110⟩ uniaxial strain.

Pure Ge channel MOSFETs are also considered as one promising option for future high performance CMOS. A compressively strained Ge channel is expected to further enhance hole mobility due to the very small effective hole mass [18]. Figure 18 shows the linear current of s-Ge PMOS with HfO_2 gate dielectrics along with the Si control device. A 2.5 × performance enhancement is observed (similar enhancement for the transconductance). For s-Ge

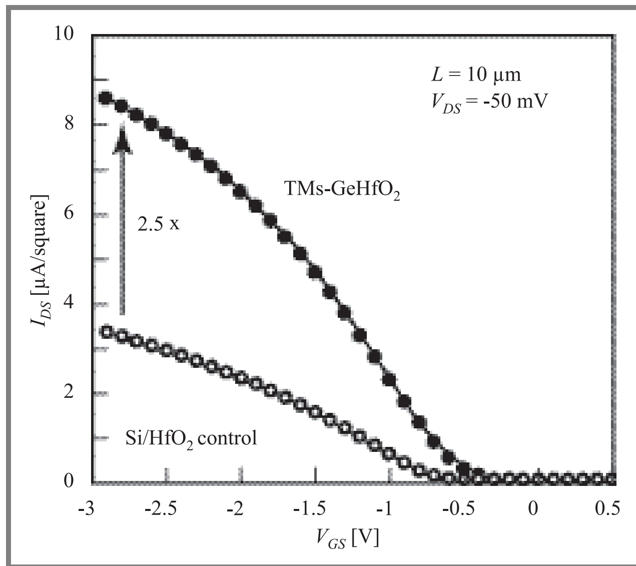


Fig. 18. Drain current of PMOSFETs with HfO_2 gate oxide on 60% Ge channel formed by local thermal mixing compared with Si PMOS control with HfO_2.

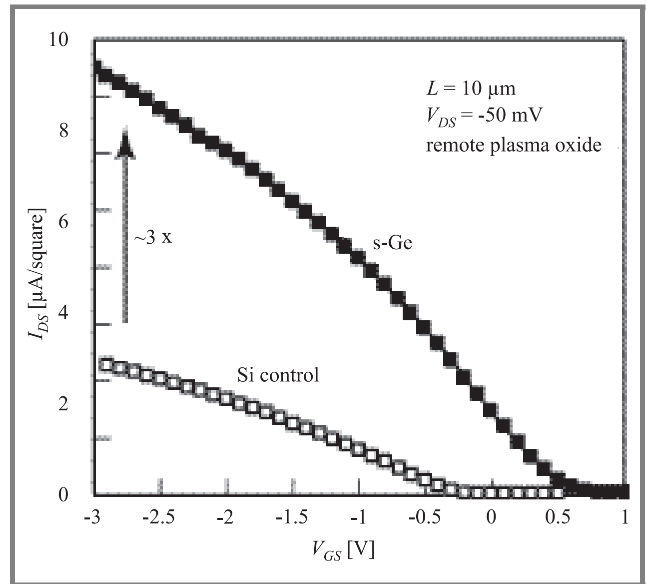


Fig. 19. Drain current of PMOSFETs with remote plasma oxide on 100% Ge channel formed by selective UHVCVD compared with Si channel PMOS control with the same oxide.

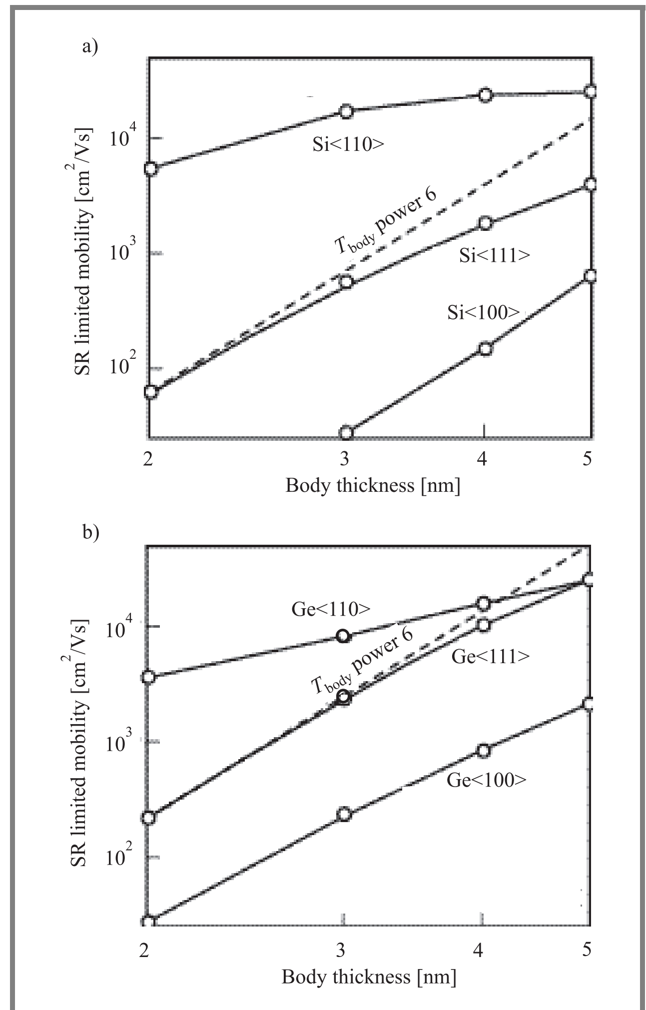


Fig. 20. Simulated surface-roughness limited hole mobility for Si (a) and Ge (b) with various orientations. Hole density $5 \cdot 10^{11}$ cm^{-2}.

P-type devices with SiO_2 gate oxide, a 3 × drive current and transconductance is obtained (Fig. 19).

The influence of surface-roughness (SR) in ultra-thin films is very important. Figure 20 shows the SR limited hole mobility as a function of body thickness for Si (SOI) and Ge (GOI) channels. The variation of hole mobility is outlined for various surface orientations [19].

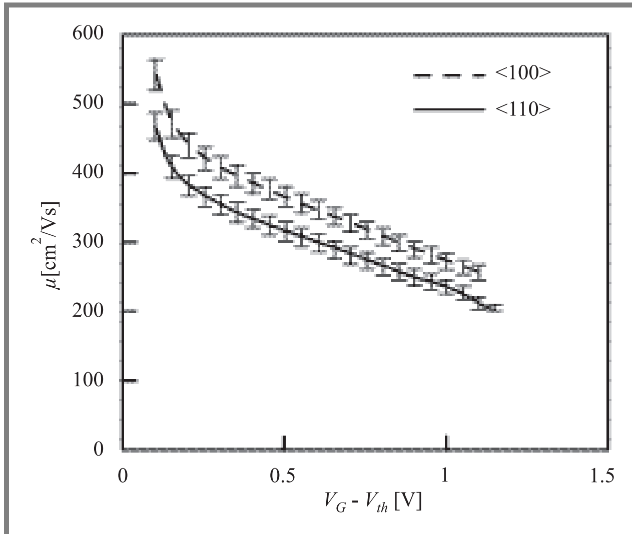


Fig. 21. Electron mobility of FinFETs with ⟨100⟩ and ⟨110⟩ fin orientation. $T_{ox} = 2$ nm, $4.5 \cdot 10^{13}$ cm^{-2} channel implantation.

Figure 21 represents electron mobilities in FinFETs with various fin orientations. An improvement of electron mobility is observed for ⟨100⟩ and an enhancement of hole mobility has also been shown for ⟨110⟩ orientation [20].

4. Comparison of the performance and physical mechanisms in multi-gate devices

Multi-gate MOSFETs realized on thin films are the most promising devices for the ultimate integration of MOS structures due to the volume inversion in the conductive layer [21].

The on-current I_{on} of the MOSFET is limited to a maximum value I_{BL} that is reached in the ballistic transport regime. Figure 22 reports the self-consistent Monte Carlo (MC) simulation of the ballistic ratio $BR = I_{on}/I_{BL}$ versus drain induced barrier lowering (*DIBL*) showing that one can increase the *BR* by scaling the gate length, thus increasing the longitudinal field at the source, but this comes at the cost of a larger *DIBL*. For a given *DIBL*, an increased ballisticity is obtained for low doping double gate SOI devices [22].

The transfer characteristics of several multiple-gate (1, 2, 3 and 4 gates) MOSFETs, calculated using the 3D Schrödinger-Poisson equation and the non-equilibrium Green's function formalism for the ballistic transport or MC simulations, have shown similar trends. The best performance (drain current, subthreshold swing) is outlined for the 4-gate (QG or GAA) structure [23, 24] (Fig. 23).

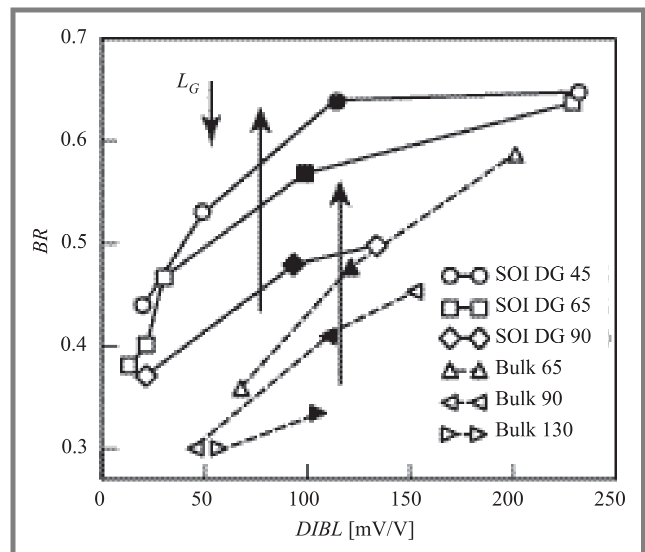


Fig. 22. Ballisticity ratio at $V_G = V_D = V_{DD}$ versus *DIBL*. Filled symbols represent transistors with the nominal gate length for the high-performance MOSFET of each technology node.

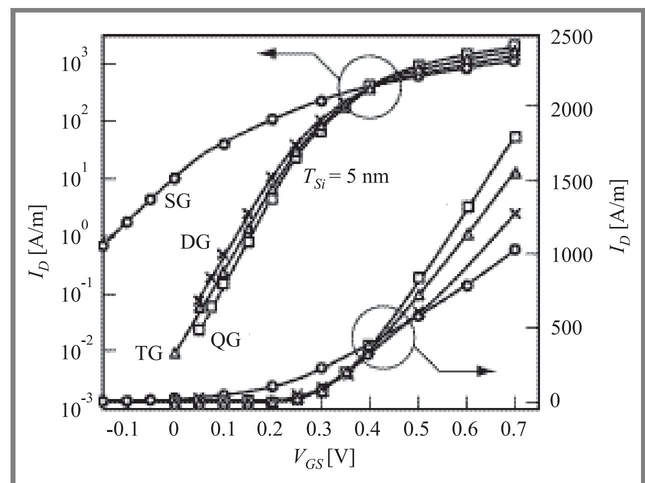


Fig. 23. I_D–V_{GS} characteristics at $V_{DS} = 0.7$ V in thin layers for different multi-gate architectures.

However, Fig. 24 demonstrates that the propagation delay in triple gate (TG) and quadruple gate (QG) MOSFETs are degraded due to a strong rise of the gate capacitance. A properly designed double-gate (DG) structure appears to be the best compromise at given I_{off} [24].

Figure 25 compares the calculated ballistic drive current for Si and Ge double-gate MOSFETs at the operation point of each generation as predicted by International Technology Roadmap for Semiconductors (ITRS) [25]. Si barely satisfies the ITRS requirement, whereas Ge offers much higher current drive. However, the simulated value of the real drain current of 2G SOI transistors is not able to satisfy the ITRS objectives, even for intrinsic devices without

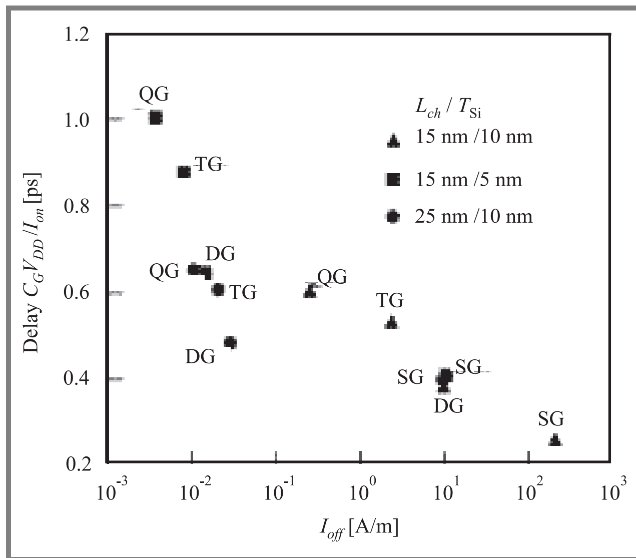


Fig. 24. Propagation delay versus I_{off} for single-gate and multi-gate SOI devices.

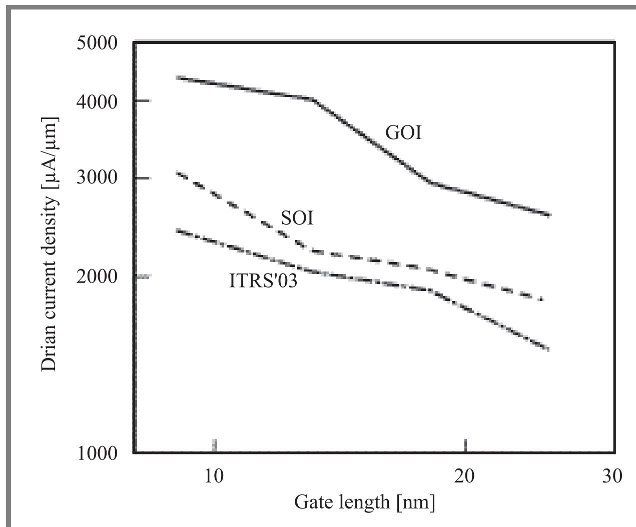


Fig. 25. Ballistic drive current for different technology nodes for SOI and GOI devices.

parasitic S/D resistances. The 2G GOI MOSFETs are able to provide the needed current drive, but parasitic resistances drastically affect the drain current (not shown here).

For a double gate device, the impact of a gate misalignment on the leakage current is important. This current is mainly due to gate induced leakage current (GIDL). This off-current is enhanced with increasing the misalignment and it is higher for a shift of the bottom gate towards the drain due to a higher V_{GD} compared to V_{GS} [26].

The impact of a gate misalignment is also significant for I_{on} in 2G MOSFETs [27]. A large back gate (BG) shift reduces the saturation current compared to the aligned case, whereas a slight BG shift towards the source increases I_{on}. This is due to a lower source access resistance. In terms of short channel effects, aligned transistors exhibit the best control while highly misaligned MOSFETs operate like single gate ones. The off-current I_{off} is much more influenced by the misalignment than I_{on} due to a degradation of the electrostatic control (Fig. 26). The oversized transistor shows attractive static performance (right hand side of Fig. 26) and a better tolerance to misalignment but the dynamic performance is rapidly degraded as the overlap length increases.

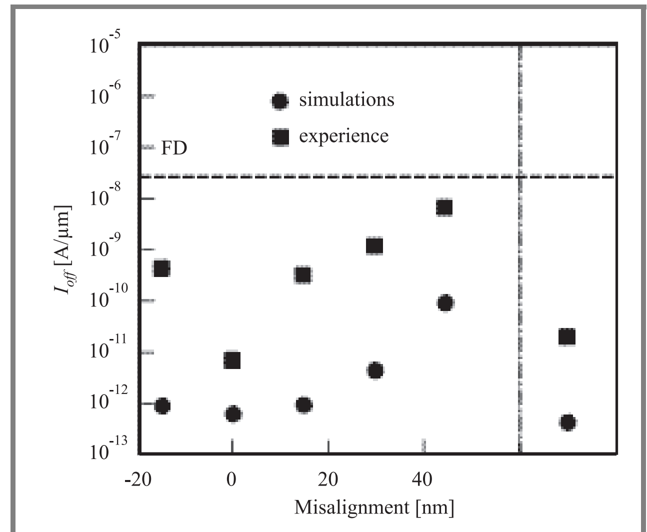


Fig. 26. Evolution of I_{off} with misalignment (experimental and simulations results, $V_D = 1.2$ V). Single gate FD results are represented by a dashed line.

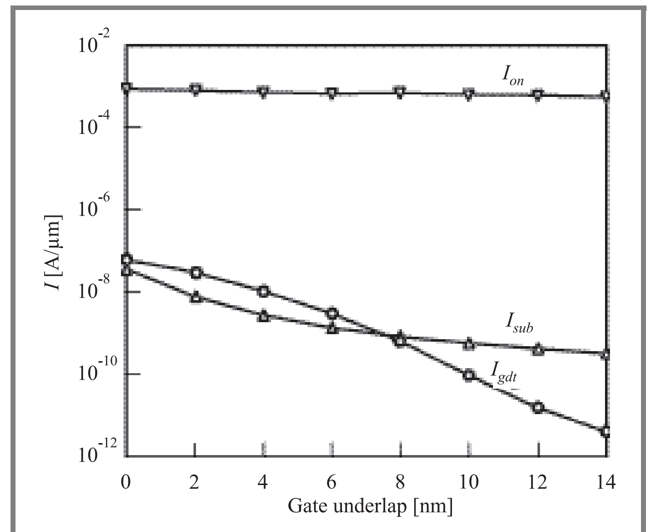


Fig. 27. I_{on}, subthreshold (I_{sub}) and gate direct tunneling (I_{gdt}) currents as a function of gate underlap.

In decananometer MOSFETs, gate underlap is a promising solution in order to reduce the *DIBL* effect. Figure 27 presents the variations of the driving current, the subthreshold current and the gate direct tunneling current versus gate underlap [28]. The on-current is almost not affected by

the gate underlap whereas the leakage currents are substantially reduced due to a decrease in DIBL and drain to gate tunneling current. A reduction of the effective gate capacitance C_G for larger underlap values at iso I_{on} has also been shown. This reduction of C_G leads to a decrease in the propagation delay and power.

Multi-bridge-channel MOSFETs (MBCFET) also present very high performance better than that of gate-all-around (GAA) devices and exceeding the ITRS roadmap requirements (Fig. 28) [29].

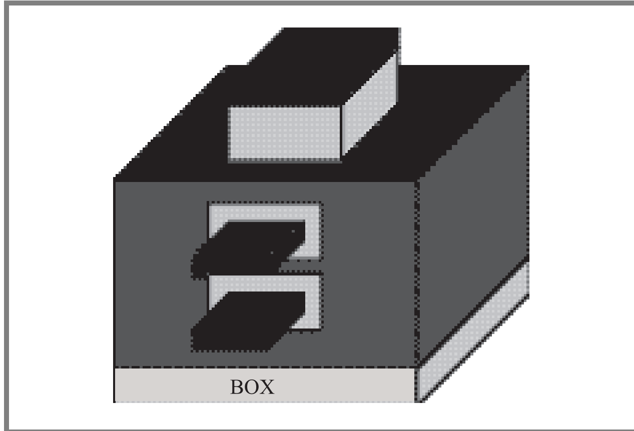


Fig. 28. Schematic diagram of MBCFET on SOI.

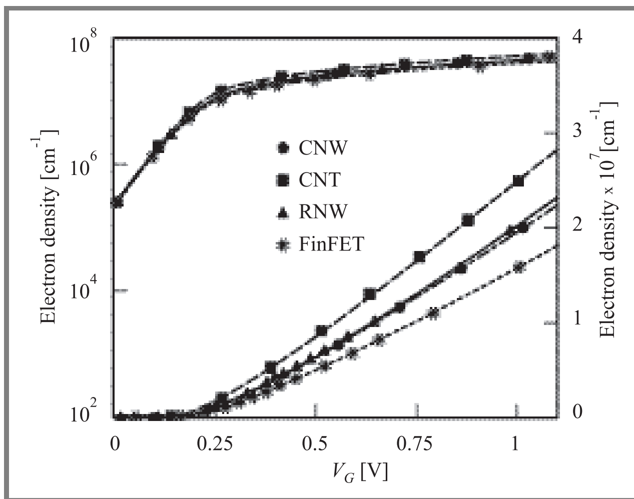


Fig. 29. Electron density per unit length for various devices (FinFET, nanowires and carbon-nanotube FET). 65 nm technology node data (EOT = 0.9 nm, $T_{Si} = 5$ nm).

Finally, FinFETs are compared with cylindrical (CNW) and rectangular (RNW) nanowires and also with gate-all-around carbon nanotubes (CNT) FET. It is shown that the CNTFET exhibits superior performance (Fig. 29) due to electron charge confinement at the surface of the nanotube, whereas in the Si-based nanowires the charge confinement at the center of the wire is responsible for an additional depletion capacitance in series with the oxide capacitance, which reduces the overall effectiveness of the gate [30].

5. Advanced SOI dynamic and non-volatile RAM

It is becoming difficult for memories to be scaled down. Indeed, traditional embedded dynamic random-access memory (DRAM) requires a complicated stack capacitor or a deep trench capacitor in order to obtain a sufficient storage capacitance in smaller cells. This leads to more process steps and thus less process compatibility with logic devices.

Capacitor-less 1T-DRAM or floating body cells have shown promising results. The operation principle is based on excess holes which can be generated either by impact ionization or by gate-induced leakage current in partially-depleted SOI MOSFETs. The GIDL current is due to band-to-band tunneling and occurs in accumulation leading to a low drain current writing and reduced power consumption together with a high speed operation. However, conventional PD SOI MOSFETs require high channel doping to suppress short-channel effects, which induces a degradation in retention characteristics. In order to overcome this problem, a DG-FinDRAM has been proposed showing superior memory characteristics (Fig. 30) [31].

Conventional floating-gate flash memory has also scaling difficulties due to nonscaling of gate-insulator stack and inefficient hot carrier injection processes at sub-50 nm gate

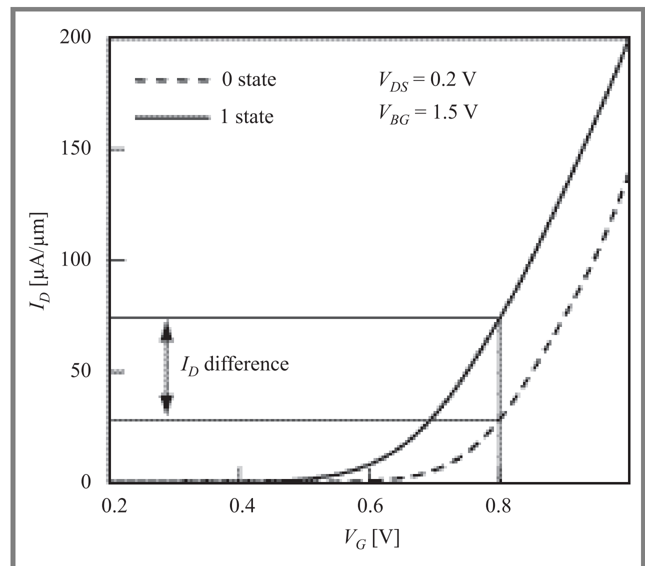


Fig. 30. I_D–V_G characteristics of the DG-FinDRAM.

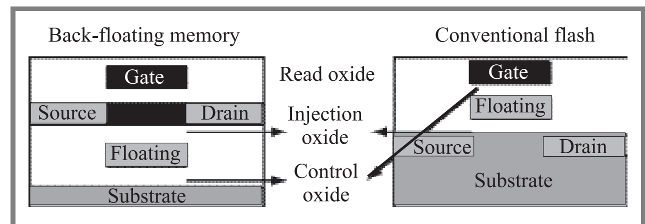


Fig. 31. Cross-sections of back floating gate and conventional front-floating gate memories.

dimensions. Back-floating gate flash memory overcomes these limitations by decoupling read and write operations and independent positioning and/or sizing of the storage element (back-floating gate) under the Si channel (Fig. 31). The charge in the back gate affects the field and the potential at the bottom interface and thus changes the threshold voltage of the device. The back-floating gate is charged by applying –10 V to the source, the drain and the front gate simultaneously, and the charges are removed from the back floating gate (erasing) with the same method but with a bias of +10 V [32].

6. Conclusion

In this paper, a review of recently explored new effects in advanced SOI devices and materials has been given. The impact of key device parameters on electrical and thermal floating body effects has been addressed for various device architectures. Recent advances in the understanding of the sensitivity of electron and hole transport to the tensile or compressive uniaxial and biaxial strains in thin film SOI have been shown. The performance and physical mechanisms have also been presented in multi-gate MOSFETs. New hot carrier phenomena have been discussed. The impact of gate misalignment or underlap, as well as the use of the back gate for charge storage in double-gate nanodevices and of capacitorless DRAM have also been outlined.

Acknowledgements

This work was partially supported by the European Network of Excellence SINANO (Silicon-based Nanodevices, FP6, IST-1-506844-NE).

References

[1] J. P. Colinge, *Silicon-on-Insulator Technology: Materials to VLSI*. Boston: Kluwer, 1991.

[2] S. Cristoloveanu and S. S. Li, *Electrical Characterization of Silicon-on-Insulator Materials and Devices*. Boston: Kluwer, 1995.

[3] F. Balestra, *SOI Devices*. Encyclopedia of Electrical and Electronics Engineering. New York: Wiley, 1999.

[4] F. Dieudonné, S. Haendler, J. Jomaah, and F. Balestra, "Gate-induced floating body effect, low frequency noise and hot carrier reliability in advanced SOI MOSFETs", *Solid-State Electron.*, vol. 48, issue 6, pp. 985–997, 2004.

[5] E. Simoen *et al.*, "Electron valence-band tunneling excess noise in twin-gate silicon-on-insulator MOSFETs", in *Proc. ULIS'2005*, Bologna, Italy, 2005, p. 113.

[6] W. Xiong *et al.*, "Full/partial depletion effects in FinFETs", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 195.

[7] L. Zafari *et al.*, "On the low frequency noise in fully depleted and double-gate SOI transistors", in *Proc. ULIS'2005*, Bologna, Italy, 2005, p. 147.

[8] N. Bresson *et al.*, "Alternative dielectrics for advanced SOI MOSFETs: thermal properties and short channel effects", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 62.

[9] E. Pop *et al.*, "Electro-thermal comparison and performance optimization of thin-body SOI and GOI MOSFETs", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 411.

[10] D. P. Ioannou *et al.*, "New insights on the hot-carrier characteristics of 55 nm PD SOI MOSFETs", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 205.

[11] J. J. Lee *et al.*, "Mobility enhancement of SSOI devices fabricated with sacrificial thin relaxed SiGe", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 139.

[12] M. Sadaka *et al.*, "Fabrication and operation of sub-50 nm strained-Si on $Si_{1-x}Ge_x$ on insulator (SGOI) CMOSFETs", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 209.

[13] T. Numata *et al.*, "Performance enhancement of partially- and fully-depleted strained-SOI MOSFETs and characterization of strained-Si devices parameters", in *Proc. IEDM'04*, San Francisco, USA, 2004, p. 177.

[14] J. Cait *et al.*, "Performance comparison and channel length scaling of strained Si FETs on SiGe-on-insulator (SGOI)", in *Proc. IEDM'04*, San Francisco, USA, 2004, p. 165.

[15] I. Aberg *et al.*, "High electron and hole mobility enhancements in thin-body strained Si/strained SiGe/strained Si heterostructures on insulator", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 173.

[16] K. Uchida *et al.*, "Experimental study of biaxial and uniaxial strain effects on carrier mobility in bulk and ultra-thin-body SOI MOSFETs", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 229.

[17] T. Guillaume *et al.*, "Influence of the mechanical strain induced by a metal gate on electron and hole transport in single and double-gate SOI MOSFETs", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 42.

[18] H. Shang, "Selectively formed high mobility strained Ge PMOSFETs for high performance CMOS", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 157.

[19] T. Low, "Impact of surface roughness on silicon and germanium ultra-thin-body MOSFETs", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 151.

[20] E. Landgraf *et al.*, "Influence of crystal orientation and body doping on trigate transistor performance", in *Proc. ULIS'2005*, Bologna, Italy, 2005, p. 15.

[21] F. Balestra *et al.*, "Double-gate silicon-on-insulator transistor with volume inversion: a new device with greatly enhanced performance", *IEEE Electron Dev. Lett.*, vol. EDL-8, p. 410, 1987.

[22] S. Eminente *et al.*, "Enhanced ballisticity in nano-MOSFETs along the ITRS roadmap: a Monte Carlo study", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 609.

[23] M. Bescond *et al.*, "3D quantum modeling and simulation of multiple-gate nanowire MOSFETs", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 617.

[24] J. Saint-Martin, A. Bournel, and P. Dollfus, "Comparison of multiple-gate MOSFET architectures using Monte Carlo simulation", in *Proc. ULIS'2005*, Bologna, Italy, 2005, p. 61.

[25] A. Khakifirooz, O. M. Nayfeh, and D. A Antoniadis, "Assessing the performance limits of ultra-thin double-gate MOSFETs: silicon vs. germanium", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 79.

[26] C. Yin and P. C. H. Chan, "Characterization and edge direct tunneling leakage of gate misaligned double gate MOSFETs", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 91.

[27] J. Widiez *et al.*, "Experimental gate misalignment analysis on double gate SOI MOSFETs, in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 185.

[28] A. Bansal, B. C. Paul, and K. Roy, "Impact of gate underlap on gate capacitance and gate tunneling current in 16 nm DGMOS devices", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 94.

[29] E.-J. Yoon *et al.*, "Sub-30 nm multi-bridge-channel MOSFET (MBCFET) with metal gate electrode for ultra high performance application", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 627.

[30] A. Marchi *et al.*, "Investigating the performance limits of silicon-nanowire and carbon-nanotube FETs", in *Proc. ULIS'2005*, Bologna, Italy, 2005, p. 99.

[31] T. Tanaka, E. Yoshida, and T. Miyashita, "Scalability study on a capacitorless 1T-DRAM: from single-gate PD-SOI to double-gate FinDRAM", in *Proc. IEDM'2004*, San Francisco, USA, 2004, p. 919.

[32] U. Avci, A. Kumar, and S. Tiwari, "Back-floating gate non-volatile memory", in *Proc. IEEE Int. SOI Conf.*, Charleston, USA, 2004, p. 133.

Francis Balestra was born in Digne, France, in 1960. He received the M.Sc. and Ph.D. degrees in electronics from the Institut National Polytechnique, Grenoble, France, in 1982 and 1985, respectively. In 1989, he obtained the habilitation diploma from the INPG authorizing him to supervise Ph.D. dissertations. He joined the Laboratoire de Physique des Composants á Semiconducteurs (LPCS), INP Grenoble, in 1982, where he has been involved in research on the characterization, modeling, and simulation of the first fully depleted silicon-on-sapphire MOS transistors. He became Chargé de Recherche C.N.R.S. (Centre National de la Recherche Scientifique) in 1985 and Directeur de Recherche CNRS in 2000. In 1993–94, he joined the Research Center for Integrated Systems at Hiroshima University as a visiting researcher, and worked on sub-0.1 μm MOSFETs and thin film SIMOX devices. Between 1996 and 1998 he has been Deputy Director of LPCS, Director of LPCS between 1999 and 2001, and he is now Director of IMEP (Institute of Microelectronics, Electromagnetism and Photonics, INP Grenoble/CNRS/UJF) which is a Laboratory of 140 researchers and technical staff. He led several research teams on deep submicron CMOS, silicon on insulator devices, ultimate Si-based devices realized with innovating architectures, low temperature electronics and advanced bipolar transistors for BiCMOS technology. He has supervised over 15 research projects and 20 Ph.D. He has coordinated European, national and regional projects. He is presently coordinating a European Network of Excellence of the 6th Framework Programme dealing with Silicon-based Nanodevices (SINANO) with 44 European partners, which represents an unprecedented collaboration in this field in Europe. He was the organizer of many international conferences, e.g., the first (1994) and fifth (2002) European Workshop on Low Temperature Electronics (founder of WOLTE), the first (2000) and second (2001) European Conference on Ultimate Integration of Silicon (founder of ULIS), and of a NATO Advanced Research Workshop on SOI in 2000. He is a member of the Advisory Committee of the Chinese Journal of Semiconductors. Doctor Balestra is a member of the European Academy of Sciences since 2003 and got the Blondel Medal (French SEE) in 2001. He has coauthored over 100 publications in international scientific journals, 150 communications at international conferences (20 invited papers and review articles), and 15 books or chapters.

e-mail: balestra@enserg.fr
Institut de Microélectronique, Electromagnétisme
et Photonique
IMEP (CNRS-INPG-UJF)
INP Grenoble-Minatec
38016 Grenoble, France

Invited paper

Special size effects in advanced single-gate and multiple-gate SOI transistors

Akiko Ohata, Romain Ritzenthaler, Olivier Faynot, and Sorin Cristoloveanu

Abstract— State-of-the-art SOI transistors require a very small body. This paper examines the effects of body thinning and thin-gate oxide in SOI MOSFETs on their electrical characteristics. In particular, the influence of film thickness on the interface coupling and carrier mobility is discussed. Due to coupling, the separation between the front and back channels is difficult in ultra-thin SOI MOSFETs. The implementation of the front-gate split C-V method and its limitations for determining the front- and back-channel mobility are described. The mobility in the front channel is smaller than that in the back channel due to additional Coulomb scattering. We also discuss the 3D coupling effects that occur in FinFETs with triple-gate and Ω-gate configurations. In low-doped or tall fins the corner effect is suppressed. Narrow devices are virtually immune to substrate effects due to a strong lateral coupling between the two lateral sides of the gate. Short-channel effects are drastically reduced when the lateral coupling screens the drain influence.

Keywords— MOSFET, SOI, ultra-thin silicon, multiple-gate, mobility, coupling effect, thin gate oxide, gate-induced floating body effect, drain-induced virtual substrate biasing.

1. Introduction

As CMOS is scaled down, the introduction of silicon-on-insulator (SOI) structures is inevitable for improving the short-channel effects, speed, and subthreshold swing. The SOI thickness should be 3–4 times smaller than the channel length and, therefore, stands as a critical parameter for the integration. On the other hand, thin-gate oxides are also necessary for improving the device performance. In this paper, we discuss a number of issues related to the downsizing of planar SOI transistors, focusing on the problems induced by the body thinning and thin gate oxide. A more advanced approach is to use multiple-gate SOI transistors like FinFETs [1] (Fig. 1a), triple-gate FETs [2] (Fig. 1b) or existing variants [3–5] (Fig. 1c). We demonstrate that interesting coupling effects in longitudinal, lateral and vertical directions arise leading to new phenomena. Due to the electrostatic influence of the top and bottom gates, the surface potential and threshold voltage can vary along the lateral gate. Then a bi-dimensional threshold voltage may be observed, depending on the location in the channel. The influence of the back-gate and drain potential, detrimental for single-gate short SOI structures, is also expected to be modified using triple-gate structures because of the three-dimensional coupling.

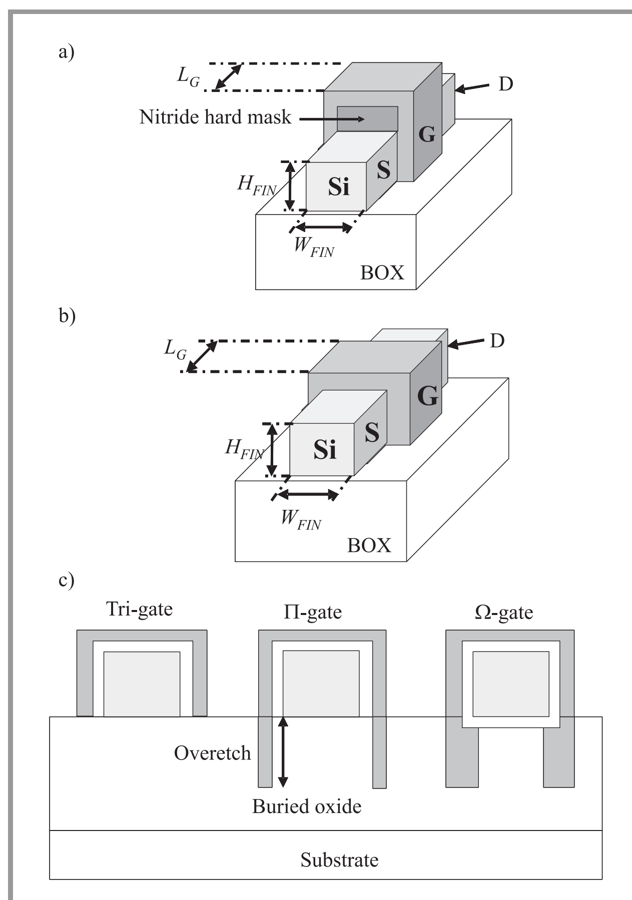


Fig. 1. FinFET (a) and triple-gate FET (b) structures. Variants of triple-gate FETs such as Π- and Ω-gate FETs (c).

2. Ultra-thin SOI MOSFETs

The SOI film thinning improves the electrostatic control and the device scalability, but also causes special coupling and transport effects.

2.1. Coupling effect

An investigation of the coupling effect between the front and back interfaces (i.e., the front-channel threshold voltage V_{T1} as a function of the back-gate voltage V_{G2}, and vice-versa $V_{T2}(V_{G1})$ [6, 7]) can provide useful information for determining whether one or two channels are created in an SOI film. This is an important point for designing double-gate MOSFETs or FinFETs, since their electrical

characteristics are affected by volume inversion [8]. Furthermore, the coupling effect is an essential operation principle for devices in which the threshold voltage is controlled by several gates.

Figures 2a and 2b show the coupling effect between the front and back interfaces in long- and short-channel SOI NMOSFETs. The transconductance curves of a short channel exhibit a single peak (Fig. 2a); however, a plateau appears for a long channel (Fig. 2b) due to the formation of a back inversion layer at the film-BOX interface before the front-channel inversion is created [9]. The back-channel characteristics at various front-gate biases are similar, i.e., the transconductance of a short channel exhibits a single peak [10].

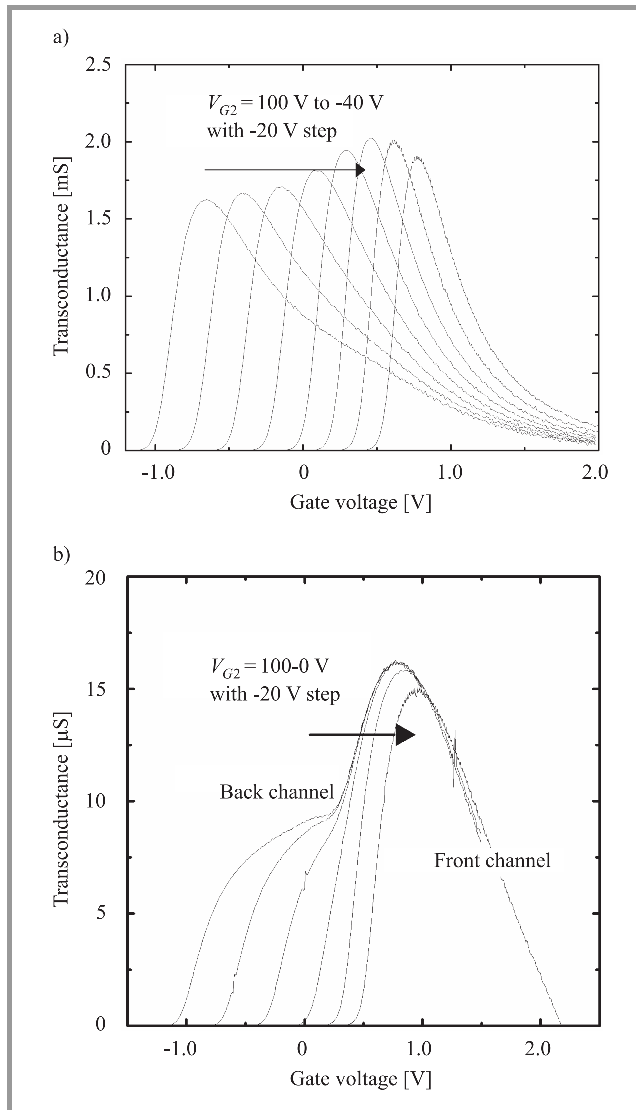


Fig. 2. Transconductance ($\delta I_d/\delta V_{G1}$) in SOI NMOSFET versus gate voltage V_{G1} with the back-gate bias V_{G2} as a parameter for (a) a short SOI MOSFET ($L = 0.5$ μm) and (b) a long SOI MOSFET ($L = 100$ μm). The plateau due to the back-channel activation is evident for a long transistor. In both cases, $T_{Si} = 15$ nm, $T_{BOX} = 400$ nm, $T_{ox} = 2$ nm, $W = 100$ μm, and $V_D = 50$ mV. The doped body ($N_A \approx 4 \cdot 10^{17}$ cm^{-3}) was thinned by sacrificial oxidation, except in the source/drain regions.

Another interesting feature is that the V_{T1} (V_{G2}) curve is superimposed on the V_{G1} (V_{T2}) one for a short channel (Fig. 3a) while the two curves are different for a long channel (Fig. 3b). The coupling curves for a long channel are described by [7]:

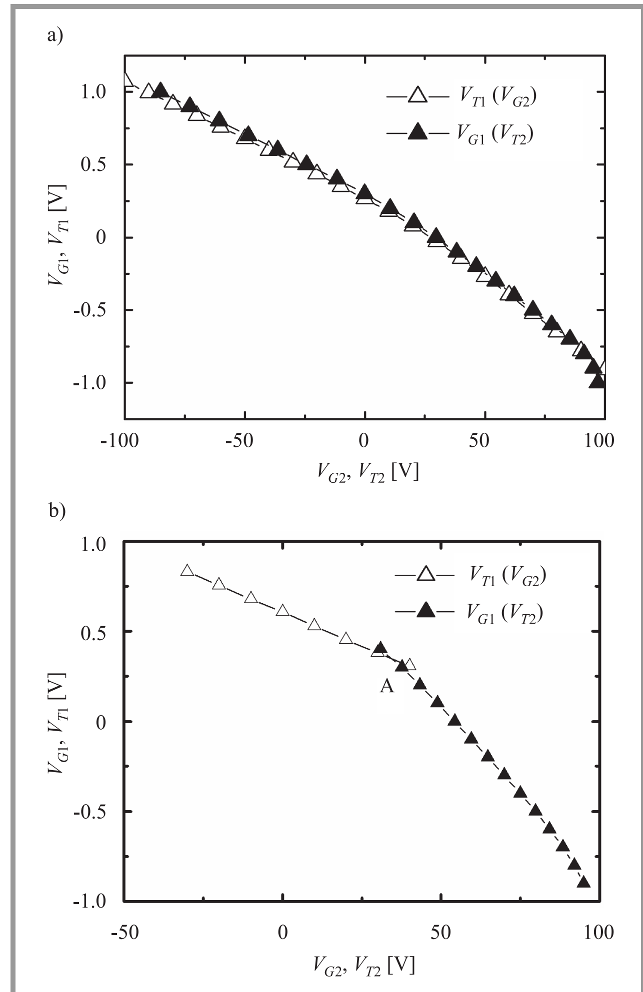


Fig. 3. (a) Front-channel threshold voltage V_{T1} versus back gate bias V_{G2} and back-channel threshold voltage V_{T2} versus front-gate bias V_{G1}. Threshold voltages were extracted from the transconductance peak. In a short channel, the two curves are different. (b) In a long channel, the two curves are different. At point A, both the front and back channels are simultaneously inverted.

$$\Delta V_{T1} \approx -\frac{C_{BOX}}{C_{ox}} \Delta V_{G2}, \quad (1)$$

$$\Delta V_{G1} \approx -\frac{C_{BOX}}{C_{ox}} \frac{C_{si} + C_{ox}}{C_{si}} \Delta V_{T2}, \quad (2)$$

where $C_{si} = \varepsilon_{si}/T_{Si}$, $C_{ox} = \varepsilon_{ox}/T_{ox}$, and $C_{BOX} = \varepsilon_{ox}/T_{BOX}$ represent the capacitances of the depleted film, gate oxide and buried oxide (BOX), respectively.

The cross-point of these curves (point A in Fig. 3b) indicates the voltage condition at which the front and back

channels are simultaneously inverted [11]. If V_{G2} is larger than this value, when sweeping V_{G1}, the back channel is inverted before the front channel. On the other hand, if V_{G2} is smaller than this value, only the front channel is inverted by sweeping V_{G1}. Thus, the inverted channel can clearly be identified from point A.

In the case of an extremely thin SOI layer ($C_{si} >> C_{ox}$), Eq. (2) can be rewritten as:

$$\Delta V_{G1} \approx -\frac{C_{BOX}}{C_{ox}} \Delta V_{T2}\,. \quad (3)$$

Since the slope of this equation is the same as that of Eq. (1), it is difficult to separate the two curves.

Furthermore, for high series resistance once the back channel is inverted, the formation of the front channel could be masked: the transconductance curve obtained by sweeping V_{G1} shows a single-peak structure due to the back channel only [10]. Thus, the results in Fig. 2a and Fig. 3a can be understood as follows: for a short channel, the back-channel plateau transforms into a peak. When measuring $V_{T1}(V_{G2})$, the beginning of the plateau appears as a peak and can be mistaken as V_{T1}. This peak shows the threshold of the back channel; it actually corresponds to $V_{G1}(V_{T2})$ as confirmed by our direct back-channel measurements $V_{T2}(V_{G1})$. The point where the slope of the curve changes (around $V_{G2} = 20$ V in Fig. 3a) corresponds to the voltage condition at which the front and back channels are simultaneously inverted.

As the SOI thickness or channel area decreases, the point where the slope of the coupling curve changes is useful to determine which channel is actually inverted; this identification is impossible from the transconductance curves due to their single-peak shape [12]. Figure 4a reproduces the transconductance curves exhibiting a single peak at various V_{G2} applied to a short-channel SOI NMOSFET. Figure 4b shows that the threshold voltage is due to the front channel when V_{G2} is smaller than point B, and to the back channel when V_{G2} is larger than point B. Note that V_{G2} at point B is less than 0 V. Hence, the transconductance peak at $V_{G2} = 0$ V occurs due to the back channel. Transconductance and subthreshold slope deteriorate because the channel is controlled through the gate oxide and silicon body. Thus, this device is not suitable for operation at $V_{G2} = 0$ V. On the other hand, the transconductance peak reaches a maximum value near the condition where both channels are simultaneously inverted ($V_{G2} \approx -5$ V in Fig. 4b).

As the SOI thickness decreases below 10 nm, the coupling effect theory should be revisited [13]. Figure 5 shows the simulated results of the surface potential variation with the front-gate voltage for a negative back-gate bias. In a thick-body MOSFET, as V_{G1} increases, the front channel is gradually inverted whereas the back-surface potential does not change, remaining in accumulation. By contrast,

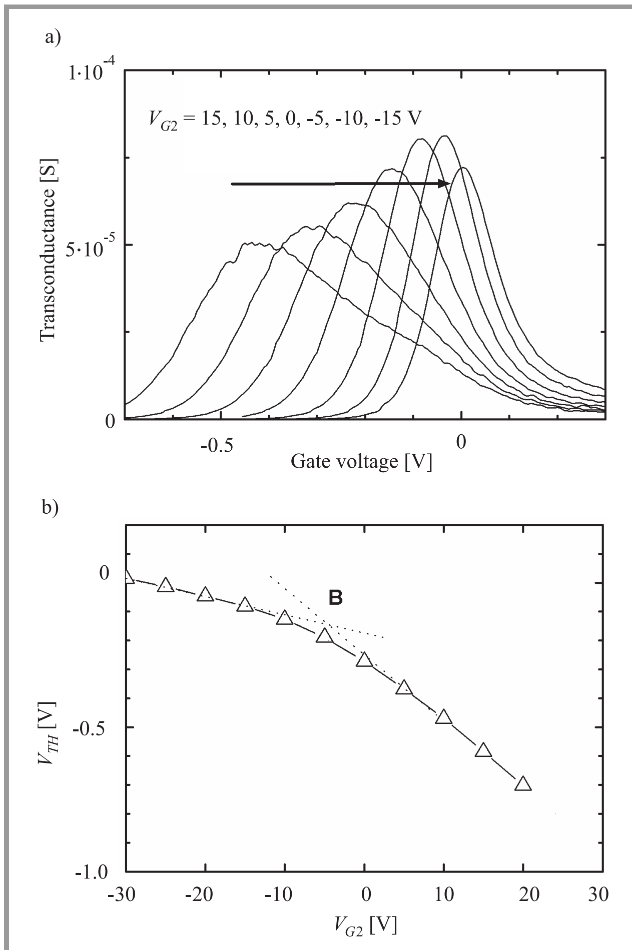


Fig. 4. (a) Transconductance ($\delta I_D/\delta V_{G1}$) versus front-gate voltage with the back-gate bias as a parameter for the undoped thin-SOI (initial wafer doping was $N_A = 5 \cdot 10^{14}$ cm^{-3}) NMOSFET. $V_D = 10$ mV, $W = 10$ μm, $L = 0.05$ μm. The average SOI thickness was 9 nm on the wafer. The gate oxide thickness is approximately 1 nm. (b) Threshold voltage V_{th} versus back gate voltage V_{G2}, $V_D = 10$ mV. At point B where the slope changes, both sides are inverted simultaneously.

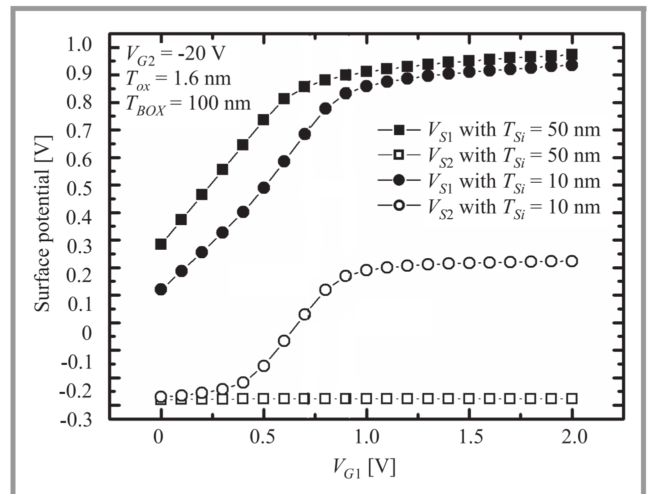


Fig. 5. Surface potential of the front (V_{S1}) and back (V_{S2}) interfaces versus front-gate voltage (V_{G1}) in the case of thick (50 nm) and thin (10 nm) SOI. Back gate voltage $V_{G2} = -20$ V.

in an ultra-thin body, both the front and back potentials increase with V_{G1}. This means that it becomes difficult to keep one channel in accumulation while inverting the opposite channel. Additional simulations show that the potential distribution in a 5 nm thick body is quasi-flat. The two channels merge in a single channel that covers the whole film volume [13]. Ultra-thin SOI devices are also subject to quantum effects: sub-band splitting which leads to an increase in the threshold voltage and a modification of the effective mass and scattering mechanisms.

2.2. Carrier mobility

The effect of the film thickness on mobility is important for further CMOS performance and scalability. Figure 6 shows the mobility-thickness correlation in 10–20 nm thick NMOSFETs with a 2-nm SiO_2 gate oxide, in which SOI thickness variations were locally caused by process fluctuations. In short-channel MOSFETs, the mobility decreases for thinner films, whereas in long MOSFETs the thickness effect is insignificant. This difference is essentially due to the series resistance. This argument applies to both the front and back channels [14].

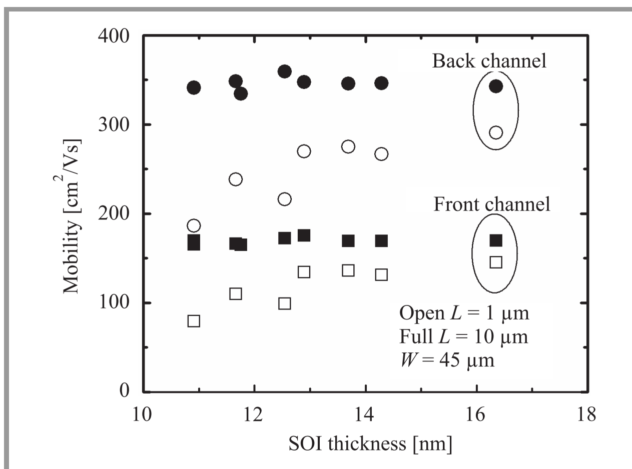


Fig. 6. Front- and back-channel field-effect mobilities versus SOI film thickness for short (1 μm) and long (10 μm) channels. The field-effect mobility was determined from the transconductance peak. All the devices were probed on the same wafer. The local film thickness was measured by ellipsometry and verified by analyzing the coupling between the front and back channels. $T_{ox} = 2$ nm and $T_{BOX} = 400$ nm. The doped body ($N_A \approx 4 \cdot 10^{17}$ cm^{-3}) was thinned by sacrificial oxidation, except in the source/drain regions.

The mobility is systematically lower at the front channel than at the back channel, even after correcting for a lowered gate-oxide capacitance due to polysilicon depletion and quantum capacitance of the inversion layer. Comparing the effective mobility at the front and back channels makes the origin of the mobility difference clear.

The front-gate split C-V method was used for determining the carrier mobility and density, N_s, at both the front and back channels. As shown in Fig. 7, a plateau is observed in

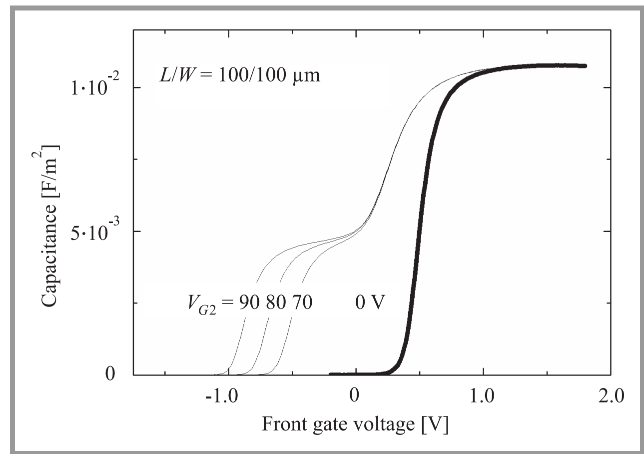


Fig. 7. Front-gate split C-V measurements for various substrate voltages. $T_{Si} \approx 16$ nm. The frequency for split C-V measurement is 5 kHz.

the C-V plot at large substrate voltages due to the creation of the back channel [15]. The carrier density N_s of the back channel can be obtained by integrating the capacitance values specific to this region [16, 17]:

$$qN_{s\,Back}\,(V_{G1}, V_{G2}) = \int_{-\infty}^{V_{G1}} C_{gc}\, \mathrm{d}V_{G1}\,. \tag{4}$$

The effective field is determined from the equations for thin SOI [18, 19]. The threshold voltage is calculated by defining a corresponding inversion charge density [17].

There is, however, a limitation in this front-gate split C-V method for evaluating the carrier density in the back channel [17]. Figure 8 shows the relationship between the back-channel threshold voltage V_{T2} and front-gate volt-

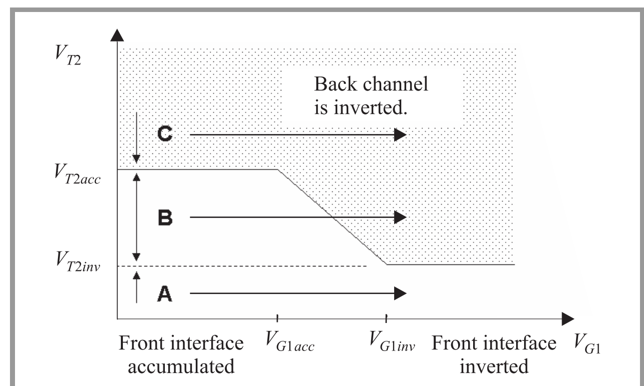


Fig. 8. The relationship between the threshold voltage of the back channel (V_{T2}) and the front-gate voltage (V_{G1}) by the model of Lim and Fossum. When the front interface is accumulated (V_{T1acc}), the threshold voltage of the back channel becomes constant. When the back gate is in region A and the front gate is swept, only the front channel is inverted. When the back gate is in region B and the front gate is swept, the back channel is turned off and on by the front gate voltage. When the back gate is in region C and the front gate is swept, the back channel cannot be turned off by the front gate voltage because of the accumulation layer at the front interface.

age V_{G1} as modeled by Lim and Fossum [7]. For the back gate bias in region A, only the front channel is activated by applying the front-gate voltage. Thus, the carrier density at the front channel can be evaluated by the front-gate split C-V method. For the back gate bias in region B, the back channel can be turned on and off by sweeping the front-gate voltage. Thus, the carrier density at the *back* channel can be evaluated by the front-gate split C-V method. On the other hand, for the back gate bias in region C, when V_{G1} is swept, it is impossible to cut off the back channel since V_{T2} becomes constant when the front interface is accumulated. Furthermore, in this region C, a careful examination is required for getting the correct C-V curves. The front-interface accumulation layer is not formed instantly as there is no source for the majority carrier in SOI films. A hysteresis in C-V and I-V is therefore observed [15, 20]. Thus, the front-gate split C-V method cannot be used for determining the carrier density at the back channel in region C.

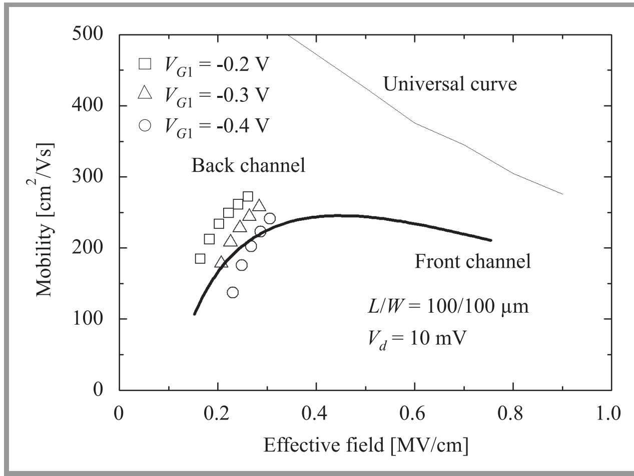


Fig. 9. Front-channel effective mobility (for $V_{G2} = 0$ V) and back-channel effective mobility (for $V_{G1} = -0.2, -0.3$, and -0.4 V) versus effective field.

Figure 9 shows the effective mobility of the front ($\mu_{eff,F}$) and back ($\mu_{eff,B}$) channels evaluated by $I_D - V_G$ and front-gate split C-V method for our devices of Fig. 6. For low $|V_{G1}|$, $\mu_{eff,B}$ is clearly higher than $\mu_{eff,F}$ at the same effective field E_{eff}. The mobility components μ_{Fl} and μ_{Bl} which indicate the deviation from the "universal mobility" curve were estimated by Matthiessen's rule:

$$\mu_{Fl,Bl}^{-1}(N_s) = \mu_{eff\,F,eff\,B}^{-1}\big(N_s(E_{eff})\big) - \mu_{universal}^{-1}(E_{eff})\,. \quad (5)$$

This procedure removes the impact of phonon scattering. Figure 10 shows that these mobility components depend on N_s, which suggests the role of Coulomb scattering. V_{G1} dependence of $\mu_{eff,B}$ in Fig. 9 is explained by the change in N_s. At a constant E_{eff}, N_s decreases with a more negative V_{G1}, which results in a reduced charge-screening effect.

Furthermore, since $\mu_{Fl} < \mu_{Bl}$, an additional scattering mechanism would affect only the front channel. We extract μ_{add} on the basis of the mobility difference between the two channels as follows:

$$\mu_{add}^{-1}(N_s) = \mu_{Fl}^{-1}(N_s) - \mu_{Bl}^{-1}(N_s)\,. \quad (6)$$

Note that the mobility limited by Coulomb scattering on film impurities is now eliminated. The dependence of μ_{add} on N_s (Fig. 10) suggests that the presence of Coulomb scat-

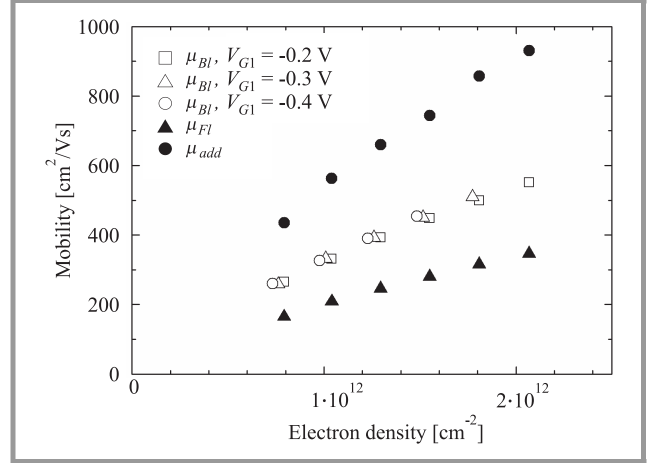


Fig. 10. Mobility-lowering components of the front channel μ_{Fl} and back channel μ_{Bl} determined from Eq. (5). The additional mobility-lowering component μ_{add} of the front channel was derived from Eq. (6).

tering centers at the front channel is responsible for the mobility difference between the two channels. Remote scattering centers from the polysilicon gate are possible candidates and should be considered for thin-gate oxides [21].

2.3. Thin gate oxide

An important effect in SOI MOSFETs with ultra-thin oxides is the gate-induced floating-body effect (GIFBE) [22]. This effect is caused by the tunneling current from the valence band of the SOI film into the gate [23], which charges the body with majority carriers. The body potential is determined by the balance between the body charging via direct tunneling current and the carrier recombination in the body or junctions. Due to the GIFBE, the transconductance curve becomes distorted giving rise to a second peak, as shown in Fig. 11.

The GIFBE occurs even in FD MOSFETs, especially when the back interface is biased close to accumulation [24]. The standard method to determine the mobility in FD SOI MOSFETs consists of eliminating the interface coupling by accumulating the opposite channel. Figure 11 shows that the second peak of transconductance due to the GIFBE becomes larger than the genuine peak. If the "maximum" transconductance is misused, then the field-effect mobility is overestimated.

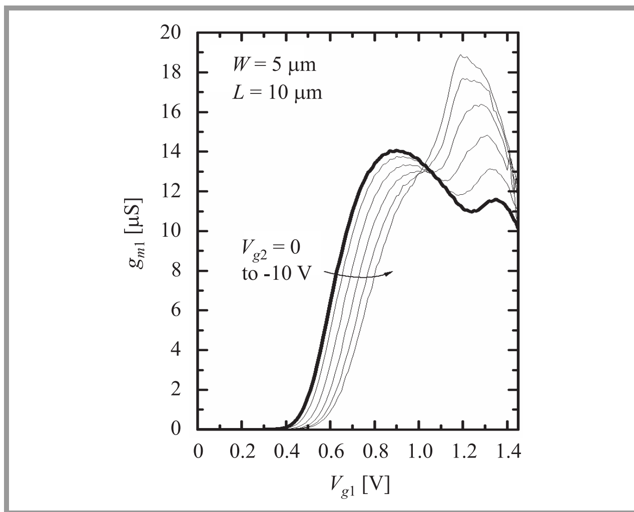


Fig. 11. Transconductance as a function of the front gate voltage for different back-gate biases from 0 V to –10 V in a FD SOI MOSFET. $T_{BOX} = 145$ nm, $T_{ox} = 1.6$ nm, and $T_{Si} = 17$ nm. The "second peak" appears due to the GIFBE.

3. Non-planar multiple-gate SOI transistors

3.1. Corner effects

Triple-gate devices feature a non planar silicon/oxide interface involving corners (Fig. 1b). In each corner region (Fig. 12a), the coupling between the adjacent gates induces a local lowering of the threshold voltage. This parasitic channel may cause an increase of the off-state leakage current. In order to characterize the threshold voltage inhomogeneity, we performed numerical simulations. The "local" threshold voltage of the different channel regions was extracted from the electrical characteristics. First, the "global" threshold voltage V_{TG} was determined from the second derivative of the drain current versus gate voltage curve. Then, the electron density for $V_G = V_{TG}$ was calculated all along the channel. Finally, using the relationship between the threshold voltage and the inversion charge, a "local" threshold voltage was calculated:

$$V_T^{LOCAL} = V_{TG} - \frac{Q_{INV}}{C_{OX}}. \qquad (7)$$

The local threshold voltage along the vertical channel for undoped FinFET and triple-gate (TriGate) transistors is plotted in Fig. 12b. The threshold of a triple-gate structure is found to decrease near the top gate. For FinFETs, the deactivation of the top channel by the nitride hard mask suppresses the drop of the threshold voltage and the profile is roughly flat along the channel.

The ΔV_T is defined as the maximum threshold lowering along the vertical axis. In Fig. 13, ΔV_T is plotted as a function of the fin height for FinFETs and triple-gate FETs, with low and high doping. As expected, tall fins

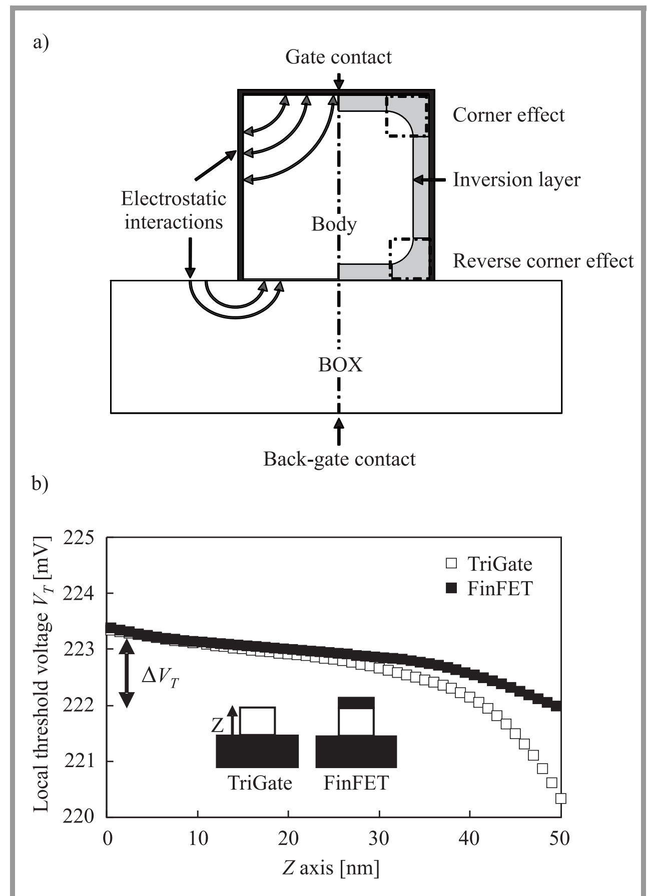


Fig. 12. (a) Cross-section of a triple-gate structure showing the electrostatic interactions leading to corner effects. (b) Local threshold voltage along the vertical channel for low doped ($N_A = 10^{15}$ cm^{-3}) FinFET and triple-gate FET with $W_{FIN} =$ 50 nm, $H_{FIN} = 50$ nm, and $L_G = 0.5$ μm.

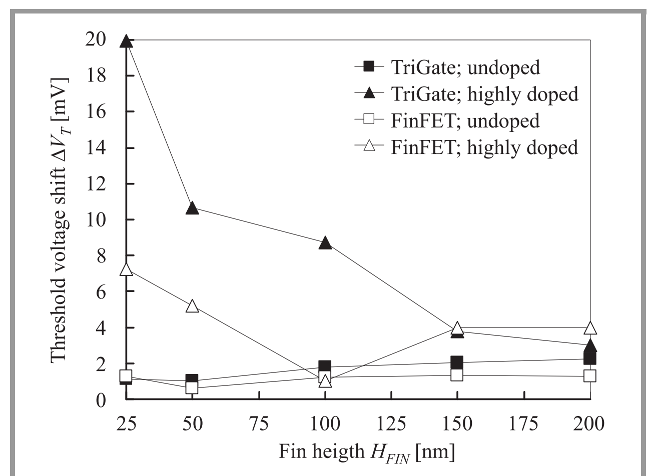


Fig. 13. ΔV_T versus fin height H_{FIN} for FinFET and triple-gate structures with low ($N_A = 10^{15}$ cm^{-3}) and high ($N_A = 10^{18}$ cm^{-3}) body doping. $W_{FIN} = 50$ nm, $L_G = 500$ nm.

exhibit a lower ΔV_T because of a smaller top-gate influence. In FinFETs the threshold voltage shift is efficiently reduced due to their quasi-double gate operation. A low

doping clearly attenuates the threshold voltage shift for both FinFET and triple-gate architectures. Although the threshold voltage non-uniformity increases with the doping level, ΔV_T shift due to the non-planar triple-gate configuration is smaller than 25 mV, indicating a weak impact of the corners.

3.2. Lateral versus vertical coupling

The ΩFETs are triple-gate transistors where the lateral gates penetrate within the BOX (Fig. 1c) [5]. A wide ΩFET behaves as a single-gate fully depleted (FD SOI) device rather than a FinFET. Changing the back-gate bias V_{G2} from -15 to $+15$ V results in a lateral shift of the current and transconductance characteristics (Fig. 14a). The back-channel activation is visible on transconductance curves as a plateau (for $V_{G2} = +15$ V and $V_G = -0.5$ V, see also Fig. 2b).

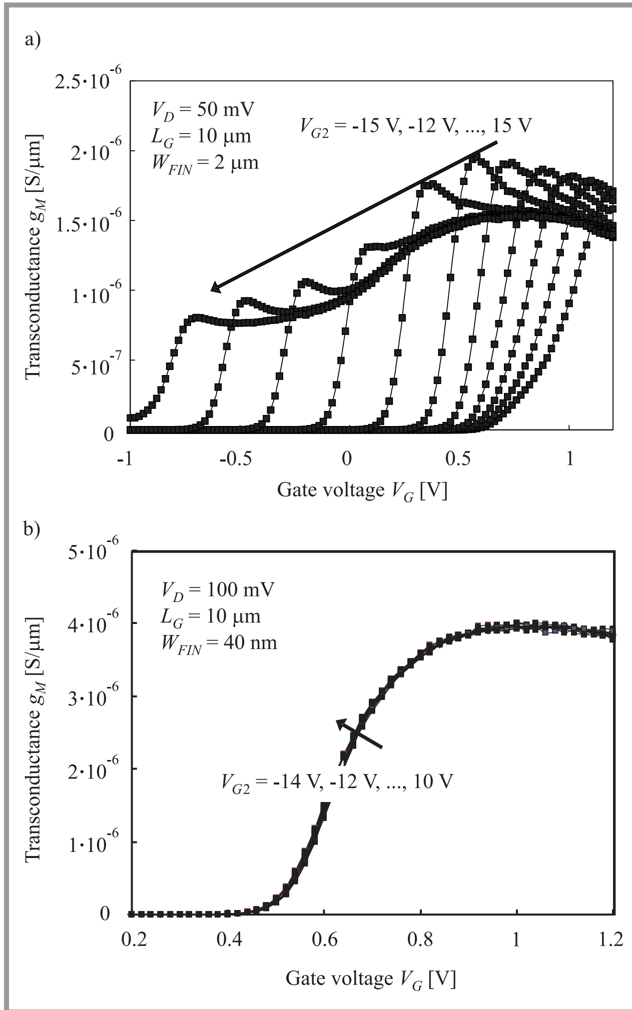


Fig. 14. (a) Transconductance g_M versus gate voltage V_G for a wide ($W_{FIN} = 2\ \mu$m) ΩFET. The back-gate voltage V_{G2} varies from -15 to 15 V. (b) Transconductance g_M versus gate voltage V_G for a narrow ($W_{FIN} = 40$ nm) ΩFET. The back-gate voltage V_{G2} varies from -14 to 10 V.

For narrow devices (Fig. 14b), the substrate bias is no longer effective. The silicon-BOX interface becomes controlled by the fringing electric field penetrating from the lateral gates into the channel (horizontal coupling) and into the BOX. For instance, when the front gate is in accumulation, it is impossible to invert the back channel even for high back-gate bias.

The threshold voltage variation with substrate bias is shown in Fig. 15. For *wide* devices, the change from substrate ac-

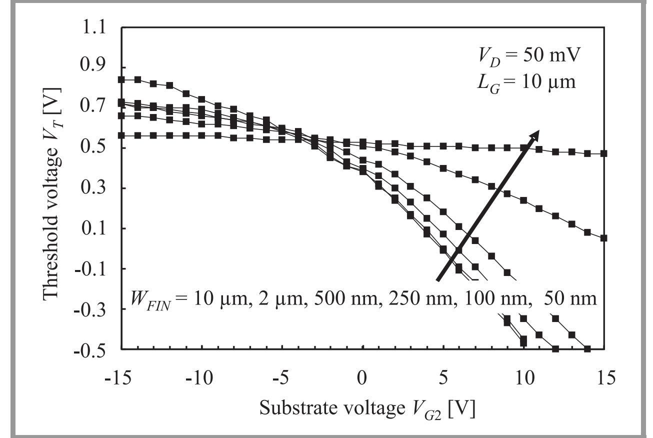


Fig. 15. Threshold voltage V_T versus back-gate bias V_{G2} for n-channel ΩFETs with various fin widths. V_T was extracted using a constant current method.

cumulation to substrate inversion is reflected by a linear decrease of the threshold voltage, followed by a more rapid drop due to the activation of the back channel (1D Lim and Fossum model, see also Fig. 3 [7]). As noted above, the lateral gates can screen the narrow silicon body from the back-gate influence. In Fig. 15, it is seen that for devices with very small fin width the threshold voltage is rather insensitive to back-gate voltage. In ΩFET devices, the control of the body and back channel by the main gate is better than in FinFETs.

The potential distribution in the channel of a triple-gate structure can be analytically calculated by adapting the model proposed for four-gate FETs [25]. The variation of the threshold voltage when the back-gate is biased into depletion (vertical coupling: Eq. (1)) can then be rewritten to take into account the influence of the lateral coupling:

$$\alpha(W,H) = \frac{\partial V_{T1}}{\partial V_{G2}} = \left[\frac{2\sqrt{2}}{\sinh\left(2\sqrt{2}\frac{H}{W}\right)}\right] \frac{\frac{C_W(W)}{C_{ox}}}{1+\frac{2\sqrt{2}}{\tanh\left(2\sqrt{2}\frac{H}{W}\right)}\frac{C_W(W)}{C_{BOX}}}, \tag{8}$$

where $C_W = \varepsilon_{Si}/W_{FIN}$, C_{ox} and C_{BOX} are respectively the lateral silicon film capacitance, front and back oxide capacitances.

In Fig. 16, the 2D coupling coefficient α, which generalizes Fossum's model [7], is plotted as a function of the channel width. For a narrow triple-gate the coefficient is close to zero; this corresponds to the situation where the lateral gates control perfectly the body, screening the influence of the back-gate. For a fully depleted SOI configuration, i.e., $W >> H$, the coupling coefficient α is close to the classical value T_{ox}/T_{BOX}, given in Eq. (1).

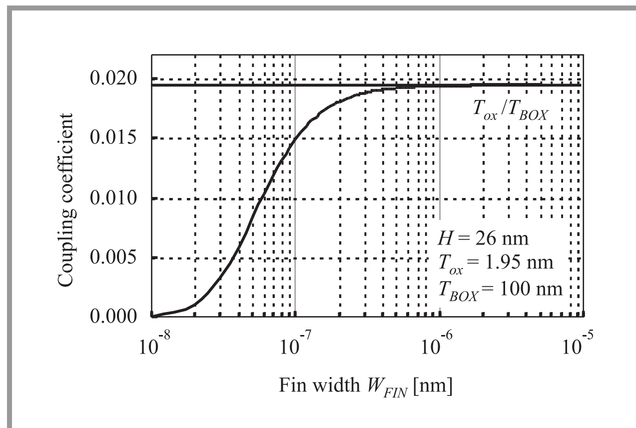


Fig. 16. Coupling coefficient $\Delta V_{T1}/\Delta V_{G2}$ (back-gate biased into depletion) as a function of the fin width in a triple-gate transistor.

Practical implications of back-gate coupling in multiple-gate FETs are related to resistance to harsh environment with radiation or hot carrier injection, where charges trapped into the BOX can modify the electrostatic potential of the back interface. Radiation tests on ΩFET devices [26] have shown that wide devices exhibit a 200 mV threshold voltage shift after 500 krad exposure. By contrast, narrow devices (50 nm) experience a quasi-null threshold voltage variation. Narrow ΩFETs are therefore intrinsically immune to radiation effects and could operate in very harsh conditions.

3.3. Lateral versus longitudinal coupling – DIVSB

A dramatic short-channel effect in fully depleted SOI structures is the penetration of the electric field from the drain into the buried oxide and underlying silicon substrate (Fig. 17a). This fringing field tends to increase the surface potential at the back interface (film-BOX). Because of the existing vertical coupling between the front and back interfaces, the properties of the front channel are degraded.

The potential profile in the buried oxide can be resolved by conformal mapping [27]. Deriving the potential at the body/BOX interface, the expressions of the back-channel transverse field components can be calculated. The source/body $C_{BS}(x)$ and drain/body $C_{BD}(x)$ capacitances

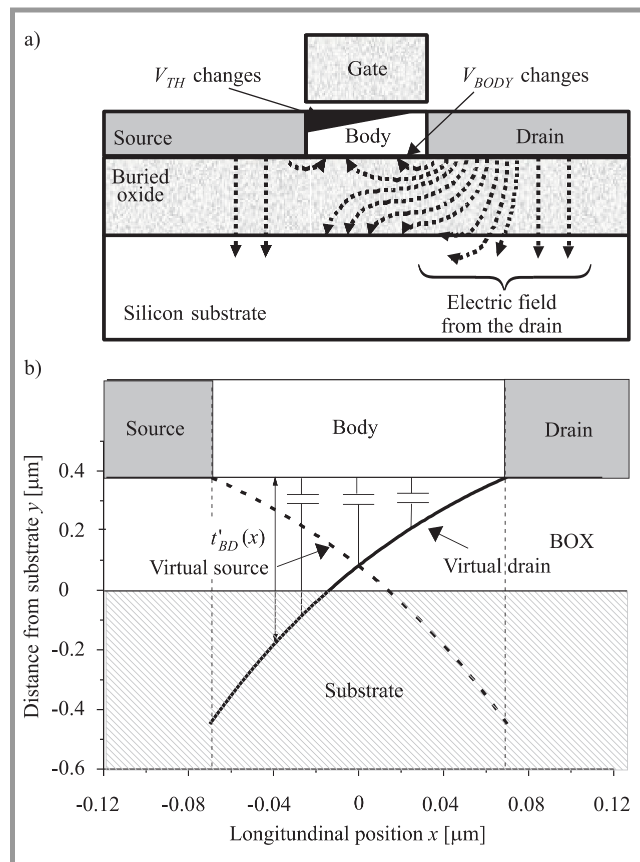


Fig. 17. (a) Schematics of the fringing fields leading to DIVSB effect and (b) equivalent representation of buried oxide fringing capacitances and source/drain virtual electrodes.

summarize the back-channel charge control by the source and drain:

$$C_{BS/BD}(x) = \frac{\varepsilon_{ox}}{T_{BOX}} \; \frac{1}{\exp\left[\dfrac{\pm\pi}{T_{BOX}}\left(x \pm \dfrac{L}{2}\right)\right] - 1}. \qquad (9)$$

A very interesting feature of this formalism is the fact that the fringe capacitances C_{BD} and C_{BS} behave like the BOX capacitance C_{BOX}. This means that drain and source influence can be described by *virtual* electrodes located at a varying distance from the body/BOX interface (Fig. 17b). This is why the phenomenon is named DIVSB (drain induced virtual substrate biasing) [27].

The 3D simulations were carried out to investigate the evolution of DIVSB effect as a function of fin width. The vertical potential profiles taken in the centre of the device (at $L_G/2$ and $W_{FIN}/2$) are presented in Fig. 18. At low drain voltage (10 mV, Fig. 18a) the potential depends exclusively on the fin width. For narrow ΩFETs, the BOX region located between the two lateral gates is totally under control and the potential exhibits a peak. For wide devices, the lateral gates have no influence. For high drain voltage (1.2 V, Fig. 18b) DIVSB effect is directly visible in a short and wide device, where the potential in the BOX is increasing as compared to a long and wide device. This

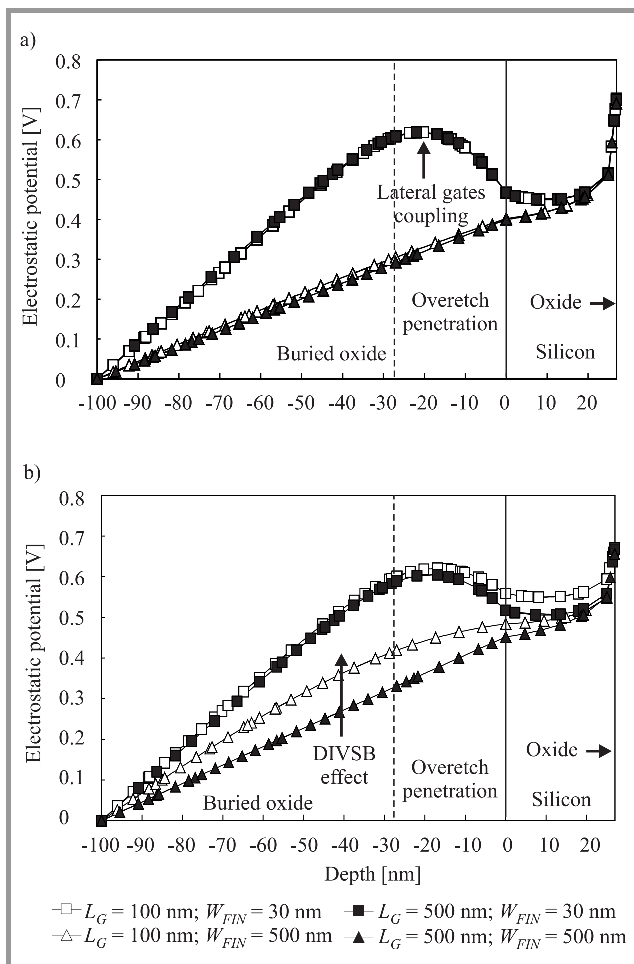


Fig. 18. Vertical potential profiles in the centre of the fin. Simulations are for short (open symbols) and long (closed symbols), as well as for narrow (squares) and wide (triangles) devices; (a) is simulated with a low drain voltage ($V_D = 10$ mV) and (b) with a high drain voltage ($V_D = 1.2$ V); $V_G - V_T = 300$ mV, and $V_{G2} = 0$ V.

critical problem is solved for narrow fins, where the potential profile in the BOX is roughly the same for long and short devices.

4. Summary

Several dimensional effects taking place in modern SOI MOSFETs have been reviewed. In planar transistors, the film thickness plays a crucial role. The coupling effects are amplified in ultra-thin films preventing the separation between the front and back channels. The coexistence of accumulation in the back channel with inversion in the front channel becomes equally difficult. In addition, the high series resistance caused by film thinning can mask the activation of one channel. The method for determining the inverted channel was described.

The carrier mobility in ultra-thin SOI MOSFETs can be extracted by adapting the front-gate split C-V method to the back channel. It was found that the mobility in the front channel is smaller than that in the back channel due to Coulomb scattering. The possibility of misevaluation of the carrier mobility, due to direct tunneling from the valence band through very thin gate oxides and related floating-body effects, was pointed out.

In non-planar FinFET-like transistors, additional size and coupling effects are induced by the multiple gates. We showed that the corner effect, i.e., the influence of a gate on the threshold voltage of an adjacent gate, is very small when using low-doped bodies or tall fins. The back-gate influence, which can modify the threshold voltage of the device and activate a back-channel, is screened for narrow fins by the strong coupling between the lateral gates. The drain-to-BOX coupling and the radiation effects are also reduced in narrow devices by the interaction of the lateral gates.

This paper addressed a number of SOI mechanisms that require further modeling, characterization and control. However, the use of SOI technology is not debatable. SOI is the necessary solution for extending the miniaturization of CMOS circuits.

Acknowledgements

We would like to thank our colleagues who have contributed to this work: J. Pretet, S. Eminente, M. Cassé, A. Vandooren, X. Mescot, F. Daugé, T. Poiroux, K. Akarvardar, T. Ernst, and A. Zaslavsky.

References

[1] D. Hisamoto *et al.*, "A folded-channel MOSFET for deep-sub-tenth micron era", in *IEDM'98 Tech. Dig.*, San Francisco, USA, 1998, pp. 1032–1034.

[2] B. Doyle *et al.*, "Tri-gate fully-depleted CMOS transistors: fabrication, design and layout", in *VLSI 2003 Tech. Dig.*, Washington, USA, 2003, pp. 132–133.

[3] J.-T. Park, J.-P. Colinge, and C. H. Diaz, "Pi-gate SOI MOSFET", *IEEE Electron Dev. Lett.*, vol. 22, no. 5, pp. 405–406, 2001.

[4] F.-L. Yang *et al.*, "25 nm CMOS omega FETs", in *IEDM'02 Tech. Dig.*, San Francisco, USA, 2002, pp. 255–258.

[5] C. Jahan *et al.*, "10 nm ΩFETs transistors with TiN metal gate and HfO_2", in *VLSI 2005 Tech. Dig.*, Kyoto, Japan, 2005, pp. 112–113.

[6] S. Cristoloveanu and S. S. Li, *Electrical Characterization of SOI Materials and Devices*. Boston: Kluwer, 1995.

[7] H. K. Lim and J. G. Fossum, "Threshold voltage of thin-film silicon-on-insulator (SOI) MOSFET's", *IEEE Trans. Electron Dev.*, vol. 30, no. 10, pp. 1244–1251, 1983.

[8] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-gate silicon on insulator transistor with volume inversion: a new device with greatly enhanced performance", *IEEE Electron Dev. Lett.*, vol. 8, no. 9, pp. 410–412, 1987.

[9] T. Ouisse, S. Cristoloveanu, and G. Borel, "Influence of series resistances and interface coupling on the transconductance of fully-depleted silicon-on-insulator MOSFETs", *Solid-State Electron.*, vol. 35, no. 2, pp. 142–149, 1992.

[10] A. Ohata, S. Cristoloveanu, M. Cassé, A. Vandooren, and F. Daugé, "Coupling effect between the front and back interfaces in thin SOI MOSFETs", *Microelectron. Eng.*, vol. 80, no. 17, pp. 245–248, 2005.

[11] A. Ohata, J. Pretet, S. Cristoloveanu, and A. Zaslavsky, "Correct biasing rules for virtual DG mode operation in SOI-MOSFETs", *IEEE Trans. Electron Dev.*, vol. 52, no. 1, pp. 124–125, 2005.

[12] A. Ohata, S. Cristoloveanu, M. Cassé, A. Vandooren, and F. Daugé, "Characterization of ultra-thin SOI MOSFETs by coupling effect between front and back interfaces", in *IEEE Int. SOI Conf.*, Hawaii, USA, 2005, pp. 63–64.

[13] S. Eminente, S. Cristoloveanu, R. Clerc, A. Ohata, and G. Ghibaudo, "Ultra-thin fully-depleted SOI MOSFETs: special charge properties and coupling effect", *Solid-State Electron.*, vol. 51, no. 2, pp. 239–244, 2007.

[14] A. Ohata, S. Cristoloveanu, and M. Cassé, "Mobility comparison between front and back channels by front-gate split capacitance-voltage method", *Appl. Phys. Lett.*, vol. 89, no. 3, p. 032104, 2006.

[15] J. Chen, R. Solomon, T. Y. Chan, K. P. Ko, and C. Hu, "Threshold voltage and CV characteristics of SiO_2 MOSFET's related to Si film thickness variations on SIMOX wafers", *IEEE Trans. Electron Dev.*, vol. 39, no. 10, pp. 2346–2353, 1992.

[16] D. Esseni, M. Mastrapasqua, G. K. Celler, C. Fiegna, L. Selmi, and E. Sangiorgi, "Low field electron and hole mobility of SOI transistors fabricated on ultrathin silicon films for deep submicrometer technology application", *IEEE Trans. Electron Dev.*, vol. 48, no. 12, pp. 2842–2850, 2001.

[17] A. Ohata, M. Cassé, and S. Cristoloveanu, "Front and back channel mobility in ultrathin SOI-MOSFETs by front-gate split CV method", *Solid-State Electron.*, vol. 51, no. 2, pp. 245–251, 2007.

[18] M. J. Sherony, L. T. Su, J. E. Chung, and D. A. Antoniadis, "SOI MOSFET effective channel mobility", *IEEE Trans. Electron Dev.*, vol. 41, no. 2, pp. 276–278, 1994.

[19] J. Choi, Y. Park, and H. Min, "Electron mobility behavior in extremely thin SOI MOSFET's", *IEEE Electron Dev. Lett.*, vol. 16, no. 11, pp. 527–529, 1995.

[20] M. Bawedin, S. Cristoloveanu, J. G. Yun, and D. Flandre, "A new memory effect (MSD) in fully depleted SOI MOSFETs", *Solid-State Electron.*, vol. 49, no. 9, pp. 1547–1555, 2005.

[21] M. S. Krishnan, Y. Yeo, Q. Lu, T. King, J. Bokor, and C. Hu, "Remote charge scattering in MOSFETs with ultra-thin gate dielectrics", in *IEDM'98 Tech. Dig.*, San Francisco, USA, 1998, pp. 571–573.

[22] J. Pretet, T. Matsumoto, T. Poiroux, S. Cristoloveanu, R. Gwoziecki, C. Raynaud, A. Roveda, and H. Brut, "New mechanism of body charging in partially depleted SOI-MOSFETs with ultra-thin gate oxides", in *Proc. 32th Eur. Solid-State Dev. Res. Conf. ESSDERC'02*, Firenze, Italy, 2002, pp. 515–518.

[23] A. Modelli, "Valence band electron tunneling in metal-oxide-silicon-structures", *Appl. Surf. Sci.*, vol. 30, pp. 298–303, 1987.

[24] M. Cassé, J. Pretet, S. Cristoloveanu, T. Poiroux, C. Fenouillet-Beranger, F. Fruleux, C. Raynaud, and G. Reimbold, "Gate-induced floating-body effect in fully-depleted SOI MOSFETs with tunneling oxide and back gate biasing", *Solid-State Electron.*, vol. 48, no. 7, pp. 1243–1247, 2004.

[25] K. Akarvardar, S. Cristoloveanu, and P. Gentil, "Threshold voltage model of the SOI 4-gate transistor", in *IEEE Int. SOI Conf.*, Charleston, USA, 2004, pp. 89–90.

[26] M. Gaillardin, P. Paillet, V. Ferlet-Cavrois, S. Cristoloveanu, O. Faynot, and C. Jahan, "High tolerance to total ionizing dose of Ω-shaped gate field-effect transistors", *Appl. Phys. Lett.*, vol. 88, p. 223511, 2006.

[27] T. Ernst and S. Cristoloveanu, "Buried oxide fringing capacitance: a new physical model and its implication on SOI device scaling and architecture", *IEEE Int. SOI Conf.*, Rohnert Park, USA, 1999, pp. 38–39.

Akiko Ohata received the M.Sc. degree in theoretical physics and Ph.D. degree in electronics from Tokyo Institute of Technology, Japan. She has engaged in research of Si quantum devices, single electron devices and SOI MOSFETs in Toshiba Corporation, RIKEN Institute, Tokyo Institute of Technology, Gunma University and IMEP.
e-mail: ohata@enserg.fr
IMEP Minatec-INPG
3 Parvis Louis Néel
Grenoble BP257, France

Romain Ritzenthaler was born in Rouen, France, in 1980. He received the M.Sc. degree from the Ecole Nationale Superieure de Physique de Grenoble (ENSPG) in 2003, Grenoble, France. He is currently pursuing his Ph.D. degree with the CEA-Laboratoire d'Electronique et Technologie de l'Information (LETI), Grenoble, France. His work is dedicated to the realization, characterization and simulation of the FinFET devices.
e-mail: ritz_romain@hotmail.com
IMEP Minatec-INPG
3 Parvis Louis Néel
Grenoble BP257, France

CEA-LETI
17 avenue des Martyrs
38054 Grenoble cedex 9, France

Sorin Cristoloveanu received the Ph.D. (1976) and the French Doctorat ès-Sciences (1981) from the National Polytechnic Institute, Grenoble, France. He joined the Centre National de la Recherche Scientifique (CNRS) and became a Director of Research in 1989. He worked at JPL (Pasadena) and at the Universities of Maryland, Florida (Gainesville) and Nashville. He served as Director of the LPCS Laboratory and of the Center for Advanced Projects in Microelectronics (CPMA Grenoble). He is the author or co-author of more than 600 technical journal papers and conference presentations (including 80 invited contributions). He authored or edited 15 books, and

organized 14 international conferences. He is a Fellow of IEEE, a Fellow of the Electrochemical Society, and Editor of "Solid-State Electronics".

e-mail: sorin@enserg.fr
IMEP Minatec-INPG
3 Parvis Louis Néel
Grenoble BP257, France

Olivier Faynot received the M.Sc and Ph.D. degrees from the Institut National Polytechnique de Grenoble, France, in 1991 and 1995, respectively. His doctoral research was related to the characterization and modeling of deep submicron fully depleted SOI devices fabricated on ultra-thin SIMOX wafers. He joined LETI (CEA-Grenoble, France) in 1995, working on simulation and modeling of deep submicron fully and partially depleted SOI devices. His main activity was the development of a dedicated partially depleted SOI model, called LETI SOI. From 2000 to 2002, he was involved in the development of a sub 0.1 μm partially depleted technology. Since 2003, he is in charge of the development of advanced single and multiple-gate fully depleted SOI technologies with high k and metal gate. He is author and co-author of more than 70 scientific publications on SOI in journals and international conferences. From 2001 to 2003, Doctor Faynot was on the IEEE International SOI Conference Committee. In 2002 and 2003, he was in the International Electron Device Meeting (IEDM) technical committee. Since 2004, Doctor Faynot joined the Solid State Device and Materials (SSDM) Conference Committee.

e-mail: ofaynot@cea.fr
CEA-LETI
17 avenue des Martyrs
38054 Grenoble cedex 9, France

Invited paper

Challenges for 10 nm MOSFET process integration

Mikael Östling, Bengt Gunnar Malm, Martin von Haartman, Julius Hållstedt, Zhen Zhang, Per-Erik Hellström, and Shili Zhang

Abstract— An overview of critical integration issues for future generation MOSFETs towards 10 nm gate length is presented. Novel materials and innovative structures are discussed. The need for high-*k* gate dielectrics and a metal gate electrode is discussed. Different techniques for strain-enhanced mobility are discussed. As an example, ultra thin body SOI devices with high mobility SiGe channels are demonstrated.

***Keywords*—sstrained silicon, silicon germanium, silicon-on-insulator (SOI), high-k dielectrics, hafnium oxide, nano-wire, low-frequency noise, mobility, metal gate.**

1. Introduction

The International Technology Roadmaps for Semiconductors (ITRS) [1] identifies a number of challenges for continued successful scaling of MOSFET technology. It is clear that new materials and process modules will be needed to meet the ITRS roadmap requirements and enhance performance for a given technology node. In this review we focus on the challenges for the 45 nm node and beyond. Although some of the discussed topics, e.g., strain-enhanced mobility, were introduced already at the 90 nm node their importance will certainly continue to increase. The main integration issues, which will be presented here, are:

- Replacing the standard SiO_2 gate oxide or oxynitride by a high-*k*/metal gate stack. This transition is required at an equivalent oxide thickness of about 1.2 nm, which has already been used in volume production from the 90 nm node.
- Strain-enhanced mobility. Process induced stress is now widely used to improve the performance of both n- and p-channel MOSFETs. Uniaxial stress, induced locally in the channel region, is the preferred integration method, while different types of strain-engineered substrates, usually with biaxial strain, are also very promising candidates.
- As the channel length is scaled down, leading to increased current density, the parasitic resistances in the extension and source/drain regions must be minimized. The ITRS roadmap clearly indicates that the main obstacle for the ultimate scaling towards 10 nm is the source/drain and contact resistance, which cannot be reduced enough in relation to the increasing current densities in sub 50 nm multi-gate devices [2]. A possible solution to this, is the use of complementary Schottky contacts to PMOS and NMOS, respectively [3].

Other integration issues include the choice of structure-planar or multigate and substrate type bulk, SOI or even strained virtual substrates. Conventional planar CMOS on bulk substrates has a significant limitation due to poor control of short channels effects (SCE). A promising alternative is ultra-thin body (UTB) SOI MOSFETs with lowly doped channels, which also offer higher mobility in addition to reduced SCE and junction leakage. Using double or multiple gates improves the electrostatic control of the channel. Of the different types of multi gate devices the FinFET [4] has received the most attention. A comprehensive analysis of FinFET structures showed that the double gate structure is preferred over more advanced triple gate or gate-all around structures [5]. One of the key metrics is the effective channel width, which can be achieved for a given layout area. The channel width is traded off against sub-threshold slope or other indicators of degraded short SCE. An innovative device structure, featuring an inverted-T channel, was recently demonstrated [6]. This device combines the thin body SOI and the FinFET structures to achieve better on-current to area ratio. For the ultimate CMOS, silicon nanowires are promising, thanks to the optimized SCE control, using a gate-all-around structure [7].

Optical lithography of 10 nm gate lines will be a serious challenge, due to effects such as line edge roughness (LER) and line width roughness (LWR). If electron beam lithography is used lateral straggle gives rise to similar issues. As we approach the 10 nm node length good control of the effective channel length must be maintained to minimize short channel effects and fluctuations in, e.g., threshold voltage. An illustration of LER is given in Fig. 1,

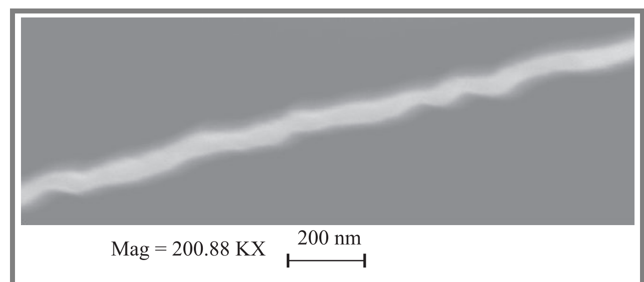


Fig. 1. High resolution SEM image of a 40 nm polysilicon nanowire, showing the effect of line edge roughness of about 4–7 nm.

which shows a poly-silicon nano-wire with approximately 40 nm width. In this case the LER was found to be correlated for the two edges, which minimizes the actual variations of LWR [8]. To form the line a combination of op-

tical lithography and etching/deposition was used-so called spacer lithography [9, 10]. This patterning technique is applicable to typical MOSFET structures, i.e., the gate stripe, but cannot directly be generalized to other patterns such as contact hole openings.
This paper is organized in three main sections – parasitic resistance, high-k and metal gate integration and strain-enhanced mobility. In each of the sections experimental results are shown from advanced nano-scale CMOS devices.

2. Parasitic resistances

The total resistance of a MOSFET transistor is determined by the sum of the channel resistance and a parasitic contribution from the source/drain regions including the actual contact resistance between silicide and highly doped silicon, sheet resistance of the silicide layer, accumulation, and spreading resistances. Typical total resistance versus gate bias curves are shown in Fig. 2 for 50 nm physical gate

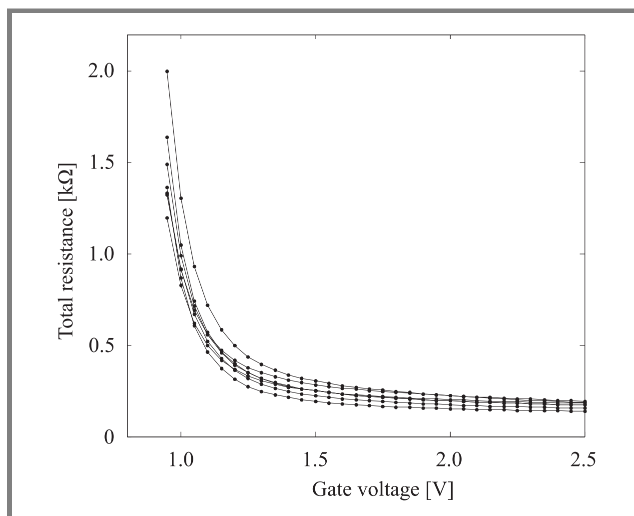


Fig. 2. Total resistance (channel + source/drain and contacts) versus gate bias for 50 nm gate length NMOS transistors.

length transistors. The spread in resistance at high gate voltage is related to the parasitic contributions, whereas the low gate voltage region includes the effect of gate length variation. For a general discussion it is convenient to consider the accumulation resistance as a (small) part of the bias dependent channel resistance. The other contributions will be discussed in more detail below. The requirements on the silicide contact resistivity to highly doped n- or p-type materials have been stated in the ITRS roadmap. For current technology generations values between $1.3–1.6 \cdot 10^{-7}\ \Omega/\text{cm}^2$ are assumed. For the year 2008, corresponding to a physical gate length of 23 nm a value of $8.3 \cdot 10^{-8}\ \Omega/\text{cm}^2$ is required. Especially for contacts to p-type material this is very hard to fulfill. Several groups have suggested that SiGe should be used in the source/drain region, to increase the solubility of boron dopant atoms and to reduce the potential barrier between metal and semiconductor [11, 12]. These values are applicable to a planar MOSFET on bulk substrates. For fully depleted UTB SOI an additional requirement on the thickness of a raised source/drain thickness is given.
To illustrate this the resistance was calculated for two different scenarios, with and without the raised source drain and compared to roadmap values, as shown in Fig. 3. The pur-

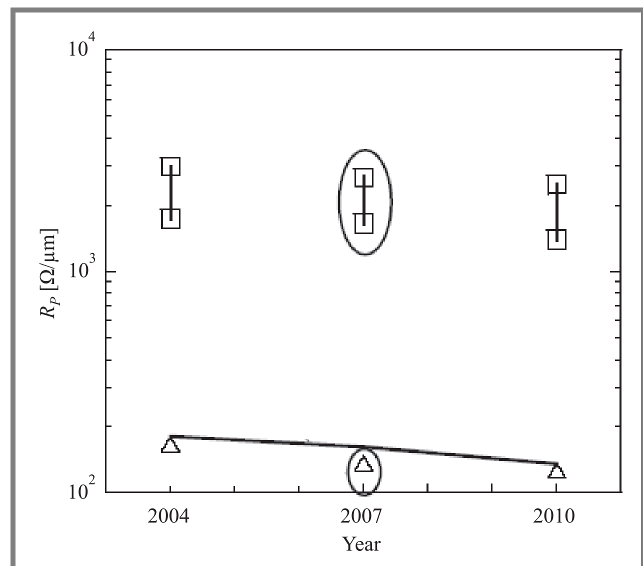


Fig. 3. Total parasitic resistance of fully silicided source/drain (squares) and raised source/drain structure (triangles). The solid line represents the ITRS roadmap recommendations [34].

pose of the raised source/drain is to provide a larger contact area, which is not limited by the thin silicon body thickness. Another related issue is the tradeoff between contact area, spreading resistance and layout density in multigate structures such as the FinFET. As discussed in [2] the contact width is shared between two or possibly three gates and the current spreading occurs in 3-dimensional way as compared to two dimensions only for the standard planar devices. One way to increase the contact area between the narrow fins and the silicide is selective epitaxy on both the top and sidewalls. This can be viewed as the 3D equivalent of the raised structure in UTB devices. A similar approach, taking advantage of the increased contact area for a wrapped contact, has also been proposed to contact carbon nano-tube field-effect transistors [8].

3. High-k and metal gate integration

Low-power applications such as battery operated handheld devices require a reduced gate leakage current. To reduce the gate leakage, standard oxynitride gate insulators will be replaced by high-k dielectrics. Among the promising candidates for the 45 nm technology node [13] are hafnium oxides (HfO_2) and hafnium silicates HfSiON with a high-k value in the range 10–15, which should be compared to 3.9 for SiO_2 and 6–7 for the oxynitrides. This leads

to significantly reduced gate leakage for the same equivalent oxide thickness (EOT). The main issues related to these types of dielectric materials, which still have to be addressed by researchers, are the high number of fixed/trapped charges and interface states. Both threshold voltage stability and low-field mobility are negatively affected by the high amount of charge present in the high-k oxides [14]. While the reduced mobility can be partially offset by strain enhancement techniques, the poor threshold voltage control and possible reliability problems cannot be accepted. An additional complication is the poor thermal stability of high-k materials. The dielectrics should be stable during high temperature processing steps (mainly source/drain activation anneals), since, e.g., re-crystallization can increase the gate leakage current. For the ultimate scaling of CMOS, below 10 nm gate length, other high-k materials such as La_2O_3, with a larger k value might be of interest [15, 16]. The choice of suitable materials is limited by the additional constraint that the band gap offset should be large enough compared to silicon. In some cases the offset to either the conduction or valence band is too small. By considering the increased fringing field, due to the higher k value, the influence on short channel effects and switching speed can be analyzed to find an optimum k value close to 30 [17].

High-k materials are often used in combination with different metal gate electrodes, e.g., TiN, TaN [18, 19]. Metal gates are important for several reasons, including the ability to control threshold voltage by tuning the work function of the gate electrode. For nitrided metal gates the tuning can be done either during the reactive sputter deposition or by subsequent nitrogen ion implantation [20, 21]. This allows reduced channel doping and hence higher mobility in both bulk and thin body SOI devices. Furthermore, metal gates do not suffer from depletion, which in turn decreases the EOT, compared to the case with a highly doped polysilicon gate electrode. For successful metal gate integration, selective etching processes, with high anisotropy, need to be developed for patterning of 10 nm gate lengths. The use of fully nickel silicided (FUSI) polysilicon gates offers a more straightforward approach in this respect, since the patterning of polysilicon gates is more mature. In this case, the work function control can be achieved by dopant pile-up at the metal gate/oxide interface. The combination of FUSI and high-k has generated a lot of attention recently [22, 23].

In the following section we discuss the influence of high-k gate dielectric materials and metal gates on the mobility and low-frequency noise in more detail. Experimental results are shown for PMOS devices, which featured a strained SiGe channel for improved hole mobility and where either TiN or poly-SiGe were used as midgap gate materials. The dielectrics discussed here are deposited by atomic-layer-deposition (ALD). The ALD technique, which is a pulsed chemical vapour deposition (CVD) process, allows arbitrary combinations of films to be deposited using a range of different precursors [24]. In other similar studies which focus on the impact on low-frequency noise hafnium based gate oxides were deposited by metal organic CVD (MOCVD) [25].

We have investigated different combinations of Al_2O_3 with k of 9 and HfO_2 with k of 25. A sandwiched structure of $Al_2O_3/HfO_2/Al_2O_3$ was investigated and compared to single layer Al_2O_3 or $HfAlO_x$. The effective dielectric constant is reduced to about 10 but the properties for integration into a standard CMOS process are much better, due to improved interfacial conditions. In this approach the aluminum oxide forms the interface to the channel region (either Si or strained SiGe) as well as to the poly-SiGe gate electrode. Another important aspect of the ALD technique is the (in situ) surface treatment before dielectric deposition [26]. The presence of a surface oxide will influence the final film quality and especially for SiGe channels a high number of interface states might be observed. While it is possible to remove the native oxide using HCl vapour an amorphous oxide might be beneficial for the formation of the Al_2O_3 interface layer. In Fig. 4. a transmission electron microscopy (TEM) cross-section of a transistor with a surface SiGe channel and a high-k gate stack is shown. Note that a SiO_2 interface layer is visible, especially close to the gate edge. The EOT (including the interface layer) was determined from C-V measurements and was found to be 1.9 nm.

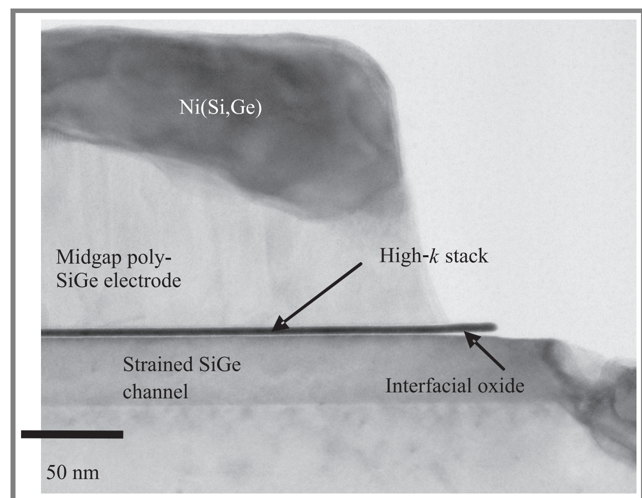


Fig. 4. TEM of a strained surface channel SiGe MOSFET with high-k $HfO_2/Al_2O_3/HfO_2$ gate stack and midgap SiGe gate electrode.

The interface state density (D_{it}) was extracted, to examine the quality of the high-k/strained SiGe channel interface. A relatively high value of $D_{it} = 7 \cdot 10^{12}$ $cm^{-2}eV^{-1}$ was obtained for devices with poly-SiGe gates. This suggests that high-temperature process steps after the high-k deposition might have degraded the film properties. For TiN metal gate devices with a reduced thermal budget excellent D_{it} values of $1.6 \cdot 10^{12}$ $cm^{-2}eV^{-1}$ were obtained in the case of SiGe channels and $3.3 \cdot 10^{11}$ $cm^{-2}eV^{-1}$ in the case of a Si-channel device with identical gate stack.

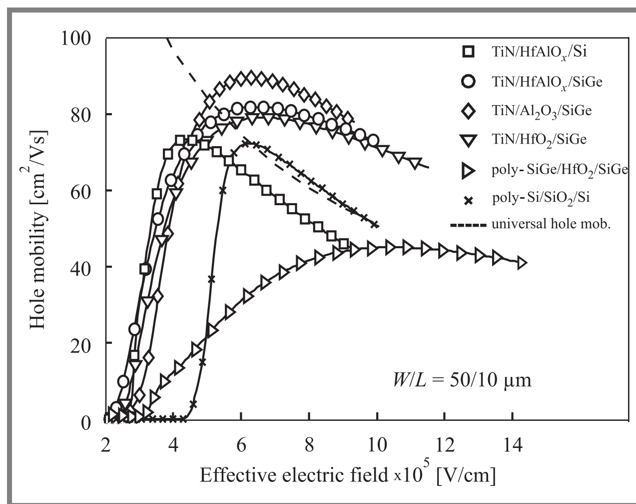


Fig. 5. Hole mobility for Si and SiGe surface channel devices with different high-k gate stacks and SiGe/TiN gate electrodes.

The purpose of introducing a surface SiGe channel is to obtain high hole mobility and carrier confinement. In Fig. 5 mobility values are compared for Si and SiGe channel devices. The SiGe devices show significantly better mobility than the Si control, which is very close to the theoretical curve. It is interesting to observe that TiN metal gate devices compare favorably to the device with poly-SiGe gate. This can be related to a reduced phonon scattering due to screening by the metal gate.

Another issue with the increased D_{it} and the number of fixed charges (N_t) in the high-k dielectrics is the possible influence on the low-frequency noise [27]. Compared to the case for buried channel SiGe devices, where the carriers are physically separated from the (oxide) interface, a much stronger influence on carrier mobility due to coulomb scattering and trapping/de-trapping in slow states will be observed for surface channel devices. In Fig. 6 low-frequency

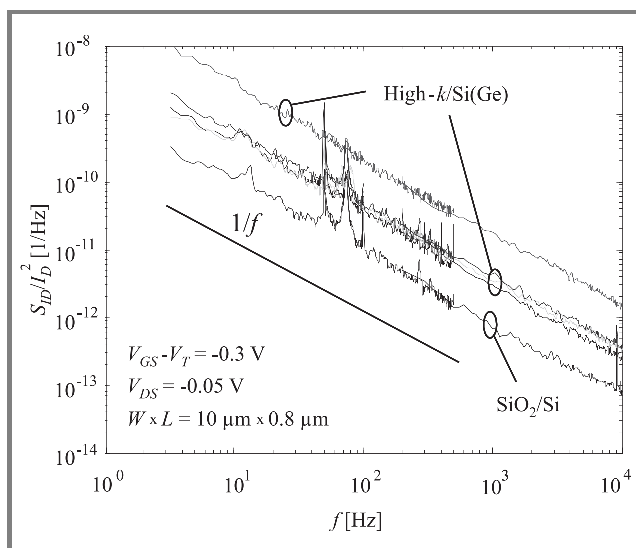


Fig. 6. Drain current noise power spectral density for high-k surface channel SiGe devices and a SiO_2/Si channel control device.

noise spectra for devices with different gate stacks on SiGe channels are compared to a SiO_2 reference. The highest noise (top curve) is observed for the case with a single layer Al_2O_3 gate. Gate stacks with either HfO_2 or $HfAlO_x$ sandwiched between Al_2O_3 layers show better noise performance. The SiO_2 reference shows the best noise performance, indicating that further optimization of the high-k gate stack is needed for low noise applications. However, the reduced phonon scattering in metal gate devices also affects the noise. It was found in [28] that the combination of high-k metal gate reduces the low-frequency noise in strong inversion.

4. Strain-enhanced mobility

For high-performance applications the challenge is mainly to maintain sufficiently high drive current for short-channel devices which suffer from short channel effects and high parasitic resistances. For higher drive current and increased switching speed the focus is on different methods of mobility enhancement, using strain. For CMOS applications both higher hole and electron mobility are desired. For PMOS the first attempts at increased channel mobility were done by selective SiGe epitaxy in the channel region [29]. However, from the 90 nm technology node, selective SiGe growth in the source/drain has emerged as the preferred method to create compressive strain in the PMOS channel [30]. Significantly increased electron mobility has also been demonstrated in NMOS devices, where a dielectric capping layer, commonly silicon nitride, introduces a tensile strain in the channel region. In this approach the strain in both PMOS and NMOS channels becomes uniaxial, which is beneficial compared to biaxial strain. It is important to note that the NMOS and PMOS can be optimized independently of each other. Very high mobility can also be achieved for both electrons and holes using so called virtual substrates, with a thin Si-channel on top of a relaxed SiGe buffer layer [31, 32]. There are several issues with the virtual substrate technique, including a poor thermal conductivity and a high intentional concentration of defects (dislocations). Furthermore the mobility increase is smaller for holes, which is not advantageous for CMOS applications, where the PMOS has the most need for performance increase.

Finally we give an example of UTB SOI devices with strained channels. The UTB devices offer significantly improved control of the short channel effects, compared to bulk devices, with the same gate length. No intentional substrate doping is needed in fully depleted devices and hence the threshold voltage is controlled only by the silicon body thickness and the gate work function. According to the scaling rules, the thickness of the thin silicon body layer should be less than one third of the gate length. Therefore, typical silicon layer thickness is in the order of 10–15 nm for a 50 nm gate length device. Such thin layers can be achieved from a starting material (SOI wafer) with silicon thickness of a few hundred nanometers by a combination of sacrificial oxidation [33] and silicon etching in HCl chem-

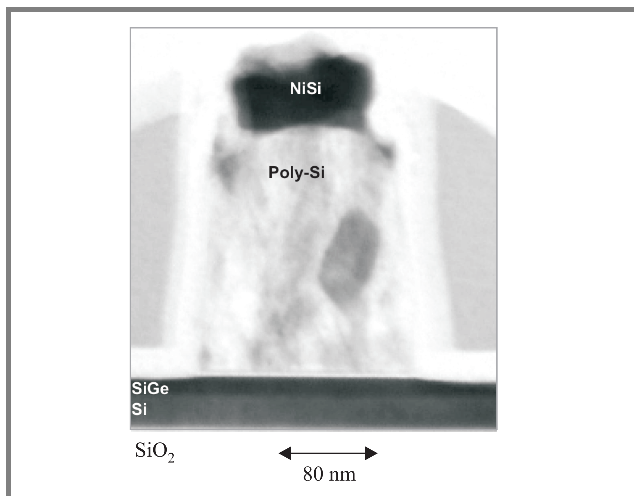


Fig. 7. Strained SiGe channel on ultra-thin body (∼20 nm) SOI substrate.

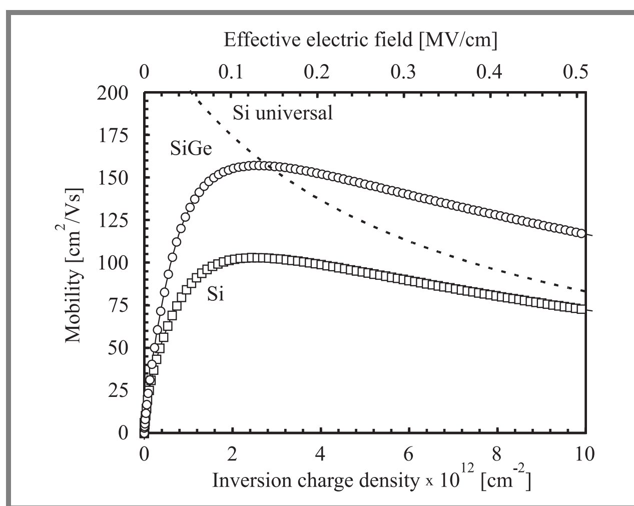


Fig. 8. The p-channel mobility in Si and strained SiGe transistors on ultra-thin body SOI.

istry. Compared to a bulk device with similar SCE control the UTB SOI devices have higher channel mobility thanks to the low doping. To further enhance mobility, strained channels can be incorporated on SOI [33], either by wafer bonding or epitaxial techniques. We have successfully implemented compressively strained SiGe and SiGeC layers in UTB SOI PMOSFETs. An example of an 80 nm gate device is shown in Fig. 7. A high quality SiGe 8 nm layer has been grown by RPCVD on top of the thinned down silicon. The fabricated devices show good performance, in terms of I-V characteristics. A significantly increased effective hole mobility, extracted from long channel split C-V measurements is demonstrated in Fig. 8. Compared to the Si control the effective hole mobility is increased by approximately 60%.

5. Conclusions

New materials and innovative device structures suitable for the ultimate scaling of CMOS to 10 nm gate lengths have been discussed. A combination of strained channels and hafnium based oxides/silicates will fulfill the drive current performance requirements for the 45 nm node. For future scaling an appropriate structure based on multiple gates will probably be needed to control the short channel effects. Among the challenges for the research community are a reduced contact resistance especially for novel multi gate devices, and gate dielectrics with higher k values based, e.g., on rare earth metals. Transistors with a combination of high-k gate dielectric and metal gate electrode show promising results both for mobility and low-frequency noise thanks to reduced phonon scattering.

Acknowledgement

The authors wish to thank the researchers and Ph.D. students at ICT/EKT for their valuable contributions to this paper. In particular we would like to mention Drs H. H. Radamson, Y.-B. Wang, J. Seger, D. Wu, and C. Isheden. We wish to thank Dr. G. Sjöblom, Dr. J. Westlinder, and Dr. J. Olsson, Uppsala University, for their contribution to the high-k work. The nano-scale CMOS development was also supported in the EU Network of Excellence, SiNano. The SSF and Vinnova are acknowledged for financial support in the high-frequency silicon and the high-speed/frequency and optoelectronics program, respectively.

References

[1] International Technology Roadmap for Semiconductors (ITRS), 2005, http://www.itrs.net/Common/2005ITRS/Home2005.htm

[2] A. Dixit, A. Kottantharayil, N. Collaert, M. Goodwin, M. Jurczak, and K. De Meyer, "Analysis of the parasitic S/D resistance in multiple-gate FETs", *IEEE Trans. Electron Dev.*, vol. 52, pp. 1132–1140, 2005.

[3] J. Kedzierski, P. Xuan, E. H. Anderson, J. Bokor, T.-J. King, and C. Hu, "Complementary silicide source/drain thin-body MOSFETs for the 20 nm gate length regime", in *Tech. Dig. IEDM*, San Francisco, USA, pp. 57–60, 2000.

[4] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm", *IEEE Trans. Electron Dev.*, vol. 47, pp. 2320–2325, 2000.

[5] J.-W. Yang and J. G. Fossum, "On the feasibility of nanoscale triple-gate CMOS transistors", *IEEE Trans. Electron Dev.*, vol. 52, pp. 1159–1164, 2005.

[6] L. Mathew, M. Sadd, S. Kalpat, M. Zavala, T. Stephens, R. Mora, S. Bagchi, C. Parker, J. Vasek, D. Sing, R. Shimer, L. Prabhu, G. O. Workman, G. Ablen, Z. Shi, J. Saenz, B. Min, D. Burnett, B.-Y. Nguyen, J. Mogab, M. M. Chowdhury, W. Zhang, and J. G. Fossum, "Inverted T channel FET (ITFET) – fabrication and characteristics of vertical-horizontal, thin body, multi-gate, multi-orientation devices, ITFET SRAM bit-cell operation. A novel technology for 45 nm and beyond CMOS", in *Tech. Dig. IEDM*, Washington, USA, 2005, pp. 731–734.

[7] S. D. Suk, S.-Y. Lee, S.-M. Kim, E.-J. Yoon, M.-S. Kim, M. Li, Ch. W. Oh, K. H. Yeo, S. H. Kim, D.-S. Shin, K.-H. Lee, H. S. Park, J. N. Han, C. J. Park, and J.-B. Park, "High performance 5 nm radius twin silicon nanowire MOSFET (TSNWFET): fabrication on bulk Si wafer, characteristics, and reliability", in *Tech. Dig. IEDM*, Washington, USA, 2005, pp. 735–738.

[8] J. Hållstedt, P.-E. Hellström, Z. Zhang, B. G. Malm, J. Edholm, J. Lu, S.-L. Zhang, H. H. Radamson, and M. Östling, "A robust spacer gate process for deca-nanometer high-frequency MOSFETs", *Microelectron. Eng.*, vol. 83, pp. 434–439, 2006.

[9] Y.-K. Choi, T.-J. King, and C. Hu, "A spacer patterning technology for nanoscale CMOS", *IEEE Trans. Electron Dev.*, vol. 49, pp. 436–441, 2002.

[10] Y.-K. Choi, T.-J. King, and C. Hu, "Spacer FinFET: nanoscale double-gate CMOS technology for the terabit era", *Solid-State Electron.*, vol. 46, pp. 1595–1601, 2002.

[11] S. Gannavaram, N. Pesovic, and C. Ozturk, "Low temperature (800°C) recessed junction selective silicon-germanium source/drain technology for sub-70 nm CMOS", in *Tech. Dig. IEDM*, San Francisco, USA, 2000, pp. 437–440.

[12] C. Isheden, P.-E. Hellström, H. H. Radamson, S.-L. Zhang, and M. Östling, "MOSFETs with recessed SiGe source/drain junctions formed by selective etching and growth", *Electrochem. Solid State Lett.*, vol. 7, pp. G53–G55, 2004.

[13] M. A. Quevedo-Lopez, S. A. Krishnan, P. D. Kirsch, H. J. Li, J. H. Sim, C. Huffman, J. J. Peterson, B. H. Lee, G. Pant, B. E. Gnade, M. J. Kim, R. M. Wallace, D. Guo, H. Bu, and T. P. Ma, "High performance gate first HfSiON dielectric satisfying 45 nm node requirements", in *Tech. Dig. IEDM*, Washington, USA, 2005, pp. 237–240.

[14] M. von Haartman, J. Westlinder, D. Wu, B. G. Malm, P. E. Hellström, J. Olsson, and M. Östling, "Low-frequency noise and Coulomb scattering in $Si_{0.8}Ge_{0.2}$ surface channel pMOSFETs with ALD Al_2O_3 gate dielectrics", *Solid-State Electron.*, vol. 49, pp. 907–914, 2005.

[15] H. Iwai, S. Ohmi, S. Akama, C. Ohshima, A. Kikuchi, I. Kashiwagi, J. Taguchi, H. Yamamoto, J. Tonotani, Y. Kim, I. Ueda, A. Kuriyama, and Y. Yoshihara, "Advanced gate dielectric materials for sub-100 nm CMOS", in *Tech. Dig. IEDM*, San Francisco, USA, 2002, pp. 625–628.

[16] A. C. Jones, H. C. Aspinall, P. R. Chalker, R. J. Potter, K. Kukli, A. Rahtu, M. Ritala, and M. Leskela, "Recent developments in the MOCVD and ALD of rare earth oxides and silicates", *Mater. Sci. Eng. B (Solid-State Mater. Adv. Technol.)*, vol. 118, pp. 97–104, 2005.

[17] N. R. Mohapatra, M. P. Desai, S. G. Narendra, and V. R. Rao, "The effect of high-*k* gate dielectrics on deep submicrometer CMOS device and circuit performance", *IEEE Trans. Electron Dev.*, vol. 49, pp. 826–831, 2002.

[18] D. Wu, A.-C. Lindgren, S. Persson, G. Sjöblom, M. von Haartman, J. Seger, P.-E. Hellström, J. Olsson, H.-O. Blom, S.-L. Zhang, M. Östling, E. Vainonen-Ahlgren, W.-M. Li, E. Tois, and M. Tuominen, "A novel strained $Si_{0.7}Ge_{0.3}$ surface-channel pMOSFET with an ALD $TiN/Al_2O_3//HfAlO_x/Al_2O_3$ gate stack", *IEEE Electron Dev. Lett.*, vol. 24, pp. 171–173, 2003.

[19] L.-Å. Ragnarsson, S. Severi, L. Trojman, D. P. Brunco, K. D. Johnson, A. Delabie, T. Schram, W. Tsai, G. Groeseneken, K. de Meyer, S. de Gendt, and M. Heyns, "High performing 8 Å EOT HfO_2/TaN low thermalbudget n-channel FETs with solid-phase epitaxially regrown (SPER) junctions", in *Tech. Dig. VLSI Symp.*, Kyoto, Japan, 2005, pp. 234–235.

[20] H.-C. Wen, K. Choi, P. Majhi, H. Alshareef, C. Huffman, and B. H. Lee, "A systematic study of the influence of nitrogen in tuning the effective work function of nitrided metal gates", in *IEEE Int. Symp. VLSI Technol.*, Hsinchu, Taiwan, 2005, pp. 105–106.

[21] H. Wakabayashi, Y. Saito, K. Takeuchi, T. Mogami, and T. Kunio, "A dual-metal gate CMOS technology using nitrogen-concentration-controlled TiNx film", *IEEE Trans. Electron Dev.*, vol. 48, pp. 2363–2369, 2001.

[22] Y. H. Kim, C. Cabral Jr., E. P. Gusev, R. Carruthers, L. Gignac, M. Gribelyuk, E. Cartier, S. Zafar, M. Copel, V. Narayanan, J. Newbury, B. Price, J. Acevedo, P. Jamison, B. Linder, W. Natzle, J. Cai, R. Jammy, and M. Ieong, "Systematic study of workfunction engineering and scavenging effect using NiSi alloy FUSI metal gates with advanced gate stacks", in *Tech. Dig. IEDM*, Washington, USA, 2005, pp. 657–660.

[23] A. Lauwers, A. Veloso, T. Hoffmann, M. J. H. van Dal, C. Vrancken, S. Brus, S. Locorotondo, J.-F. de Marneffe, B. Sijmus, S. Kubicek, T. Chiarella, M. A. Pawlak, K. Opsomer, M. Niwa, R. Mitsuhashi, K. G. Anil, H. Y. Yu, C. Demeurisse, R. Verbeeck, M. de Potter, P. Absil, K. Maex, M. Jurczak, S. Biesemans, and J. A. Kittl, "CMOS integration of dual work function phase controlled Ni FUSI with simultaneous silicidation of NMOS (NiSi) and PMOS (Ni-rich silicide) gates on HfSiON", in *Tech. Dig. IEDM*, Washington, USA, 2005, pp. 661–664.

[24] M. Leskelä and M. Ritala, "Atomic layer deposition (ALD): from precursors to thin film structures", *Thin Solid Films*, vol. 409, pp. 138–146, 2002.

[25] C. Claeys, E. Simoen, A. Mercha, L. Pantisano, and E. Young, "Low-frequency noise performance of HfO_2 based gate stacks", *J. Electrochem. Soc.*, vol. 152, pp. F115–F123, 2005.

[26] D. Wu, H. Radamson, P.-E. Hellström, S.-L. Zhang, M. Östling, E. Vainonen-Ahlgren, E. Tois, and M. Tuominen, "Influence of surface treatment prior to ALD high-*k* dielectrics on the performance of SiGe surface-channel pMOSFETs", *IEEE Electron Dev. Lett.*, vol. 25, pp. 289–291, 2004.

[27] M. von Haartman, D. Wu, B. G. Malm, P. E. Hellstrom, S. L. Zhang, and M. Östling, "Low-frequency noise in $Si_{0.7}Ge_{0.3}$ surface channel pMOSFETs with ALD HfO_2/Al_2O_3 gate dielectrics", *Solid-State Electron.*, vol. 48, pp. 2271–2275, 2004.

[28] M. von Haartman, B. G. Malm, and M. Östling, "Comprehensive study on low-frequency noise and mobility in Si and SiGe pMOSFETs with high-*k* gate dielectrics and TiN gate", *IEEE Trans. Electron Dev.*, vol. 53, pp. 836–843, 2006.

[29] M. von Haartman, A.-C. Lindgren, P.-E. Hellström, M. Östling, T. Ernst, L. Brévard, and S. Deleonibus, "Influence of gate width on 50 nm gate length $Si_{0.7}Ge_{0.3}$ channel PMOSFETs", in *Proc. 33rd Eur. Solid-State Dev. Res. Conf. ESSDERC'03*, Estoril, Portugal, 2003, pp. 529–532.

[30] S. E. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, G. Glass, T. Hoffman, J. Klaus, Z. Ma, B. Mcintyre, A. Murthy, B. Obradovic, L. Shifren, S. Sivakumar, S. Tyagi, T. Ghani, K. Mistry, M. Bohr, and Y. El-Mansy, "A logic nanotechnology featuring strained-silicon", *IEEE Electron Dev. Lett.*, vol. 25, pp. 191–193, 2004.

[31] E. Kasper and K. Lyutovich, "Strain adjustment with thin virtual substrates", *Solid-State Electron.*, vol. 48, pp. 1257–1263, 2004.

[32] P.-E. Hellström, J. Edholm, M. Östling, S. Olsen, A. O'Neill, K. Lyutovich, M. Oehme, and E. Kasper, "Strained-Si NMOSFETs on thin 200 nm virtual substrates", in *Solid State Dev. Res. Symp.*, Bethesda, USA, 2005.

[33] J. Seger, P. E. Hellström, J. Lu, B. G. Malm, M. von Haartman, M. Östling, and S. L. Zhang, "Lateral encroachment of Ni-silicides in the source/drain regions on ultrathin silicon-on-insulator", *Appl. Phys. Lett.*, vol. 86, pp. 253507–1, 2005.

[34] J. Seger, "Interaction of Ni with SiGe for electrical contacts in CMOS technology", Ph.D. thesis, KTH – Royal Institute of Technology, Sweden, 2005.

Mikael Östling holds a position as Professor in solid state electronics. Since 2001 he is Head of the Department of Microelectronics and Information Technology. In December 2004 he was appointed Dean, School of Information and Communication Technology, KTH, Sweden. He was a senior visiting Fulbright Scholar 1993–94 with the center for integrated systems at Stanford University, and a visiting Professor with the University of Florida,

Gainesville. He initiated and was appointed program Director by the Swedish Foundation for Strategic Research for a silicon nanoelectronics national program 2000–2007. His research interests are silicon/silicon germanium devices and process technology for very high frequency, as well as device technology for wide band-gap semiconductors with special emphasis on silicon carbide and nitride based structures. He has supervised 20 Ph.D. theses work, and been the author of 8 book chapters and about 300 scientific papers published in international journals and conferences. He is an editor of the "IEEE Electron Device Letters" and a fellow of the IEEE.
e-mail: ostling@imit.kth.se
School of Information Technology
KTH – Royal Institute of Technology
P.O. Box 229
164 40 Kista, Sweden

Bengt Gunnar Malm was born in Stockholm, Sweden, in 1972. He received the M.Sc. degree in engineering physics and radiation science from Uppsala University, Uppsala, Sweden, in 1997, and the Ph.D. degree in solid-state electronics from Royal Institute of Technology (KTH), Stockholm, in 2002. Currently, he is a Senior Researcher at the Solid-State Device Lab, School of Information and Communication Technology, KTH. His research interests include characterization, modeling, and process development of Si- and SiGe-based devices and circuits for RF/wireless and high-speed applications, optimization and modeling of RF properties, noise and distortion and thermal effects. He has co-supervised 1 Ph.D. student project, and has published or co-authored more than 20 scientific papers in international journals and conferences.
School of Information Technology
KTH – Royal Institute of Technology
P.O. Box 229
164 40 Kista, Sweden

Martin von Haartman was born in Södertälje, Sweden, in 1976. He received the M.Sc. degree in electrical engineering and the Ph.D. degree in solid state electronics from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2001 and 2006, respectively. His research interests include device physics, characterization and modeling of Si and SiGe based CMOS and bipolar devices with the main focus on low-frequency noise. Doctor Martin von Haartman was a recipient of the IEEE Electron Devices Society Graduate Student Fellowship in 2004.
e-mail: mvh@kth.se
School of Information Technology
KTH – Royal Institute of Technology
P.O. Box 229
164 40 Kista, Sweden

Julius Hållstedt was born in Stockholm, Sweden, in 1977. He received the M.Sc. degree in materials science from Royal Institute of Technology (KTH), Stockholm, Sweden, in 2002. He is currently working toward the Ph.D. degree in solid state electronics at KTH. His current research focuses on SiGe and SiGeC integration in novel MOSFET device structures.
e-mail: juliush@kth.se
School of Information Technology
KTH – Royal Institute of Technology
P.O. Box 229
164 40 Kista, Sweden

Zhen Zhang was born in Anhui, China, in 1979. He received the B.Sc. degree at University of Science and Technology of China (USTC) in 2000 and the M.E. degree at Shanghai Institute of Ceramics, Chinese Academy of Sciences (SICCAS) in 2003. Since 2003, he has been a Ph.D. student of solid-state electronics at Royal Institute of Technology (KTH), Stockholm, Sweden. His current research interests include: silicon based nanostructures and nano-devices covering materials, processing, integration and novel structures; Schottky source/drain MOSFET; field-effect sensors based on silicon nanowires for detection of bio-chemical molecules.
School of Information Technology
KTH – Royal Institute of Technology
P.O. Box 229
164 40 Kista, Sweden

Shili Zhang works as a Professor of solid-state electronics at Royal Institute of Technology (KTH), Stockholm, Sweden. He got his B.Sc. in 1982 and M.Sc. in 1985, then Ph.D. in 1990 at KTH. His major research interest focuses on exploiting and utilizing the well-established silicon technology for applications in nanoelectronics and field-effect sensing, specifically: silicon based nanostructures and nano-devices covering materials, processing, integration and novel structures; carbon nanotube based nanoelectronics with focus on tube separation, device fabrication and tube-metal contacts; field-effect sensors based on silicon nanowires and carbon nanotubes for detection of bio-chemical molecules; printed RF-ID sensor solutions based on carbon nanotubes.
School of Information Technology
KTH – Royal Institute of Technology
P.O. Box 229
164 40 Kista, Sweden

Per-Erik Hellström (Hellberg) was born in Stockholm, Sweden, in 1970. He received the M.Sc. and Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1995 and 2000, respectively. His Ph.D. thesis dealt with polycrystalline $Si_{1-x}Ge_x$ as gate material for CMOS technology. Since 2000, he has been a Research Associate at the Department of Microelectronics and Applied Physics at KTH. His current research interests include nano-scaled MOSFET devices and $Si_{1-x}Ge_x$ in CMOS technology. In 2006 he received the "docent" degree at KTH.

Department of Microelectronics and Applied Physics
KTH – Royal Institute of Technology
P.O. Box 229
164 40 Kista, Sweden

Invited paper

Review and perspective of high-*k* dielectrics on silicon

Stephen Hall, Octavian Buiu, Ivona Z. Mitrovic, Yi Lu, and William M. Davey

Abstract— The paper reviews recent work in the area of high-*k* dielectrics for application as the gate oxide in advanced MOSFETs. Following a review of relevant dielectric physics, we discuss challenges and issues relating to characterization of the dielectrics, which are compounded by electron trapping phenomena in the microsecond regime. Nearly all practical methods of preparation result in a thin interfacial layer generally of the form SiO_x or a mixed oxide between Si and the high-*k* so that the extraction of the dielectric constant is complicated and values must be qualified by error analysis. The discussion is initially focussed on HfO_2 but recognizing the propensity for crystallization of that material at modest temperatures, we discuss and review also, hafnia silicates and aluminates which have the potential for integration into a full CMOS process. The paper is concluded with a perspective on material contenders for the "end of road map" at the 22 nm node.

Keywords— high-k dielectrics, dielectric constant, interfacial layer, hafnia, aluminates, silicates.

1. Introduction

The challenges around the search for a replacement for silicon dioxide as the gate dielectric in the ubiquitous CMOS technology are well known to the community. The unique and excellent intrinsic properties of SiO_2 together with its compatibility with high temperature manufacturing process and the natural abundance of silicon, have underpinned the entire development of the \$200B silicon industry. However, the relentless miniaturization or scaling of the MOS transistor and associated infrastructure on chip create the demand for ever thinner gate oxides. There is a school of thought that the rate of technology scaling is exceeding the rate at which circuit designers can fully exploit the advantages of a given technology node but the momentum behind the industry and the customer demand for ever faster processing, sets the paradigm. The well-known issue for the gate oxide then, is that it must become vanishingly thin to control adequately the electrostatics of the MOSFET channel and so win in the competition with the drain voltage encroachment, to minimise undesirable short channel effects. In fact, the International Technology Roadmap for Semiconductors (ITRS) predicts equivalent oxide thicknesses of 1 nm in 2007, reducing to 0.35 nm for the 22 nm node [1]. Notwithstanding other issues, at least three mono-layers of SiO_2 are required so that "bulk" like properties can be achieved giving a lower limit for the native oxide in any event, of about 0.7 nm [2]. Such oxide thickness reduction comes at a price as the quantum mechanical current leakage through the gate becomes prohibitively high and so therefore is the stand by power dissipation in chips containing a billion individual transistors. The gate leakage must be reduced without compromising the current drive (I_{ON}) of the transistor so materials with higher dielectric constant (k) are sought to allow a thicker oxide for the same gate capacitance, so mitigating the leakage problem.

Silicon dioxide is a hard act to follow and any contender must satisfy stringent requirements. We can summarize the requirements [3] as relating to:

- thermodynamic stability in contact with Si;
- a high enough k to warrant the cost of R&D – including a propensity to be scaled;
- band offsets for electrons and holes > 1 eV which translates to band gap energies $(E_g) > 5$ eV taking into account the inverse relationship between E_g and k;
- stability through a high temperature CMOS manufacturing process and finally, acceptable reliability and wear-out attributes.

With these constraints in mind, the periodic table reveals (perhaps not surprisingly) relatively few contenders. In the short to medium term, taking account of ITRS performance requirements, the metallic oxide HfO_2 is the main contender and its silicates and aluminates can reduce the tendency for crystallization occurring at temperatures beyond about 450°C, at the expense of a slight reduction in the k values. Looking at the requirements for the 22 nm node, contenders such as Pr, La look to be promising, while Gd, Ce and Sm oxides are also worthy of consideration in many respects.

A vast array of metrological techniques has been developed over the years for characterizing SiO_2 both in terms of very fundamental physics and also engineering perspectives. The techniques represent a self-consistent methodology for engineering highly reliable gate oxides in a mass production environment and there is considerable confidence in this technology. It soon became clear that the new dielectrics have properties that require the experimental techniques to be reexamined and recalibrated. Taking the case of HfO_2, the rapid electron trapping which gives rise to a significant threshold instability has required the establishment of high bandwidth measuring systems with data capture times of the order of microseconds or less. Turning to physical characterization, spectro-ellipsometry (SE) represents a very powerful tool for obtaining fundamental parameters and properties such as oxide thickness, dielectric

constant and band gap. Studies of the losses (complex part of the permittivity or absorption coefficient) can provide information regarding defect levels in the oxide band gap and importantly, at the band edges. The SE together with standard C-V and I-V measurements can provide powerful self-consistent, non-destructive schemes for characterizing these materials.

To address the above points, we have structured the paper as follows. In Section 2 we outline the dielectric physics that underpins the engineering of the k value and in Section 3 we discuss methodologies for accurately determining experimental k values from C-V, spectroscopic ellipsometry and complementary techniques. We present a brief overview of trapping effects in high-k materials in Section 4. Section 5 contains a review and appraisal of aluminates and silicates of hafnia that allow for higher process temperatures. Section 6 presents the case for likely materials for the 22 nm node and the paper is concluded in Section 6.

2. How to increase k – dielectric physics

Figure 1 presents a useful description of the frequency dependence of the dielectric function over a wide range of frequencies. In general, the "zero frequency" value of the dielectric constant can be seen to have two components: a "high – frequency" one, where the contribution of electronic polarization dominates and one related to the ionic

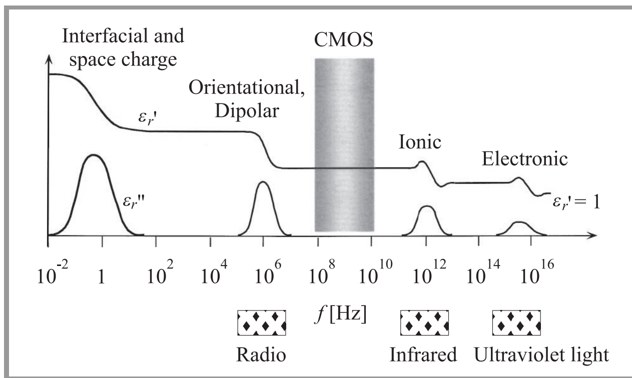


Fig. 1. The frequency dependence of the dielectric function ($\varepsilon_r = \varepsilon_r' + j\varepsilon_r''$, where ε_r' is the real part and ε_r'' – imaginary part of the complex dielectric permittivity) [4].

contribution [5]. In the CMOS frequency window, we can see that electronic and ionic processes contribute to k and we consider that the permittivity is given by the relation:

$$\varepsilon_{ox} = \varepsilon_\infty + \varepsilon_{latt} \,, \tag{1}$$

where ε_{ox} is equivalent to k.

The electronic component, which arises from simple polarization of the atoms, is the main component for SiO_2 and the simple relationship $n \sim \sqrt{\varepsilon_\infty}$ links the refractive index, readily measurable in ellipsometry, to the permittivity, giving $\varepsilon_{ox} \sim \varepsilon_\infty$. The essence of increasing k then is to choose materials that can contribute a large lattice component. Table 1 shows some values of these parameters for different crystalline forms of hafnia. We can see that ε_{latt} can vary from about 2 to over 25 depending on the crystalline form. Without going into details of the crystallography, we can simply make the point that the permittivity can vary over a wide range depending on the form of the material and hence the method used to prepare it. Furthermore, amorphous forms are preferred for processing in any event. The variability of k with the structure of various metallic oxides is pointed out from another perspective in [6], by consideration of the Clausius-Mossotti (C-M) theory which links the k to the polarizability α, and the volume of the unit cell, V_m as described in Eq. (2):

Table 1
The electronic (ε_∞) and lattice (ε_{latt}) permittivity components for different crystalline forms of hafnia [5]

Crystalline phase	ε_∞	ε_{latt}	$\varepsilon_{ox}(k)$
c-HfO_2	5.37	20.80	26.17
t-HfO_2: parallel	5.13	14.87	20.00
t-HfO_2: perpendicular	5.39	27.42	32.81
m-HfO_2: yy	–	10.75	10.75
m-HfO_2: xx	–	11.70	11.70
m-HfO_2: zz	–	7.53	7.53
m-HfO_2: xz	–	1.82	1.82

$$\varepsilon_r = \frac{\left(1 + \frac{2}{3} 4\pi \frac{\alpha}{V_m}\right)}{1 - \frac{1}{3} 4\pi \frac{\alpha}{V_m}} \,. \tag{2}$$

In essence, larger atoms yield more polarization and hence higher k values. The C-M equation reveals that k raises steeply as the ratio α/V_m increases demonstrating the strong connection with the structure and nature of the material.

3. How to measure k – methodologies

The simplest, most convenient and appropriate way to measure k is from a C-V plot although care is required to ensure that a genuine response is obtained which usually means adjusting the data for a variety of frequency dependent parasitic phenomena such as series resistance [7], leakage current [8] and lossy interfacial capacitance [9]. Furthermore, in the case of ultra-thin gate dielectrics, the accumulation capacitance does not readily saturate to the oxide capacitance C_{ox}, because the oxide capacitance is large compared to that of the space charge in accumulation and also due to the quantization of energy levels in the accumulation layer. The difference between the measured accumulation capacitance and the true oxide capacitance must be taken into account in the extraction of capacitance equivalent thickness (CET) and the effective dielectric constant.

The Maserjian technique [10] provides a simple method to extract the oxide capacitance from a C-V plot under accumulation conditions. Computer code is available to account for accumulation layer related quantum mechanical effects and oxide leakage [11]. Having obtained a genuine accumulation and hence oxide capacitance, C_{ox} (considered here per unit area) and if no transition layer (SiO_x) is present, the permittivity can simply be obtained from the relation that $k = C_{ox} t_{ox}$, where t_{ox} has been measured from ellipsometry (see later). In practice, and usually intentionally, a so-called transitional layer (TL) is present between the substrate and the high-k layer and a two-capacitor model (with perfect, planar interfaces, i.e., no roughness) can be used to analyse the MOS structure which may be written:

$$\frac{EOT}{\varepsilon_{SiO_2}} = \frac{t_{TL}}{\varepsilon_{TL}} + \frac{t_{hi-k}}{k}, \tag{3}$$

where we extend the definition of equivalent oxide thickness (EOT) to incorporate the TL. The electronic properties of TL are dependent on the nature of its formation and it can be designated SiO_x in general where often $x = 2$ is used and the permittivity of 3.9 is then considered. However, it is important to note that x-values greater or less than 2 can arise therefore affecting the permittivity value; for instance, $x > 2$ for oxides that are heavily strained, and $x < 2$ for "unintentional" oxides that grow after an HF dip treatment. It is important therefore to understand the nature of the oxide and if possible measure its electrical and optical properties independently. We have illustrated the importance of this issue in a recent publication [12] and the main points are summarized here. We considered four samples of varying Hf stoichiometry with TLs produced by rapid thermal oxidation (RTO) and so-called chemical oxidation associated with SC1/SC2 cleaning procedures. The chemical nature of the TL plays a major role in the growth dynamics of the HfO_2 layer; it has been shown that the use of chemical oxides, which are characterized by higher OH concentration, results in almost linear growth, while obtaining a two-dimensional uniform coverage with HfO_2 [13]. The thickness of the TL was measured in situ prior to the deposition by angle resolved X-ray photoelectron spectroscopy (ARXPS), that of the hafnia layer by spectro-ellipsometry and the Hf content by Rutherford backscattering spectrometry (RBS). The SE measurements were performed in the 184–1700 nm spectral range, at three various angles of incidence (65°, 70°, and 75°) for an increased sensitivity. A simple model was used for establishing the thickness of the HfO_2 film. The model incorporated a Si substrate and a TL for which the optical properties were established on a control sample together with a Cauchy layer for describing the hafnia layer. The thickness of the hafnia layer was extracted in the spectral region where the HfO_2 layer was transparent.

The maximum capacitance was measured and the model of [12] used to extract k using four values for ε_{TL}, in the 3–4.3 range. This choice of values for the ε_{TL} has been observed in TL's in our experience and also reported in the literature. Figure 2 shows the results of the extraction and it is striking that the spread of the k-values is relatively large (10.6–19). When increasing the [Hf] concentration in the layers, the spread of results is reduced significantly from factors of ~ 7 to ~ 1. Furthermore, for the same thickness of TL and nearly the same concentration of Hf, the samples with the chemical oxide TL have significantly higher relative dielectric constant: (14–19), as compared to 12–15. These results demonstrate the sensitivity of

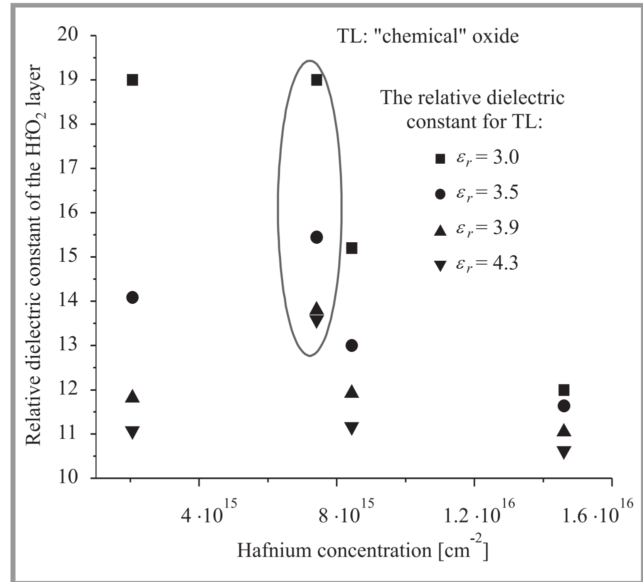


Fig. 2. Relative dielectric constant of the HfO_2 layer versus Hf content calculated for different values of relative permittivity of transitional layer [12].

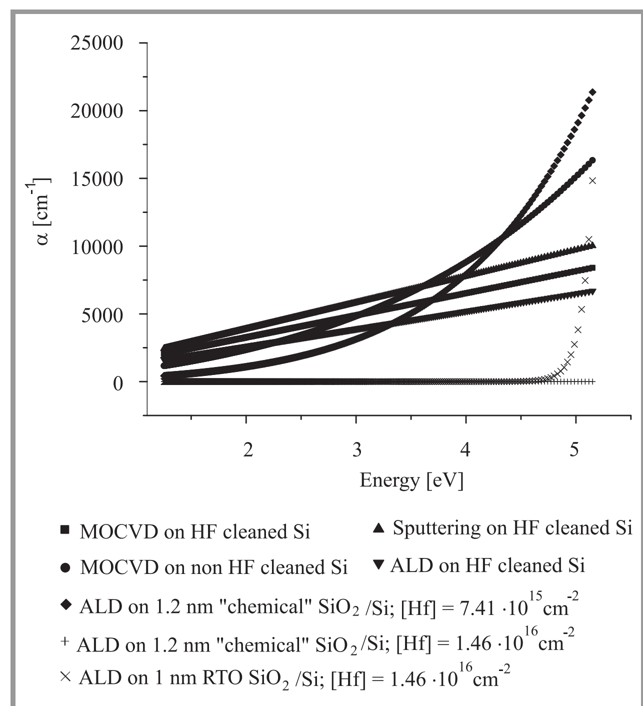


Fig. 3. The absorption coefficient α versus energy for hafnia layers prepared by different techniques [14].

the extraction technique to the TL characteristics. Clearly for low Hf density, there is doubt that the model of Eq. (3) is valid and suggests a non-uniform or mixed TL and possible poor morphology of the hafnia layer. This point was pursued in the study of [14] where the effects of pre-treatment were investigated. Figure 3 shows the absorption coefficient (α) extracted from the imaginary part of the complex permittivity, measured with SE. The best result (lowest α) is obtained with atomic layer deposition (ALD) on chemical oxide TL with ALD on RTO exhibiting a sharp increase in absorption at an energy E $\sim$ 4.7 eV. There is some evidence that such an energy level is associated with the oxygen vacancy in hafnia films [15]. The worst case is for a film on chemical oxide with sub-stoichiometric Hf. Other samples in the study incorporated HF surface preparation and inferior properties are apparent for both ALD and metal organic chemical vapour deposition (MOCVD) hafnia deposition.

We can summarise this section by reinforcing the importance of taking careful consideration of the TL when extracting *k* from C-V data and would advocate the use of error bars and a clear description of methodology when quoting experimental values.

4. Parasitic charges: measurement challenges

A key advantage of the SiO_2 system is the excellent electrical properties in terms of electron and hole traps. As-grown and appropriately annealed thermal oxide contains very low trap levels with relatively small capture cross-sections. As well as being virtuous for integrated circuit engineering, it has made far easier the characterization and study of the properties of these traps. Investigation of trapping in SiO_2 has been a major activity for nearly 50 years with specialist conferences (e.g., INFOS, SISC) over much of this time. It soon became apparent, when studying hafnia and other high-*k* dielectrics, that electron trapping in particular was extremely severe and there was a need for specialized measurement configurations to characterize the extremely fast trapping kinetics. An analogue based technique whereby the drain current is monitored across a small drain load resistance and fed to an oscilloscope is a typical set-up [16] although the technique has been refined further [17].

A study by Zhao *et al.* [18] illustrates effectively the trapping time constants and it can be seen that data capture of the order of 10's of microsecond are required to capture the full extent of the trapping, as shown in Fig. 4. Translating the time-dependence of the voltage shifts with first order trapping theory reveals for as-grown electron traps, two effective capture cross-sections of the order of 10^{-15} cm^2 with concentrations of the order of 10^{12} cm^{-2}; these being very large values relative to SiO_2. It is important to point out that such trap concentrations confined within such thin films imply very closely spaced traps – of the order of nm's which makes the use of first order trapping theory controversial. However, the values at least convey the rapid nature of the trapping and are useful for providing a representation of the time constants associated with the phenomenon.

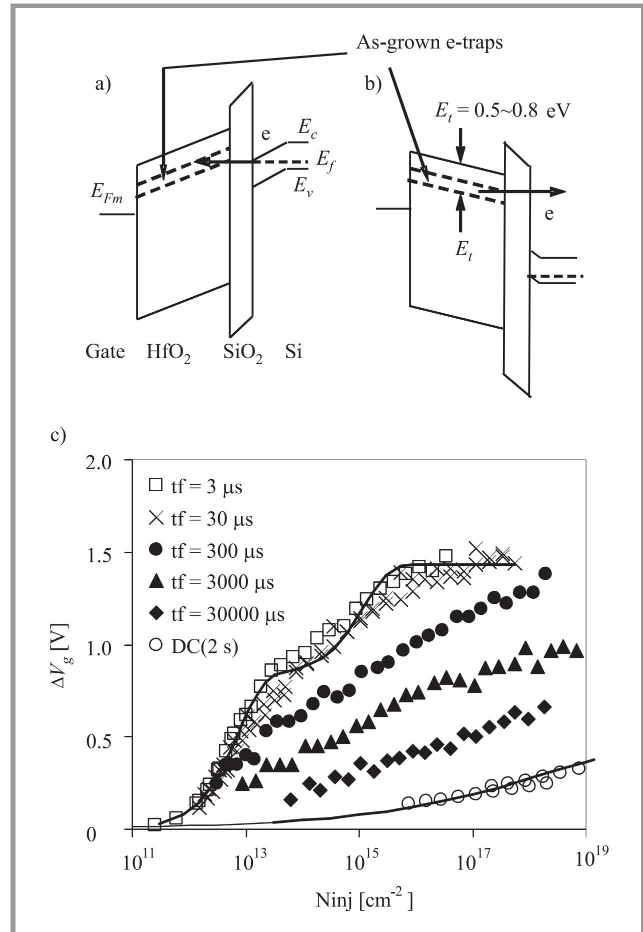


Fig. 4. Energy band diagrams of as-grown e-traps in HfO_2. (a) e trapping at $V_g > 0$ V. (b) e detrapping at $V_g < 0$ V. (c) Dynamic behavior of electron trapping measured by different techniques. The data represented by symbol "o" was measured by the traditional DC $I_d - V_g$ with V_g increasing with a step of 0.1 V for each point. The data in other symbols were obtained by the pulsed $I_d - V_g$ technique [18].

Other trapping studies show also that the films are rich in fixed positive charge with similar concentrations, as shown in Fig. 5. It is possible also to create positive charge by stressing, with similar concentrations to the as-grown ones [19]. The measurements are usually carried out on MOSTs but there is a great advantage to employing MOS capacitors due to the simplicity of the structure. Capacitor based measurements can be employed for rapid screening of new materials. We have developed a novel measurement system based on pulsing MOS capacitors [20]. Using this technique, which involves a deep-depleting voltage step, we can observe a positive

charge which would not readily be apparent from transistor based measurements which involve an inverted surface with a ready supply of minority carriers. The rapid removal of compensating electrons in the film reveals the influence of positive charge and the associated centroid induces an image charge in the substrate which is realized by a further extension of the depletion region.

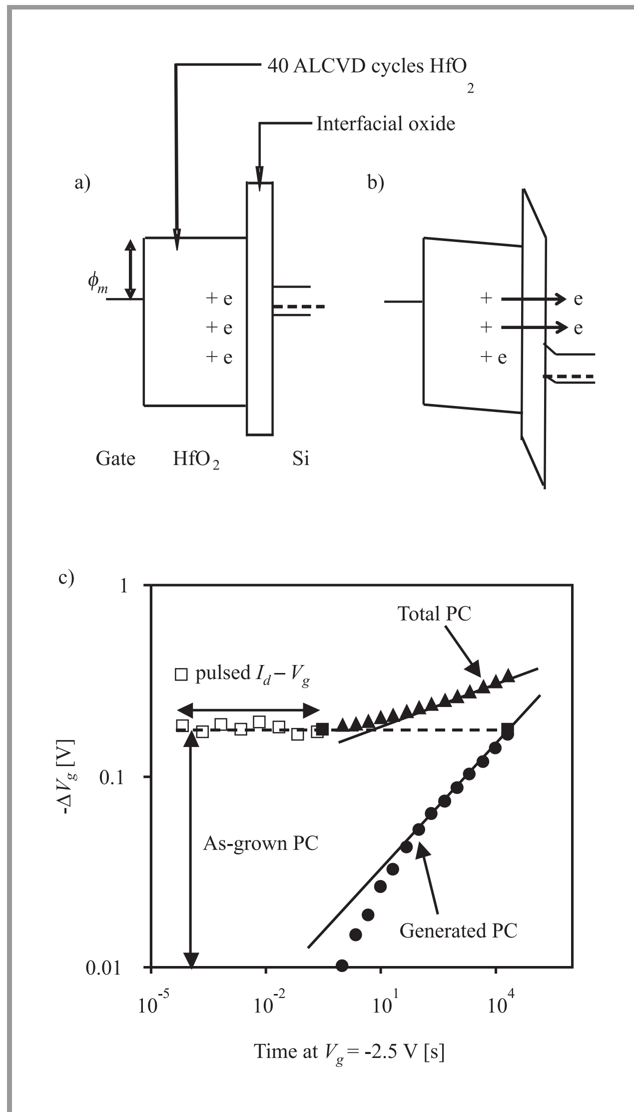


Fig. 5. A schematic illustration of the defect and physical process responsible for the ΔV_{th} of PMOSFETs. (a) Under $V_g = 0$ V, the donor-like defects are neutral. (b) Under $V_g < 0$ V, electrons tunnel away, leaving positive charges in the dielectric. (c) As-grown positive charge (PC) and generated PC. As-grown PC was measured by pulsed $I_d - V_g$ technique and generated PC was measured by traditional DC $I_d - V_g$ technique (after C. Z. Zhao *et al.*, unpublished).

This image charge then manifests itself as an extension to the depletion edge causing an undershoot in the capacitance (see Fig. 6b). The undershoot region can be further interrogated to reveal the rate at which the electrons are de-trapped; that is to say, the rate at which the positive charge is uncovered. As time progresses, the capacitance relaxes as electron-hole pairs are created in the depletion region. The oxide field also increases during the relaxation as the voltage across the depleted semiconductor is transferred to the oxide allowing tunnelling of minority

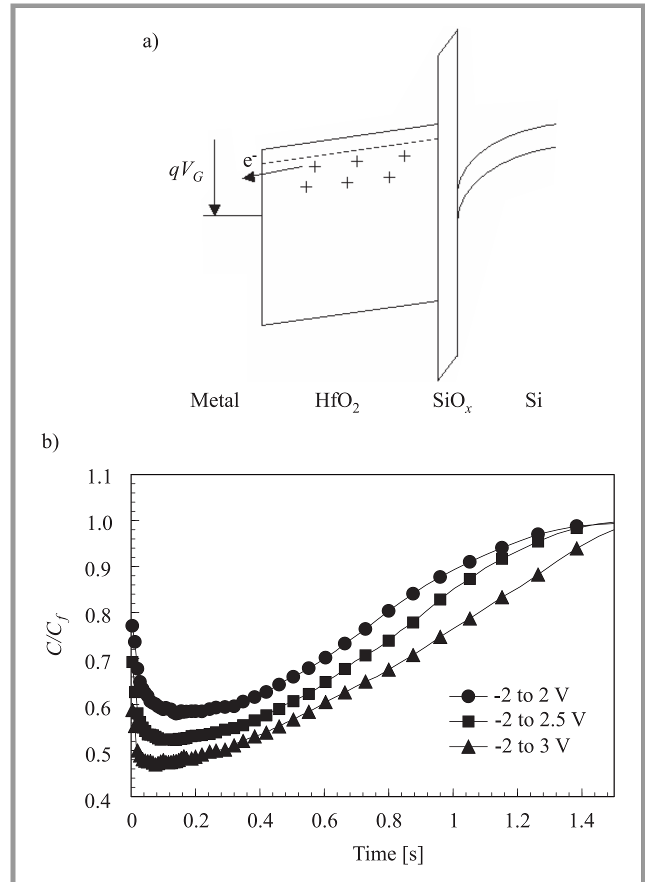


Fig. 6. (a) Energy band diagram showing positive charges are generated by electrons detrapping from pre-existed oxide defects. (b) Capacitance-transient curves of an HfO_2 sample showing undershoots [20].

carriers into the oxide and associated compensation of the positive charge. The method could be easily employed in "stress and sense" methodologies, for investigating trap creation.

5. Materials for manufacturability

Despite the advantages that HfO_2 possesses as a candidate for alternative high-k dielectric, one major problem associated with HfO_2 is the thermal instability. To minimise electrical and mass transport along grain boundaries and stabilise the interface between Si and metal oxide, it is preferable that the gate oxide remains amorphous throughout CMOS processing. Unfortunately, HfO_2 films crystallize at low temperatures of 450°C when deposited by molecular beam epitaxy (MBE) [21] to 530°C when deposited by ALD [22].

5.1. Aluminates

It has been reported that the crystalline temperature of HfO_2 can be increased by the incorporation of Al_2O_3 to form an HfAlO alloy, which will still have a relatively high dielectric constant (typically $k \sim 15$) whilst remaining amorphous up to high processing temperatures [23, 24]. It has been shown that Hf-aluminate film with 7% Al deposited by MOCVD remains amorphous up to 900°C [24]. Another group [25] also reported that Hf-aluminate film deposited by ALD can stay amorphous up to 1000°C rapid thermal anneal.

When incorporating Al_2O_3 into HfO_2, the large band gap of Al_2O_3 ($\sim$ 9 eV) also increases the band gap of the compound. Yu *et al.* [26] showed that the band gap of the HfAlO films (estimated by XPS) varies linearly from 5.25 to 6.52 eV with Al concentration from 9.6 to 33.9%. Using the spectro-ellipsometry we demonstrated that the band gap of HfAlO films deposited by MOCVD can be increased up to 7.9 eV with 38% Al [27]. When combined with electrical measurements, the ellipsometry data can provide valuable information related to the relative dielectric constant of the layers. Our results demonstrated the possibility of adjusting the relative dielectric constant of the layers in a wide range (9–17), when the aluminium concentration varies between 4.5% and 38%. This result is consistent with the results of Zhu *et al.* [28], who also reported the dielectric constant of HfAlO films deposited by jet vapour deposition decreases from 19.6 for HfO_2 to 7.6 for Al_2O_3.

Another feature for HfAlO is the high density of fixed oxide charge. Results reported by Bae *et al.* [29] show negative fixed charge of $1.5 \cdot 10^{12}$ cm^{-2} for HfAlO with 38% Al and $1 \cdot 10^{12}$ cm^{-2} for HfAlO with 20% Al. Our results [30] extended this relationship to HfAlO with Al concentration from 4.5 to 38%; the fixed charge density varies almost linearly from $4.8 \cdot 10^{11}$ to $1.1 \cdot 10^{12}$ cm^{-2}. This feature shows the possibility to adjust the threshold voltage simply by adjusting the ratio of precursors. Recent work [31] successfully demonstrated the attainment of symmetry threshold voltage in HfAlO based complementary MOSFETs by adjusting the Hf/Al ratio.

A few papers [24, 32] have reported large amounts of hysteresis observed in C-V measurements, indicating high densities of oxide traps in the HfAlO films. The measurements of the Doppler broadening spectra of annihilation radiation and the lifetime spectra by Uedono *et al.* [33] shows strong oxygen deficiency in the compound. Driemeier *et al.* [25] found that the oxygen deficiency increases with increasing Al/Hf ratio. The oxide traps may induce additional leakage and therefore the advantages of larger band gap and thermal stability may be traded off. In [27], the HfAlO film with 31.7% Al shows significant higher leakage than the film with 6.8% Al at as-deposited status. However after anneal in N_2 at 700°C, the film with 6.8% Al shows a sudden increase in leakage current. The leakage current of the HfAlO with 31.7% Al remains low due to the improved thermal stability by higher density of Al introduced. Our results [30] showed that HfAlO film with 22% Al has the lowest leakage current (@–1 V); further increase of Al concentration results in excessive leakage.

Hong *et al.* [34] conducted annealing studies on 7.3–7.8 nm thick HfAlO films with 14% Al deposited by thermal sputtering. They annealed the samples in an N_2 ambient for 5 min at temperatures from 500 to 900°C. The HRTEM images showed the interfacial layer growth between the HfAlO film and the Si substrate after anneal.

Cho *et al.* [35] carried out the annealing studies of ultra thin (1.3 nm) HfAlO films. The films were first deposited by ALD and subsequently annealed at 700°C for 60 s in an NH_3 atmosphere. They found that the near-edge x-ray absorption fine structure spectra of the HfO_2 components remained the same while the spectra of Al_2O_3 were changed after the anneal. This result indicates that the change in the bonding characteristics as the result of N incorporation is mainly caused by N incorporation into Al oxide.

Torii *et al.* [36] proposed the employment of HfAlO/SiON stack as gate dielectric and demonstrated successful integration into a standard CMOS process. The transistor achieved encouraging properties such as low EOT (1.1 nm), low leakage ($\sim 10^{-2}$ A/cm^2), low interface density ($2 \cdot 10^{11}$ eV^{-1}cm^{-2}), symmetrical threshold voltage and 92% electron mobility ($V_g = 1.1$ V) of those for SiO_2.

5.2. Silicates

Hafnium silicate films, $(HfO_2)_x(SiO_2)_{1-x}$, are being studied as an alternative to pure hafnium oxide due to comparable advantages such as an increased crystallization temperature [4], stable amorphous structure [37–39] which also resists oxygen diffusion [4, 37], reduced growth of interfacial layers at the silicon/high-k interface and higher values of band gap and effective electron rest mass resulting in reduced leakage [37]. Hafnium silicate films do however have the disadvantage of having a lower k value ($\sim 11-15$) [40, 41] than the pure oxide ($\sim 21-25$) [42, 43] reducing the scalability of the material.

Takeuchi and King in 2004 [44] compared the compositional dependency of the electrical properties of hafnium silicate films from published studies and found that there was a nonlinear dependency of the permittivity of the film with the permittivity decreasing with increasing concentration of incorporated silicon. The same work also reviewed experimental band gap results for hafnium silicate films of varying composition and observed that the compositional dependence of the band gap of hafnium oxide films has two distinct regions. The band gap of hafnium silicate films decreases linearly at a an approximate rate of 50 meV/% when the hafnium oxide content is increased, until the hafnium oxide content reaches 64%. At this stage the band gap becomes independent of hafnium oxide content and stays

constant at a value of 5.7 eV. The theoretical conduction and valence band offsets for an $Hf_{0.5}Si_{0.5}O_2$ film were shown to be 1.5 eV and 3.4 eV, respectively.

Cho *et al.* in 2005 [45] studied the dependence of hafnium silicate phase separation on the composition of the film using XPS finding that a silicon-dioxide-rich hafnium silicate sample ($x = 0.25$) could withstand temperatures greater than 900°C for 1 min in a nitrogen ambient without phase separation but that a hafnium rich hafnium silicate sample ($x = 0.75$) phase-separates at a temperature of 800°C. Thermal stability is a required property for high-k dielectrics due to current processes requiring the gate oxide to remain unaffected by an annealing temperature of 1000°C for 5 s to activate the polysilicon gate [46]. Wilk, Wallace, and Anthony in 2000 [47] were able to anneal silicon-rich (i.e., $x = 0.2$) hafnium silicate samples of thickness 3 nm for 20 s at temperatures of 1050°C in nitrogen without visible grain boundaries formation and proposed the resistance to crystallization may continue even for hafnium silicate films of hafnium content up to 30%.

Nitrogen incorporation into hafnium silicate films is known to be beneficial to their electrical properties such as further increase of the phase separation temperature of hafnium silicates [48], increased permittivity [43] and reduced boron penetration [49]. Cho *et al.* [45] annealed hafnium rich ($x = 0.75$) $\sim$ 3.5 nm thick hafnium silicate samples for 1 min at 900°C in either NH_3 or N_2. The sample annealed in pure nitrogen was seen to phase-separate to contain monoclinic HfO_2 grains, whereas the sample annealed in NH_3 remained stable with no visible phase separation [50]. It was seen that annealing in both atmospheres increased the Si/high-k interfacial layer by less than 1 nm, however annealing in N_2 caused the growth of a 1.4 nm overlayer which seriously increased the effective oxide thickness (EOT) of the film. In the same paper, Cho *et al.* showed results from 3 nm hafnium silicate samples of hafnium content ($x = 0.5$) annealed for 60 s in either NH_3 at a temperature of 750°C or N_2 at a temperature of 950°C. Cho *et al.* also reported that N_2 increased the EOT compared to the as-deposited film whereas NH_3 reduced the EOT compared to the as-deposited film. The samples annealed in N_2 however had superior electrical qualities having leakage currents an order of magnitude ($\sim 10^{-9}$ A/cm^2) lower than those of the NH_3 annealed samples ($\sim 10^{-8}$ A/cm^2) and having a higher effective mobility [43].

Nitrogen incorporation has, however, also been reported to reduce the conduction band and valence band offsets for a $(HfO_2)_{0.40}(SiO_2)_{0.60}$ by 0.33 eV and $\sim$ 1.2 eV, respectively, and reduce the band gap of the film by $\sim$ 1.50 eV [49]. The reduction in band offsets is not serious enough to affect the viability of nitrogen incorporating films, however leakage current will increase through such a film.

In our own laboratories, $(HfO_2)_x(SiO_2)_{1-x}/SiO_2$ $(0 < x < 1)$ gate stacks grown by MOCVD at IMEC were investigated using spectroscopic ellipsometry and electrical characterization techniques [51]. The optical constants, thickness of the layers and optical band gap for hafnium silicates of four concentrations were assessed using UV – NIR and deep UV spectral regions. The permittivity was seen to decrease from $\sim$ 21 for HfO_2 layer to $\sim$ 8 for Hf-silicate with $x = 0.3$. The results suggest that an Hf content above 60% is required to yield a permittivity higher than 10.

6. The way forward to the 22 nm node

Nag [52] explored the relationship between the mean atomic number of atoms constituting different semiconductors, $|Z|$, and the dielectric constant of these materials.

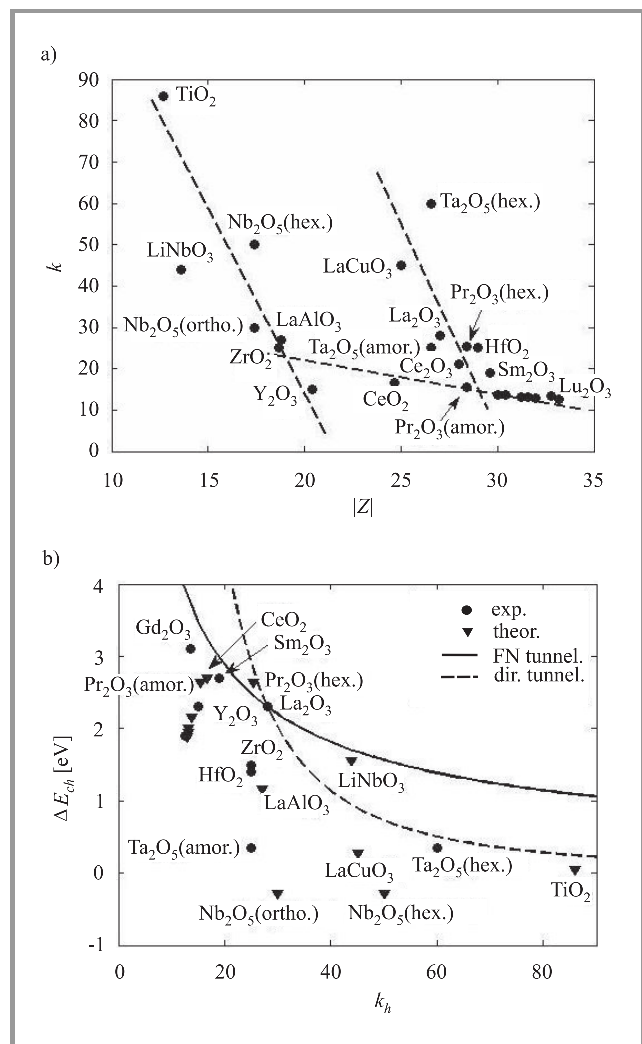


Fig. 7. (a) Experimental relative permittivity for some experimental gate dielectrics plotted against their mean atomic number [6]. (b) Conduction band offset versus the relative permittivity for experimental high-k dielectrics [6].

Busani and Devine [53] and Xue *et al.* [54] also pointed out corresponding relations for the rare earth oxides, but

beside any changes in polarizability, α, both cases suggested that a change in the molecular volume, V_m, was responsible for the relationship between the permittivity value and $|Z|$. Both cases modelled the relative permittivity value using the Clausius-Mossotti equation (see relation (2)).

Figure 7a shows the relative permittivity values versus the mean atomic number of some well-known and potentially suitable metal oxides, after Engstrom *et al.* [6]. In the same work, Pauling electronegativities were considered to allow prediction of the conduction band offsets for the oxides for which values were experimentally unknown, allowing predictions of conduction band offsets versus relative permittivity, as shown in Fig. 7b. Boundaries were established assuming the most stringent requirements for the 22 nm node, namely EOT = 0.5 nm and leakage $< 10^{-2}$ A/cm^2 at 1 V, considering both purely direct tunnelling and Fowler-Nordheim and are included on the plot. By assuming the most pessimistic scenario, which was that any materials with relative permittivity lower than that of lanthanum would suffer from direct tunnelling and above this Fowler-Nordheim tunnelling, it was predicted that only a few materials would be able to meet the requirements for the 22 nm node, namely Pr_2O_3 in the hexagonal phase, La_2O_3 and $LiNbO_3$. It was also suggested that Sm_2O_3, Ce_2O_2 and Gd_2O_3 were worthy of consideration. Of the lanthanides, La_2O_3 has been extensively studied by Iwai *et al.* and is perhaps the most serious contender for this node [55].

Kwo *et al.* [56] reported encouraging results of amorphous Gd_2O_3 films. The films were attained by electron beam evaporation using powder packed ceramic Gd_2O_3 sources. The scanning transmission electron microscopy (STEM) results showed that the film was 4.5 nm thick and no interfacial layer was observed. The C-V and I-V measurements showed the EOT of the film was 1.65 nm and the leakage (@ 1 V) was 10^{-4} A/cm^2, much lower comparing with SiO_2 with similar EOT. They also showed that these amorphous dielectrics could withstand annealing tests to a temperature of 850°C, as corroborated by the XPS analysis.

Ohmi *et al.* [57] made a comparative study of rare earth oxides grown by E-beam deposition. They found La_2O_3 possessed the lowest leakage and smooth interface among the rare earth oxides investigated. Other material such as Dy_2O_3 and Lu_2O_3 also showed good electrical properties but highly dependent on deposition processes [57].

Acknowledgements

The work has been funded by EPSRC, UK, with contributions from EU SINANO and PULLNANO. The authors acknowledge their collaborators within the SINANO framework for useful discussions and also acknowledge IMEC (Belgium) for supplying silicate samples and Prof. Paul Chalker and Dr R. J. Potter of Liverpool University, Materials Science Division for the hafnium aluminate samples associated with our results within the paper.

References

[1] International Technology Roadmap for Semiconductors (ITRS), www.itrs.net.

[2] D. A. Muller, T. Sorsch, S. Mocclo, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, "The electronic structure at the atomic scale of ultrathin gate oxides", *Nature*, vol. 399, pp. 758–761, 1999.

[3] J. Robertson, "Interfaces and defects of high-*k* oxides on silicon", *Solid-State Electron.*, vol. 49, pp. 283–293, 2005.

[4] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High-*k* gate dielectrics: current status and materials properties considerations", *J. Appl. Phys.*, vol. 89, no. 10, pp. 5243–5275, 2001.

[5] G.-M. Rignanese, "Dielectric properties of crystalline and amorphous transition metal oxides and silicates as potential high-*k* candidates: the contribution of density-functional theory", *J. Phys. Cond. Matt.*, vol. 17, no. 7, pp. R357–R379, 2005.

[6] O. Engstrom, B. Raeissi, S. Hall, O. Buiu, M. C. Lemme, H. D. B. Gottlob, P. K. Hurley, and H. Cherkaoui, "Navigation aids in the search for future high *k* dielectrics: physical and electrical trends", *Solid-State Electron.*, vol. 51, pp. 622–626, 2007.

[7] D. K. Schroder, *Semiconductor Material and Device Characterization*. New York: Wiley, 1998.

[8] K. J. Yang and C. Hu, "MOS capacitance measurements for high-leakage thin dielectrics", *IEEE Trans. Electron Dev.*, vol. 46, no. 7, pp. 1500–1501, 1999.

[9] K. S. K. Kwa, S. Chattopadhyay, N. D. Jankovic, S. H. Olsen, L. S. Driscoll, and A. G. O'Neill, "A model for capacitance reconstruction from measured lossy MOS capacitance-voltage characteristics", *Semicond. Sci. Technol.*, vol. 18, pp. 82–87, 2003.

[10] J. Maserjian, G. Petersson, and C. Svensson, "Saturation capacitance of thin oxide MOS structures and the effective surface density of states of silicon", *Solid-State Electron.*, vol. 17, no. 4, pp. 335–339, 1974.

[11] QMCV simulator developed by UC Berkeley Device Group, http://www-device.eecs.berkeley.edu/qmcv/index.shtml

[12] O. Buiu, S. Hall, O. Engstrom, B. Raeissi, M. Lemme, P. K. Hurley, and K. Cherkaoui, "Extracting the relative dielectric constant for high-*k* layers from CV measurements – errors and error propagation", *Microelectron. Reliab.*, vol. 47, pp. 678–681, 2007.

[13] M. L. Green, M.-Y. Ho, B. Busch, G. D. Wilk, T. Sorsch, T. Conard, B. Brijs, W. Vandervorst, P. I. Räisänen, D. Muller, M. Bude, and J. Grazul, "Nucleation and growth of atomic layer deposited HfO_2 gate dielectric layers on chemical oxide (Si-O-H) and thermal oxide (SiO_2 or Si-O-N) underlayers", *J. Appl. Phys.*, vol. 92, no. 12, pp. 7168–7174, 2005.

[14] O. Buiu, Y. Lu, I. Z. Mitrovic, S. Hall, P. Chalker, and R. J. Potter, "Spectro-ellipsometric assessment of HfO_2 thin films", *Thin Solid Films*, vol. 515, no. 2, pp. 623–626, 2006.

[15] K. Xiong and J. Robertson, "Point defects in HfO_2 high-*k* gate oxide", *Microelectron. Eng.*, vol. 80, pp. 408–411, 2005.

[16] A. Kerber, E. Cartier, L. Pantisano, R. Degraeve, G. Groeseneken, H. E. Maes, and U. Schwalke, "Charge trapping in SiO_2/HfO_2 gate dielectrics: comparison between charge-pumping and pulsed $I_D - V_G$", *Microelectron. Eng.*, vol. 72, no. 1-4, pp. 267–272, 2004.

[17] J. Mitard, C. Leroux, G. Ghibaudo, G. Reimbold, X. Garros, B. Guillaumot, and F. Boulanger, "Investigation on trapping and detrapping mechanisms in HfO_2 films", *Microelectron. Eng.*, vol. 80, pp. 362–365, 2005.

[18] C. Zhao, M. B. Zahid, J. F. Zhang, G. Groenseneken, R. Degraeve, and S. De Gendt, "Properties and dynamic behavior of electron traps in HfO_2/SiO_2 stacks", *Microelectron. Eng.*, vol. 80, pp. 366–369, 2005.

[19] C. Z. Zhao, J. F. Zhang, M. B. Zahid, G. Groeseneken, R. Degraeve, and S. De Gendt, "Impact of gate materials on positive charge formation in HfO_2/SiO_2 stacks", *Appl. Phys. Lett.*, vol. 89, pp. 023507-1-3, 2006.

[20] S. Hall, O. Buiu, and Y. Lu, "Direct observation of anomalous positive charge and electron trapping dynamics in high-k films using pulsed MOS capacitor measurements", *IEEE Trans. Electron Dev.*, vol. 54, no. 2, pp. 272–278, 2007.

[21] J.-H. Hong, T.-H. Moon, and J.-M. Myoung, "Microstructure and characteristics of the HfO_2 dielectric layers grown by metalorganic molecular beam epitaxy", *Microelectron. Eng.*, vol. 75, no. 3, pp. 263–268, 2004.

[22] E. P. Gusev, C. Cabral, M. Copel, C. D'Emic, and M. Gribelyuk, "Ultrathin HfO_2 films grown on silicon by atomic layer deposition for advance gate dielectric applications", *Microelectron. Eng.*, vol. 69, no. 2-4, pp. 145–151, 2003.

[23] M. Cho, H. B. Park, J. Park, and C. S. Hwang, "Thermal annealing effects on the structural and electrical properties of HfO_2/Al_2O_3 gate dielectric stacks grown by atomic layer deposition on Si substrate", *J. Appl. Phys.*, vol. 94, no. 4, pp. 2563–2571, 2003.

[24] R. J. Potter, P. A. Marshall, P. R. Chalker, S. Taylor, A. C. Jones, T. C. Q. Noakes, and P. Bailey, "Characterization of hafnium aluminate gate dielectrics deposited by liquid injection metalorganic chemical vapor deposition", *Appl. Phys. Lett.*, vol. 84, no. 20, pp. 4119–4121, 2004.

[25] C. Driemeier, K. P. Bastos, L. Miotti, I. J. R. Baumvola, N. V. Nguyen, S. Sayan, and C. Krug, "Compositional stability of hafnium aluminates thin films deposited on Si by atomic layer deposition", *Appl. Phys. Lett.*, vol. 86, pp. 221911-1-3, 2005.

[26] H. Y. Yu, M. F. Li, B. J. Cho, C. C. Yeo, and M. S. Joo, "Energy gap and band alignment for $(HfO_2)_x(Al_2O_3)_{1-x}$ on (100) Si", *Appl. Phys. Lett.*, vol. 81, no. 2, pp. 376–378, 2002.

[27] O. Buiu, Y. Lu, S. Hall, I. Z. Mitrovic, R. J. Potter, and P. R. Chalker, "Investigation of optical and electronic properties of hafnium aluminate films deposited by metal-organic chemical vapour deposition", *Thin Solid Films*, vol. 515, iss. 7-8, pp. 3772–3778, 2007.

[28] W. J. Zhu, T. Tamagawa, M. Gibson, T. Furukawa, and T. P. Ma, "Effect of Al inclusion in HfO_2 on the physical and electrical properties of the dielectrics", *IEEE Electron Dev. Lett.*, vol. 23, no. 11, pp. 649–651, 2002.

[29] S. H. Bae, C. H. Lee, R. Clark, and D. L. Kwong, "MOS characteristics of ultrathin CVD HfAlO gate dielectrics", *IEEE Electron Dev. Lett.*, vol. 24, no. 9, pp. 556–559, 2003.

[30] Y. Lu, O. Buiu, S. Hall, I. Z. Mitrovic, W. Davey, R. J. Potter, and P. R. Chalker, "Tuneable electrical properties of hafnium aluminate gate dielectrics deposited by metal organic chemical vapour deposition", *Microelectron. Reliab.*, vol. 47, pp. 722–725, 2007.

[31] M. Kadoshima, A. Ogawa, H. Ota, K. Iwamoto, M. Takahashi, N. Mise, S. Migita, M. Ikeda, H. Satake, T. Nabatame, and A. Toriumi, "Symmetrical threshold voltage in complementary metal-oxidesemiconductor field-effect transistors with $HfAlO_x$(N) achieved by adjusting Hf/Al compositional ratio", *J. Appl. Phys.*, vol. 99, pp. 054506-1-9, 2006.

[32] W. L. Liu, P. F. Lee, J. Y. Dai, J. Wang, H. L. W. Chan, C. L. Choy, Z. T. Song, and S. L. Feng, "Self-organized Ge nanocrystals embedded in HfAlO fabricated by pulsed-laser deposition and application to floating gate memory", *Appl. Phys. Lett.*, vol. 86, pp. 013110-1-3, 2004.

[33] A. Uedono, K. Ikeuchi, K. Yamabea, T. Ohdaira, M. Muramatsu, R. Suzuki, A. S. Hamid, T. Chikyow, K. Torii, and K. Yamadaa, "Annealing properties of open volumes in $HfSiO_x$ and $HfAlO_x$ gate dielectrics studied using monoenergetic position beams", *J. Appl. Phys.*, vol. 98, pp. 023506-1-5, 2005.

[34] Y. E. Hong, Y. S. Kim, K. Do, D. Lee, D. H. Koa, J. H. Ku, and H. Kim, "Thermal stability of Al- and Zr-doped HfO_2 thin films grown by direct current magnetron sputtering", *J. Vac. Sci. Technol. A*, vol. 23, no. 5, pp. 1413–1418, 2005.

[35] M.-H. Cho, D. W. Moon, S. A. Park, Y. K. Kim, K. Jeong, S. K. Kang, D.-H. Ko, S. J. Doh, J. H. Lee, and N. I. Lee, "Interfacial characteristics of N-incorporated HfAlO high-k thin films", *Appl. Phys. Lett.*, vol. 84, no. 25, pp. 5243–5245, 2004.

[36] K. Torii, R. Mitsuhashi, H. Ohji, T. Kawahara, and H. Kitajima, "Nitrogen profile engineering in the interfacial SiON in a HfAlO/SiON gate dielectric by NO re-oxidation", *IEEE Trans. Electron Dev.*, vol. 53, no. 2, pp. 323–328, 2006.

[37] G. D. Wilk and R. M. Wallace, "Electrical properties of hafnium silicate gate dielectrics deposited directly on silicon", *Appl. Phys. Lett.*, vol. 74, no. 19, pp. 2854–2856, 1999.

[38] M. Ritala, K. Kukli, A. Rahtu, P. I. Räisänen, M. Leskelä, T. Sajavaara, and J. Keinonen, "Atomic layer deposition of oxide thin films with metal alkoxides as oxygen sources", *Science*, vol. 288, no. 5464, pp. 319–321, 2006.

[39] Z. M. Rittersma, E. Naburgh, T. Dao, A. H. C. Hendriks, W. F. A. Besling, E. Tois, E. Vainonen-Ahlgren, M. Tuominen, and S. Haukka, "Physical and electrical properties of Zr-silicate dielectric layers deposited by atomic layer deposition", *Electrochem. Solid-State Lett.*, vol. 6, no. 7, pp. F21–F23, 2003.

[40] M. Lemberger, A. Paskaleva, S. Zürcher, A. J. Bauer, L. Frey, and H. Ryssel, "Electrical properties of hafnium silicate films obtained from a single-source MOCVD precursor", *Microelectron. Reliab.*, vol. 45, no. 5-6, pp. 819–822, 2005.

[41] H.-J. Cho, H. L. Lee, H. B. Park, T. S. Jeon, S. G. Park, B. J. Jin, S. B. Kang, Y. G. Shin, U.-I. Chung, and J. T. Moon, "Effects of post-deposition annealing on the electrical properties of HfSiO films grown by atomic layer deposition", *Jpn. J. Appl. Phys.*, vol. 44, no. 4B, pp. 2230–2234, 2005.

[42] M. Balog, M. Schieber, S. Patai, and M. Michman, "Thin films of metal oxides on silicon by chemical vapor deposition with organometallic compounds", *J. Cryst. Growth*, vol. 17, pp. 298–301, 1972.

[43] P. J. Harrop and D. S. Campbell, "Selection of thin film capacitor dielectrics", *Thin Solid Films*, vol. 2, no. 4, pp. 273–292, 1968.

[44] H. Takeuchi and T.-J. King, "Process optimization and integration of HfO_2 and Hf-silicates", in *Proc. Mater. Res. Soc. Symp.*, San Francisco, USA, 2004, vol. 811, pp. D7.6.1–D7.6.12.

[45] M.-H. Cho, K. B. Chung, C. N. Whang, D. W. Lee, and D.-H. Ko, "Phase separation and electronic structure of Hf-silicate film as a function of composition", *Appl. Phys. Lett.*, vol. 87, pp. 242906-1-3, 2005.

[46] J. Robertson, "High dielectric constant gate oxides for metal oxide Si transistors", *Rep. Progr. Phys.*, vol. 69, pp. 327–396, 2006.

[47] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "Hafnium and zirconium silicates for advanced gate dielectrics", *J. Appl. Phys.*, vol. 87, no. 1, pp. 484–492, 2000.

[48] M. Koyama, A. Kaneko, T. Ino, M. Koike, Y. Kamata, R. Iijima, Y. Kamimuta, A. Takashima, M. Suzuki, C. Hongo, and S. Inumiya, "Effects of nitrogen in HfSiON gate dielectric on the electrical and thermal characteristics", in *Int. Electron Dev. Meet. Tech, Dig.*, San Francisco, USA, 2002, pp. 849–852.

[49] M. A. Quevedo-Lopez, M. R. Visokay, J. J. Chambers, M. J. Bevan, A. LiFatou, L. Colombo, M. J. Kim, B. E. Gnade, and R. M. Wallace, "Dopant penetration studies through Hf silicate", *J. Appl. Phys.*, vol. 97, pp. 043508-1-15, 2005.

[50] K. B. Chung, C. N. Whang, M.-H. Cho, C. J. Yim, and D.-H. Ko, "Suppression of phase separation in Hf-silicate films using NH_3 annealing treatment", *Appl. Phys. Lett.*, vol. 88, pp. 081903-1-3, 2006.

[51] I. Z. Mitrovic, O. Buiu, S. Hall, C. Bungey, T. Wagner, W. Davey, and Y. Lu, "Electrical and structural properties of hafnium silicate thin films", *Microelectron. Reliab.*, vol. 47, pp. 645–648, 2007.

[52] B. R. Nag, "Empirical formula for the dielectric constant of cubic semiconductors", *Appl. Phys. Lett.*, vol. 65, no. 15, pp. 1938–1939, 1994.

[53] T. Busani and R. A. B. Devine, "The importance of network structure in high-*k* dielectrics: $LaAlO_3$, Pr_2O_3, and Ta_2O_5", *J. Appl. Phys.*, vol. 98, pp. 044102-1-5, 2005.

[54] D. Xue, K. Betzler, and H. Hesse, "Dielectric constants of binary rare-earth compounds", *J. Phys. Cond. Matt.*, vol. 12, no. 13, pp. 3113–3118, 2000.

[55] H. Iwai, S. Ohmi, S. Akama, C. Ohshima, A. Kikuchi, I. Kashiwagi, J. Taguchi, H. Yamamoto, J. Tonotani, and Y. Yoshihara, "Advanced gate dielectric materials for sub-100 nm CMOS", in *Int. Electron Dev. Meet. Tech. Dig.*, San Francisco, USA, 2002, pp. 625–628.

[56] J. Kwo, M. Hong, A. R. Kortan, K. L. Queeney, Y. J. Chabal, R. L. Opila, D. A. Muller, S. N. G. Chu, B. J. Sapjeta, T. S. Lay, J. P. Mannaerts, T. Boone, H. W. Krautter, J. J. Krajewski, A. M. Sergnt, and J. M. Rosamilia, "Properties of high *k* gate dielectrics Gd_2O_3 and Y_2O_3 for Si", *J. Appl. Phys.*, vol. 89, no. 7, pp. 3920–3927, 2001.

[57] S. Ohmi, S. Akama, A. Kikuchi, I. Kashiwagi, C. Ohshima, K. Sato, K. Oshima, and H. Iwai, "Rare earth oxide gate thin films prepared by E-beam deposition", in *Int. Worksh. Gate Insul. IWGI 2001*, Tokyo, Japan, 2001, pp. 200–204.

Stephen Hall received the Ph.D. degree from the University of Liverpool, UK, in 1987, for work on a new form of integrated injection logic in the GaAs/AlGaAs materials system. He then took up a lecturing post and undertook work on SOI materials characterization and device physics in the SIMOX and oxidised porous Si systems. The work was in collaboration with UK university and industrial collaborators. Other areas of activity concerned fabrication and electrical assessment of cobalt disilicide Schottky diodes together with silicon-germanium materials characterization and device physics for both SOI HBT bipolar transistor and MOSFET. His current major activities concern novel design concepts for vertical MOSFETs, high-*k* dielectrics and novel device and circuit blocks for third generation neural network application. He is currently Head of the Department of Electrical Engineering and Electronics and serves on a number of UK national advisory groups.

e-mail: s.hall@liverpool.ac.uk
Department of Electrical Engineering and Electronics
Brownlow Hill, University of Liverpool
Liverpool L69 3GJ, UK

Octavian Buiu received the B.Sc. and M.Sc. degrees in physics from the University of Bucharest, Romania, in 1985 and 1987, respectively. In 1997, he received the Ph.D. degree (*magna cum laude*) in atomic and molecular physics from Babes-Bolyai University of Cluj, Romania. He is a Lecturer in the Solid State Electronics Research Group, Department of Electrical Engineering and Electronics, University of Liverpool, UK. Between 1987 and 1997 he was a Research Scientist and Senior Scientist at the Research Institute for Electronic Devices and the National Research and Development Institute for Microtechnology, Bucharest. Between 1997 and 2000, he was a Research Fellow in the UK, working on degradation mechanisms in submicron LDD MOSFETs and deposition methods for dielectric and semiconductor layers. His research interests include optical and electrical properties of thin films (porous silicon, high-*k* dielectrics, SiO_2, SiON, SiGe and SiGeC) and their associated interfaces, electronic circuits and microsensors for biomedical applications, reliability tests on MOSFETs devices. He has published more than 35 papers on these topics.

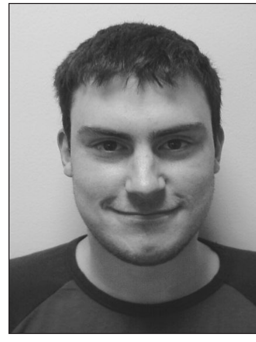
e-mail: O.Buiu@liverpool.ac.uk
Department of Electrical Engineering and Electronics
Brownlow Hill, University of Liverpool
Liverpool L69 3GJ, UK

Ivona Z. Mitrovic received the B.Sc. E.E. degree in microelectronics in 1997 from the Faculty of Electronic Engineering, University of Nis, Serbia, and the M.Sc. degree in materials science in 2002 from the University of Belgrade, Serbia. She is currently pursuing the Ph.D. degree at the Department of Electrical Engineering and Electronics, University of Liverpool, UK, with a thesis in the area of SiGeC HBTs on insulator for radio frequency communications. Her research work in the period 1997–2001 was part of a project of the Serbian Academy of Sciences and Arts, and included the study of dielectrics, in particular $BaTiO_3$-ceramics. Since 2001, she has been a Research Assistant at the Department of Electrical Engineering and Electronics, University of Liverpool. Her research interests include materials and electrical characterization study of SiGe, SiGeC and high-*k* dielectrics, as well as device modeling and characterization (heterojunction bipolar transistors).

e-mail: ivona@liverpool.ac.uk
Department of Electrical Engineering and Electronics
Brownlow Hill, University of Liverpool
Liverpool L69 3GJ, UK

Yi Lu received the B.Sc. degree in electrical engineering from the North China Electrical Power University, Baoding, China, in 2001, and the M.Sc. degree in intelligence engineering from the University of Liverpool, UK, in 2002. He is currently pursuing his Ph.D. degree at the Department of Electrical Engineering and Electronics, the University of Liverpool. His research interests include optical and electrical characterization of Hf-based high-*k* dielectrics, as well as MOS device modeling.

e-mail: luyi@liverpool.ac.uk
Department of Electrical Engineering and Electronics
Brownlow Hill, University of Liverpool
Liverpool L69 3GJ, UK



William Mark Davey received the M.E. degree in electronics from the University of Liverpool, UK, in 2005. He is currently studying for a Ph.D. at the Department of Electrical Engineering and Electronics, the University of Liverpool. His current research interests are the physical, electrical and optical characterization of hafnium related compounds and lanthanide series oxides for advanced CMOS application.

e-mail: W.Davey@liverpool.ac.uk
Department of Electrical Engineering
and Electronics
Brownlow Hill, University of Liverpool
Liverpool L69 3GJ, UK

Invited paper

Semiconductor cleaning technology for next generation material systems

Jerzy Ruzyllo

Abstract— This paper gives a brief overview of the challenges wafer cleaning technology is facing in the light of advanced silicon technology moving in the direction of non-planar device structures and the need for modified cleans for semiconductors other than silicon. In the former case, the key issue is related to cleaning and conditioning of vertical surfaces in next generation CMOS gate structure as well as deep 3D geometries in MEMS devices. In the latter, an accelerated pace at which semiconductors other than silicon are being introduced into the mainstream manufacturing calls for the development of material specific wafer cleaning technologies. Examples of the problems related to each challenge are considered.

Keywords— III-V compounds, FinFET, IC manufacturing, MEMS, MOS gate stack, semiconductor cleaning.

1. Introduction

Wafer cleaning is the most frequently applied processing step in high-end silicon IC manufacturing. As such, chemistry and implementation of Si cleaning operations are very well established and backed by many years of extensive research, as well as significant industrial tool base. As a result, silicon cleaning technology is by far the most mature among all semiconductors of any practical importance. The first complete, based on scientific considerations cleaning recipe specifically designed to clear Si surface from particles, metallic, and organic contaminants was proposed in 1970 [1]. Since then, silicon cleaning technology was undergoing continuous evolutionary modifications. Surprisingly, state-of-the-art Si cleaning still relies on roughly the same set of chemical solutions, but the way they are prepared and delivered to the wafer is very different from the one proposed originally. In addition, selected surface cleaning/conditioning functions that were traditionally performed by wet cleaning chemistries are now carried in the gas-phase [2].

What is important to the point being made in this paper, however, is that as advanced as they currently are, silicon cleaning methods cannot meet all the diversified emerging needs of semiconductor technology across the spectrum of materials and device structures both in terms of implementation methods and chemistries. Two key challenges are reviewed in this paper. First challenge is related to the growing importance of non-planar silicon devices such as next generation MOS gate stacks, micro-electro-mechanical system (MEMS) devices, and nanowires. Cleaning operations implemented in the traditional way may not be entirely effective in these cases. Second challenge results from the increasingly broad use in practical applications of semiconductor materials other than silicon. The re-emergence of germanium (Ge) as a possible replacement for silicon in selected applications, growing importance of IV-IV compounds (SiGe, SiC), and inevitable continued growth of technology of III-V compounds such as GaAs, GaN, and InSb for instance, underscores this trend.

2. Non-planar silicon based devices

The issue of non-planarity of silicon surfaces in device fabrication is likely to challenge standard wafer cleaning technology on at least three different fronts.

2.1. Next generation CMOS technology

A challenge at hand in cutting edge digital CMOS technology is to maintain adequate capacitance density of gate structure needed to sustain high drive current. One approach is to use gate dielectrics featuring dielectric constant higher than that of SiO_2, while the other is to increase gate area without increasing the area of the cell by structuring MOS gate 3-dimensionally. As the latter approach appears to be the one that will provide better long-term solutions, the interest in processing 3D MOS gate structures is growing. Regardless of what specific configuration will become a standard, all will involve pre-gate oxidation cleaning and conditioning of post-RIE vertical walls engraved in silicon. For example, one possible solution considered involves formation of the thin "fin"-like strip of Si and forming an MOS gate structure around it as shown in Fig. 1. Starting with silicon-on-insulator (SOI) substrate (Fig. 1a), the fin as shown in Fig. 1b is formed by reactive ion etching (RIE).

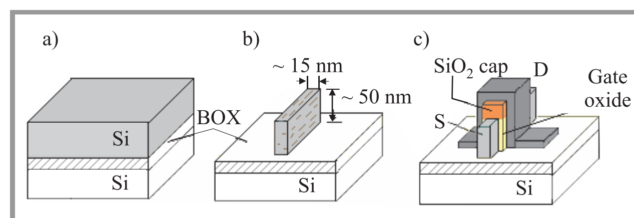


Fig. 1. The SOI substrate (a) in which a "fin" is defined by RIE (b), and then MOS gate is built around it (c).

Surrounded at both sides by the gate oxide and gate contact such a "fin" will eventually become a channel in the FinFET structure (Fig. 1c).
In the processing of the vertical surfaces of the "fin" in FinFETs [3] (Fig. 1b), or in other structures with working sidewalls [4], or in the processing of U-shaped trenches in UMOSFETs [5], the challenge is to assure defect-free SiO_2-Si interface formed on the surfaces defined by the damaging RIE process. Furthermore, a drastic departure from surface flatness in next generation MOS gate structures creates obvious problems in particle removal.

2.2. MEMS technology

Due to the advantageous mechanical properties of Si increasingly complex micro-electro-mechanical systems are possible. Specific feature of MEMS manufacturing is that it includes deep etching of elaborate 3D features, as well as demanding release processes such as those shown in Fig. 2 where buried oxide (BOX) in SOI wafer is deep etched laterally. Removal of possible etch residues from such extremely confined geometrical features and assuring stiction-free operation of beams and membranes cannot be accomplished using conventional wet cleaning and etching technology. The anhydrous HF/methanol (AHF/MeOH) [6] sacrificial oxide etch process has been investigated as a viable solution to the latter [7]. Solution to the MEMS cleaning problems comes in the form of a supercritical fluid cleaning technology [8]. Once in the supercritical state, a fluid features essentially no surface tension, and hence, features no limitations regarding geometries it can penetrate. Due to these characteristics, a supercritical cleaning technology becomes a standard in the processing of highly confined semiconductor structures. Figure 3 shows under what conditions in terms of temperature and pressure liquids and gases can be transformed into a state of supercritical fluid. The most common supercritical carrier of cleaning chemistries is CO_2 for which a critical point is at 31°C and 73 atm.

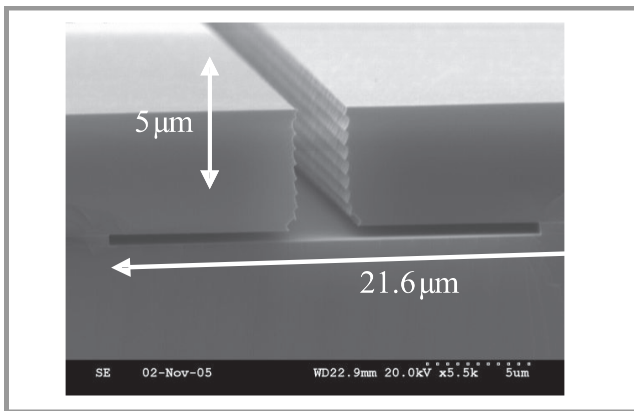


Fig. 2. SOI wafer with "buried oxide" etched laterally.

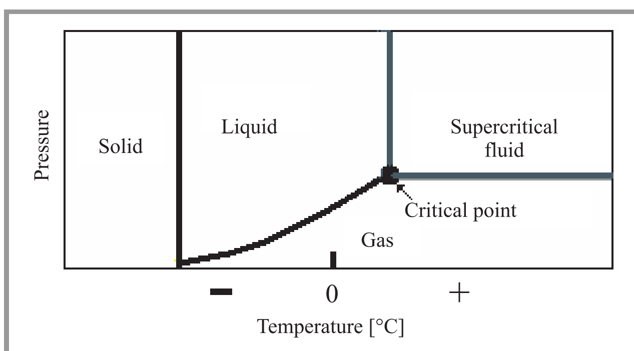


Fig. 3. Definition of the supercritical fluid phase.

2.3. Silicon nanowires

In the continued push toward faster and more efficient switching devices, silicon nano-geometry structures that depart from conventional planar technology, such as silicon nanowires, are aggressively pursued. Making functional devices out of nanowires requires subjecting them to a standard fabrication sequence. Figure 4 shows loosely scattered Si nanowires released after an anisotropic bottom-up growth process and a single wire mounted in between two metal contacts. Considering extreme fragility of nanowires and a size that makes their handling very difficult, the use of conventional fabrication methods is in this case severely restricted. In particular, those restrictions apply to wet cleaning operations which, due to the problems of nanowire han-

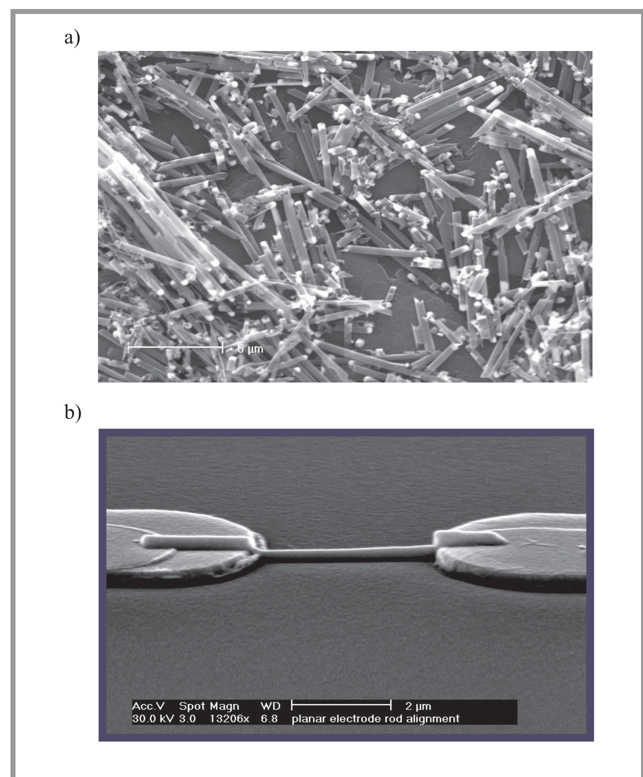


Fig. 4. (a) Silicon wires after growth and release and (b) single wire mounted between two contacts (*Courtesy Redwing Research Group, Penn State University*).

dling in liquids, are incompatible with nanowire processing. Also, anisotropic dry cleaning methods (derivatives of sputtering and RIE) are not suitable in this application. The isotropic gas-phase methods based on HF vapor, e.g., AHF/MeOH process [6, 9] or remote plasma [10] may offer workable solutions to the problem of nanowire cleaning and surface conditioning.

3. Semiconductors other than silicon

Due to its outstanding crystal quality, excellent oxidation characteristics, manufacturability, abundance, relatively low cost, and adequate electronic properties silicon was for the last 40 years, and will remain in the future, a dominant semiconductor used in device manufacturing. However, growing needs for improved performance in specific electronic (e.g., high-temperature, high-power, as well as ultra-high speed) and photonic (e.g., emission of blue light or UV detection) applications, require significantly improved manufacturing technology in the range of semiconductors other than silicon. Examples of such materials include germanium, Ge, due to its electron mobility higher than that of Si and prospects for integration with high-*k* gate dielectrics, silicon germanium, SiGe, needed to process strained-channel Si MOSFETs, as well as silicon carbide, SiC, for its wide energy gap. Also of growing interest are III-V semiconductors beyond the most advanced GaAs such as GaN for its wide, direct band gap, and indium antimonide, InSb, for its electron mobility of 80 000 cm^2/Vs to name just two. Figure 5 shows key characteristics of selected elemental and compound semiconductors [11].

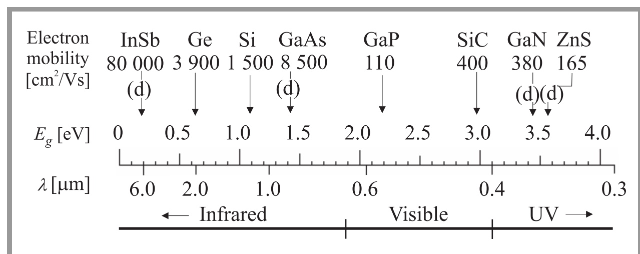


Fig. 5. Energy band gap, type of the energy gap, and cut-off wavelengths for various semiconductors.

Surface cleaning is becoming a growing issue in the processing of semiconductors other than silicon. This is because for the most part inferior quality of the substrate crystal rather than cleanliness of its surface was until recently a dominant factor defining manufacturing yield in those materials. With the improvements in the quality of single-crystal substrates of various semiconductors the paradigm is shifting and much closer attention to the cleaning technology is being paid.

In general, technology of surface cleaning of materials other than silicon attempts to draw from the pool of Si cleaning chemistries and to use high performance cleaning infrastructure developed over the last 40 years for silicon. While there is no significant barrier regarding the latter task, implementation of the former is not a straightforward matter. This is because due to the differences in the chemical compositions, various semiconductors not always respond in the desired fashion to the cleaning chemistries successfully used in silicon processing. To illustrate the nature of the problem we shall consider cleaning-related issues in the case of germanium, Ge, and silicon carbide, SiC.

Germanium is re-emerging as an alternative to silicon semiconductor in those applications in which outstanding characteristics of silicon's native oxide, SiO_2, are not coming into play and in which higher electron mobility of Ge may be beneficial. Specifically, Ge in conjunction with high-*k* gate dielectrics may offer advantages over the Si based MOS gates. However, processing of Ge-HfO_2 gate stacks for instance, requires surface termination prior to high-*k* deposition different than in the case of silicon. One approach is to Si-passivate germanium surface through an anneal in SiH_4. Also, plasma PH_3 treatment at 400°C given to Ge in situ prior to HfO_2 deposition was reported to improve the characteristics of both NMOS and PMOSFETs [12].

As far as standard cleaning operations such as native oxide etching, particle and metallic contaminant removal are concerned, the response of Ge surface to the Si cleaning chemistries varies depending on application. In the case of particle deposition and removal for instance, it was established that Ge surface acts in the same way as Si surface [13]. Situation is different in the case of Ge native oxide, GeO_2, which in contrast to Si native oxide SiO_2, cannot be removed entirely using HF-based chemistries [14, 15]. Furthermore, the metallic contaminant deposition and removal was shown to be driven in the case of Ge by somewhat different mechanisms than in the case of Si substrates. Most notably, in the case of Ge it does depend on the pH of solution, and, unlike in the case of Si, all common metallic contaminants can be removed from the Ge surface using HF:H_2O solution [16].

In contrast to elemental semiconductors such as Si and Ge, silicon carbide, SiC, represents a class of man-made binary semiconductor compounds in which each element features often drastically different chemical characteristics. In the case of SiC for instance, oxidized Si forms a solid SiO_2, while oxidized carbon forms gaseous compounds CO and CO_2. Hence, the response of compound semiconductors to cleaning chemistries may not be entirely isotropic. In spite of it, due to the fact that SiC is a chemical derivative of Si, cleaning chemistries used in Si processing are rather arbitrarily adopted to process SiC surfaces. It turns out that such an automatic transfer of cleaning technology from Si to SiC may not necessarily produce the desired results. To exemplify this point let us refer to the results of the experiments in which roughness of SiC surface exposed to various cleaning chemistries was monitored [17]. The results, summarized in Fig. 6, indicate sensitivity of SiC

surface roughness to various cleaning chemistries different both quantitatively and qualitatively than that of Si.

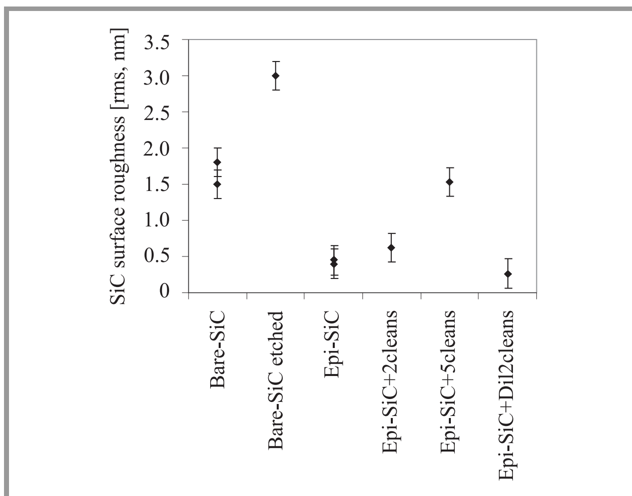


Fig. 6. Changes of SiC surface roughness as a result of various chemical surface treatments.

An analysis supported by experiments similar to the one given above for SiC should be carried out for other compound semiconductors, III-V in particular. The results are very likely to demonstrate that the needs regarding cleaning technology vary from material to material, and hence, for each semiconductor material of interest, the dedicated cleaning recipes should be developed.

4. Summary

The purpose of this overview was to demonstrate the challenges cleaning and surface conditioning technology is facing as on the one hand silicon technology goes non-planar and on the other, semiconductors other than silicon are being pursued more actively than ever before in a range of applications. The discussion presented leads to the following observations:

- silicon cleaning technology both in terms of chemistries as well as tools used is a foundation upon which any new developments responding to the emerging needs of semiconductor cleaning will be based;
- silicon cleaning chemistries are not always compatible with all semiconductors that may be of the commercial importance, and hence, dedicated cleaning technology must be investigated and developed for each of them;
- innovative solutions are needed to cope with surface cleaning and conditioning needs in emerging non-planar device manufacturing such as 3D MOS gates, MEMS, nanowires, and nanotubes.

References

[1] W. Kern and D. Puotinen, "Cleaning solution based on hydrogen peroxide for use in semiconductor technology", *RCA Rev.*, vol. 31, pp. 187–206, 1970.

[2] J. Ruzyllo, "Second look at role of dry water cleaning technology", in *Proc. Symp. Contam. Free Manuf. SEMICON West 1998*, San Francisco, USA, 1998, pp. L1–L6.

[3] K. Endo, S. Noda, M. Masahara, T. Kubota, T. Ozaki, S. Samukawa, Y. Liu, K. Ishii, Y. Ishikawa, E. Sugimata, T. Matsukawa, H. Takashima, H. Yamauchi, and E. Suzuki, "Fabrication of Fin-FETs by damage-free neutral-beam etching technology", *IEEE Trans. Electron Dev.*, vol. 53, pp. 1826–1833, 2006.

[4] J. P. Campbell, "Analysis of sidewall cleans relating to titanium salicide filaments", *IEEE Trans. Semicond. Manuf.*, vol. 17, pp. 603–611, 2004.

[5] C. T. Wu, R. S. Ridley Sr., G. Dolny, T. Grebs, C. Knoedler, S. Suliman, B. Venkataraman, O. Awadelkarim, and J. Ruzyllo, "Growth and reliability of thick gate oxide in U-trench for power MOSFET's", in *Proc. 14th Int. Symp. Pow. Semicond. Dev. ICs*, Osaka, Japan, 2002, pp. 149–153.

[6] J. Ruzyllo, K. Torek, C. Daffron, R. Grant, and R. Novak, "Etching of thermal oxides in low-pressure anhydrous HF/CH_3OH gas mixture at elevated temperature", *J. Electrochem. Soc.*, vol. 140, pp. L64–L67, 1993.

[7] M. Erdamar, "Studies of MEMS release processes using anhydrous HF/methanol mixtiure". M.S. thesis, Penn State University, 2005.

[8] R. B. Farmer, B. D. Jones, K. P. Gupta, I. H. Jafri, and D. M. Dispensa, "Supercritical phase water drying/cleaning system", US Patent 6,067,728, May 30, 2000.

[9] K. Torek, J. Ruzyllo, R. Grant, and R. Novak, "Reduced pressure etching of thermal oxides in anhydrous HF/alcoholic gas mixtures", *J. Electrochem. Soc.*, vol. 142, pp. 1322–1328, 1995.

[10] J. Ruzyllo, D. Frystak, and R. Bowling, "Dry cleaning procedure for silicon IC fabrication", in *IEEE Int. Electron. Dev. Meet. IEDM Tech. Dig.*, Washington, USA, 1990, pp. 409–412.

[11] N. Wu, Q. Zhang, D. S. Chan, N. Balasubramanian, and C. Zhu, "Gate-first germanium nMOSFET with CVD HfO_2 gate dielectric and silicon surface passivation", *IEEE Electron Dev. Lett.*, vol. 27, pp. 479–481, 2006.

[12] S. J. Whang, S. J. Lee, F. Gao, N. Wu, C. X. Zhu, J. S. Pan, L. J. Tang, and D. L. Kwong, "Germanium p- and n-MOSFETs fabricated with novel surface passivation (plasma-PH_3 and thin AlN) and TaN/HfO_2 gate stack", in *IEEE Int. Electron Dev. Meet. IEDM Tech. Dig.*, San Francisco, USA, 2004, pp. 307–310.

[13] S. Sioncke, M. Lux, W. Fyen, M. Meuris, P. Mertens, and A. Theuwis, "Particle deposition and removal from Ge surfaces", in *8th Int. Symp. Ultra Clean Proces. Semicond. Surf. UCPSS 2006*, Antwerp, Belgium, 2006.

[14] J. Kim, K. Saraswat, and Y. Nishi, "Study of germanium surface in wet chemical solutions for surface cleaning applications", *ECS Trans.*, vol. 1, pp. 214–219, 2005.

[15] J. Kim, K. McVittie, K. Saraswat, and Y. Nishii, "Passivation studies of germanium surface", in *8th Int. Symp. Ultra Clean Proces. Semicond. Surf. UCPSS 2006*, Antwerp, Belgium, 2006.

[16] S. Sioncke, B. Onsia, K. Struys, J. Rip, R. Vos, M. Meuris, P. Mertens, and A. Theuwis, "Metal deposition on Ge surfaces", *ECS Trans.*, vol. 1, pp. 220–227, 2005.

[17] K. Chang, T. Witt, A. Hoff, R. Woodin, R. Ridley, G. Dolny, K. Shanmugasundaram, E. Oborina, and J. Ruzyllo, "Surface roughness in silicon carbide technology", *ECS Trans.*, vol. 1, pp. 228–233, 2005.

Jerzy Ruzyllo is a Professor of electrical engineering and materials science and engineering at Penn State University in the United States and holds a title of Professor in Poland. He received his Ph.D. degree in 1977 from the Warsaw University of Technology, Poland, where he also served on the faculty before joining Penn State. His experience and professional interests are in research and teaching in the area of semiconductor materials and device processing. He is a Fellow of IEEE and a Fellow of the Electrochemical Society.

e-mail: jxr6@psu.edu
Department of Electrical Engineering
Penn State University
University Park, PA 16802, USA

Invited paper

Development of 3C-SiC MOSFETs

Mietek Bakowski, Adolf Schöner, Per Ericsson, Helena Strömberg, Hiroyuki Nagasawa, and Masayuki Abe

Abstract— The paper reviews the development of the 3C-SiC MOSFETs in a unique development project combining the material and device expertise of HAST (Hoya Advanced Semiconductor Technologies) and Acreo, respectively. The motivation for the development of the 3C-SiC MOSFETs and the summary of the results from the lateral and vertical devices with varying size from single cell to 3×3 mm^2 large devices are reviewed. The vertical devices had hexagonal and square unit cell designs with 2 μm and 4 μm channel length. The p-body was aluminum implanted and the source was nitrogen or phosphorus implanted. Low temperature Ti/W contacts were evaluated.

Keywords— vertical MOSFET, 3C-SiC, channel mobility.

1. Introduction

The 3C-SiC is a cubic polytype of SiC and can be grown on Si. This means that large 3C-SiC crystals equal in diameter to the commercially available Si wafers can be readily made. Processing of large area wafers greatly improves cost per manufactured device. The 6” diameter 3C-SiC wafers have been demonstrated [1].

The 3C-SiC is a promising material for MOSFET devices because of high channel mobility due to lower density of interface states compared to 4H-SiC. Poor performance of 6H- and 4H-SiC MOSFETs is related to the interface states located in the band gap close to the conduction band edge limiting the transport of electrons in the channel [2]. Due to the smaller band gap of 3C-SiC, the interface states observed in 6H- and 4H-SiC are located in the conduction band and have no effect on the transport properties of the channel. Channel mobility values of 75 to 260 cm^2/Vs have been reported [3–6].

The 3C-SiC polytype has lower critical electric field value due to the lower band gap. It means that the drift region doping corresponding to a given blocking voltage will be lower compared to the hexagonal 4H- and 6H-SiC polytypes. This also means that the specific junction capacitance will be lower in the 3C-SiC devices as compared to the 4H- and 6H-SiC ones. This is an advantage from the point of view of high speed MOSFETs [7].

Considering all of the above and given large-area substrates of good quality, 3C-SiC may well be the material of choice for medium voltage (600 V to 1200 V) MOSFETs.

We have reported earlier on large area lateral 3C-SiC MOSFETs [8] and on vertical large area 3C-SiC MOSFETs with varying size from a single unit cell to 3×3 mm^2 [9]. We have also reported earlier on the impact of technology on the characteristics of vertical 3C-SiC implanted MOSFETs (DMOSFETs). The MOSFET devices investigated here have a single implanted p-body profile. They will be, however, referred to as DMOSFETs since they represent the most simple and close approximation of double-diffused Si MOSFET concept in SiC technology. The exact translation of the DMOSFET concept into SiC technology is sometimes called DIMOSFET (double implanted MOSFET). Devices with phosphorus and nitrogen implanted source were compared. The impact of low temperature Ti/W contacts on the device characteristics was evaluated and compared to Ni-silicide contacts annealed at 950°C [10]. In this review we include some of the most recent results of voltage blocking characteristics and the temperature dependence of the channel mobility.

2. Lateral MOSFETs

2.1. Experimental

The 2” 3C-SiC $\langle 001 \rangle$ substrates were manufactured by HAST (Hoya Advanced Semiconductor Technologies). A 2 μm thick p-type epi layer with an Al-doping of 10^{16} cm^{-3} was grown at Acreo. For epitaxial LT MOSFETs (lateral trench MOSFETs), an additional n+ source and drain layer with nitrogen doping concentration of 10^{19} cm^{-3} and thickness of 0.3 μm was grown. The LDD MOSFETs (lightly-doped-drain MOSFETs) were implanted with nitrogen at 500°C in order to create two box profiles with the doping of 10^{18} cm^{-3} and 10^{20} cm^{-3} in the low doped drain region and source and drain contact regions, respectively. The gate oxide was grown thermally for 90 min at 1100°C in dry oxygen followed by a 3-hour post-oxidation anneal in wet oxygen at 950°C. The resulting oxide thickness was about 60 nm. The devices have two level metallization with oxide/nitride isolated bridges between gate and source interconnections. The 1×1 mm^2, 2×2 mm^2 and 3×3 mm^2 devices contain 220, 880 and 1980 unit cells, respectively. A post-processing annealing was done for 30 min at 400°C in nitrogen.

2.2. Results

Typical output characteristics of the lateral LDD MOS devices are shown in Fig. 1. Both the drain current and the leakage current scale linearly with the device size up to the maximum investigated device size of 3×3 mm^2, as shown in Fig. 2. We could obtain a current of about 0.3 A and 1.2 A with the current flow in the $\langle 110 \rangle$ and $\langle \bar{1}10 \rangle$ direction, respectively, from the largest 3×3 mm^2 devices with gate voltage of 20 V.

Lateral 3C-SiC MOSFET devices were fabricated with the area up to 3×3 mm^2. The LDD MOSFET devices

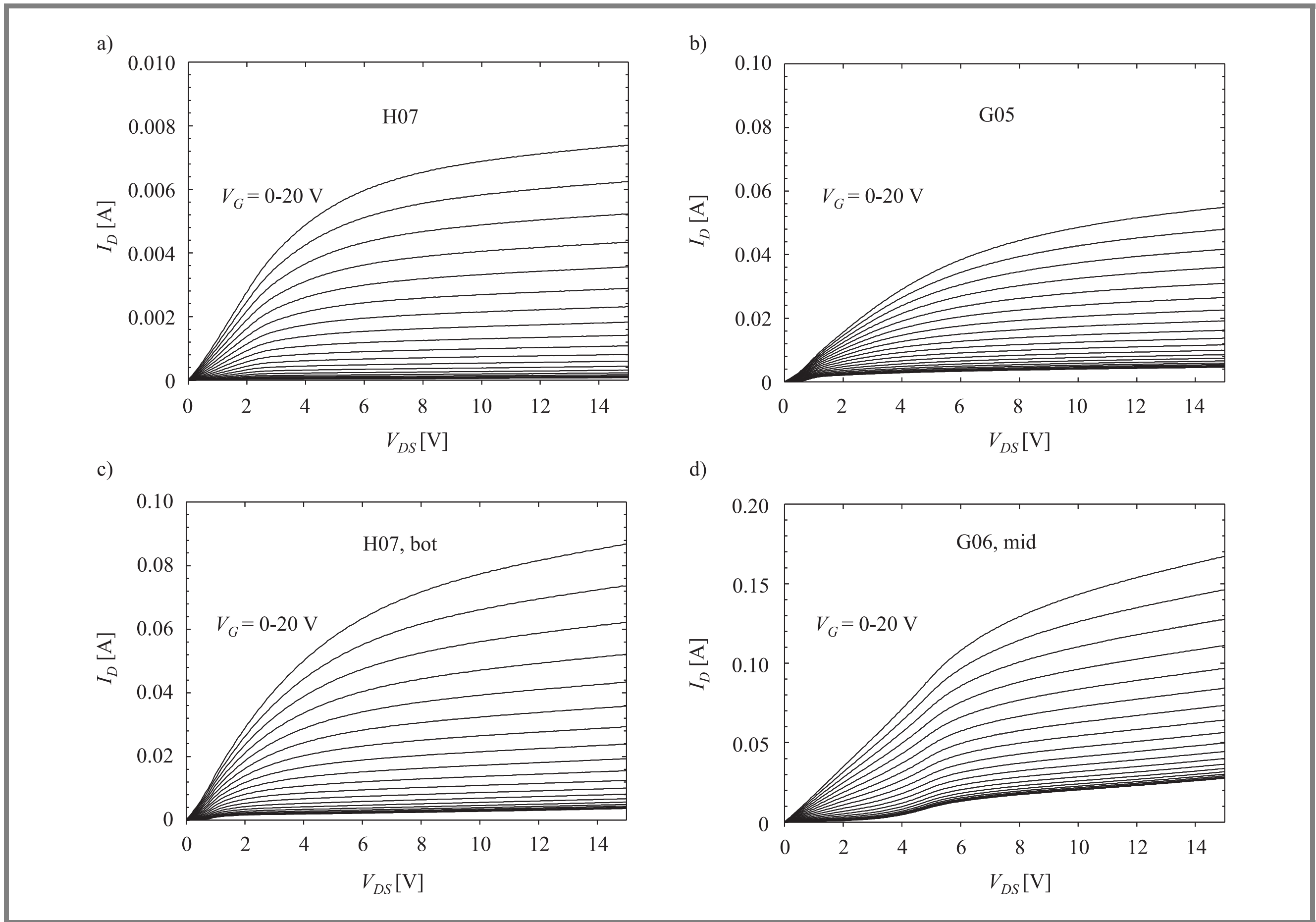


Fig. 1. Output characteristics of large-area LDD MOS devices containing 12×5 (a); 22×10 (b); 44×10 (c); 66×10 (d) cells; 4 μm, 400°C, corresponding to 0.5×0.5 mm^2, 1×1 mm^2, 2×1 mm^2 and 3×1 mm^2 device area, respectively. Only data with the current flow along the $\langle 110 \rangle$ direction are shown. Note leakage current due to the implanted drain junction.

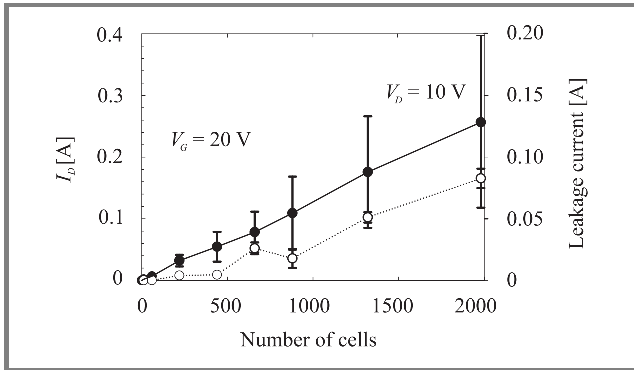


Fig. 2. Drain current (solid line) and leakage current (dotted line) as a function of the number of unit cells for LDD MOS devices with 4 μm channel length and current flow along the $\langle 110 \rangle$ direction.

have blocking capability of 100 V and channel mobility 2–3 times higher compared to the average 4H-SiC devices with current flow along the $\langle \bar{1}10 \rangle$ direction [1]. Channel mobility of devices with current flowing along the $\langle 110 \rangle$ direction is comparable to that of an average 4H-SiC device. The properties of the fabricated MOSFETs, of both LT- and LDD-type are dominated by high density of interface states of the order of 10^{13} cm^{-2}eV^{-1}. This is at

Table 1
Summary of lateral MOSFET parameters

Parameter	LT MOS (w3) After 400°C anneal $\langle 110 \rangle / \langle \bar{1}10 \rangle$	LDD MOS (w4) After 400°C anneal $\langle 110 \rangle / \langle \bar{1}10 \rangle$
V_{Th} [V]	5/7.5	3/3–5
μ_{eff} [cm^2/Vs]	8–10/2–6 ($V_G = 25$ V)	2–4/5–8 ($V_G = 20$ V)
μ_{FE} [cm^2/Vs]	15–20/6–15 ($V_G = 25$ V)	2–6/10–15 ($V_G = 20$ V)
D_{it} [cm^{-2}/eV^{-1}]	$2 \cdot 10^{13}/3 \cdot 10^{13}$	$2{-}4 \cdot 10^{13}/3{-}6 \cdot 10^{13}$
Note the dependence of the mobility on the direction of current flow and high density of interface states, D_{it}, obtained from the subthreshold slope.		

present the main reason for the poor performance of 3C-SiC devices compared to what was reported in [5]. The linear scaling of the device performance with the device area confirms the potential of the 3C-SiC material for fabrication of large-area devices. A correlation of high leakage current and high number of extended crystal defects (mainly stacking faults) has been observed. The results are summarised in Table 1.

3. Vertical MOSFETs

3.1. Experimental

The 10 μm thick, $5 \cdot 10^{15}$ cm^{-3} doped n-type epilayers were grown on 2” 3C-SiC $\langle 001 \rangle$ substrates manufactured by HAST [1]. Box profile with the depth of approximately 1 μm and maximum doping of 10^{18} cm^{-3} obtained by means of aluminum implantation combining five energies from 30 to 700 keV was used to define the p-body region and the ring termination. The n+ source region is defined by a box profile obtained by means of either nitrogen implantation combining three energies of 50, 90 and 150 keV or phosphorus implantation combining three energies of 70, 120 and 200 keV. The depth of this box profile is 0.4 μm and maximum doping is $4 \cdot 10^{19}$ cm^{-3}. All implantation processes were performed at room temperature. The gate oxide was grown thermally for 90 min at 1100°C in dry oxygen followed by a 3-hour post-oxidation anneal in wet oxygen at 950° C . The resulting oxide thickness was about 60 nm. All wafers underwent a shallow nitrogen implantation (30 keV, $5 \cdot 10^{12}$ cm^{-2}) in the gate oxide region prior to the thermal oxidation, which has been reported to reduce the interface-state density D_{it} [11]. The fabricated MOSFETs have two level metallization with oxide/nitride isolation between gate and source interconnections. Devices with the area of 1 mm^2 contain up to 976 hexagonal and 660 square unit cells, while those with the area of 3×3 mm^2 contain up to 12000 hexagonal and 8000 square unit cells.

3.2. Results

An example of the output characteristics and device lay-out is shown in Fig. 3.

The difference in performance between devices with phosphorus (P) and nitrogen (N) implanted source is illustrated in Fig. 4, where a comparison between P and N doped 400×400 μm^2 MOSFETs containing 102 hexagonal cells and ring termination (active area $7.3 \cdot 10^{-4}$ cm^2), is shown. A comparison of channel mobility extracted from the output characteristics of devices with N doped source and two metallization technologies, nickel silicide and low temperature Ti/W ohmic contacts, is shown in Fig. 5. It is clearly seen that the channel mobility is degraded in the case of nickel silicide contacts due to the 950°C silicidation step performed after gate oxide formation (Fig. 5c).

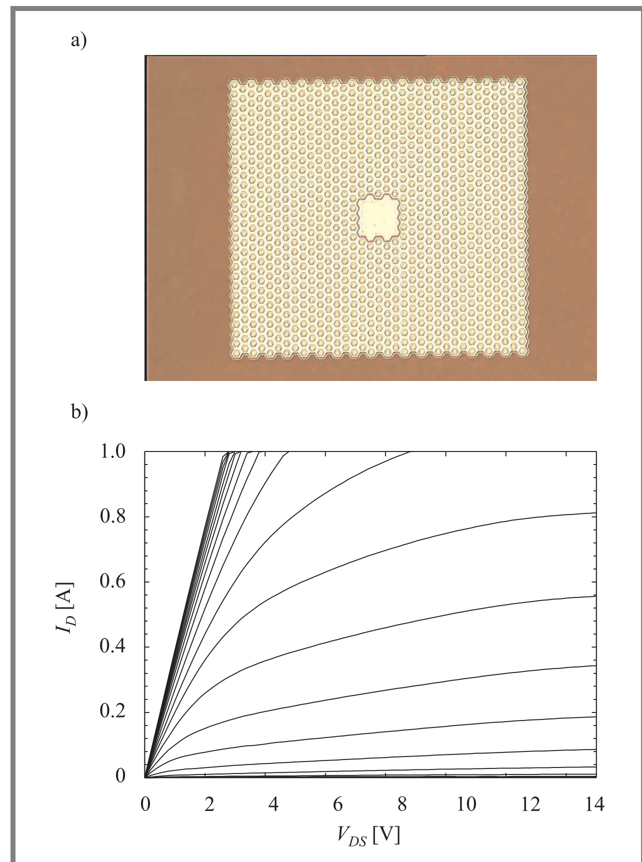


Fig. 3. Micrograph (a) and output characteristics of a 1×1 mm^2 3C-SiC MOSFET containing 976 hexagonal cells with 2 μm channel length and P implanted source (b). The cell pitch for this device is 28 μm. High temperature steps following the thermal oxidation were avoided and the ohmic contact to the source was made with as-deposited Ti/W.

The channel mobility values are improved by reducing the thermal budget of the processing steps following the gate oxidation (Fig. 5b). The channel mobility ranges from 30 to 40 cm^2/Vs and from 20 to 30 cm^2/Vs for P and N doped source devices, respectively.

Table 2
Comparison of vertical MOSFET parameters for devices with P and N implanted source

Parameter	P impl. source Ti/W no anneal	N impl. source Ti/W no anneal
V_{Th} [V]	–5 to 3	0
μ_{eff} [cm^2/Vs] μ_{FE} [cm^2/Vs]	30–40	20–30
D_{it} [cm^{-2}eV^{-1}]	$1 \cdot 10^{13}$	$7 \cdot 10^{12}$
Sub-threshold slope [mV/dec]	750	750
R_{on} [mΩcm^2] ($V_G = 15$ V)	17 ($L_{ch} = 2$ μm) 30 ($L_{ch} = 4$ μm)	24 ($L_{ch} = 2$ μm) 46 ($L_{ch} = 4$ μm)

The devices have high density of interface states of the order of 10^{13} $cm^{-2}eV^{-1}$ and sub-threshold slope of 750 mV per decade as illustrated in Fig. 6. In spite of high D_{it} values relatively high channel mobility was observed. From SiO_2/3C-SiC MOS interface studies it is expected that D_{it} values can be reduced by at least one order of magnitude, which again indicates the potential of 3C-SiC for high current MOSFET devices [12].

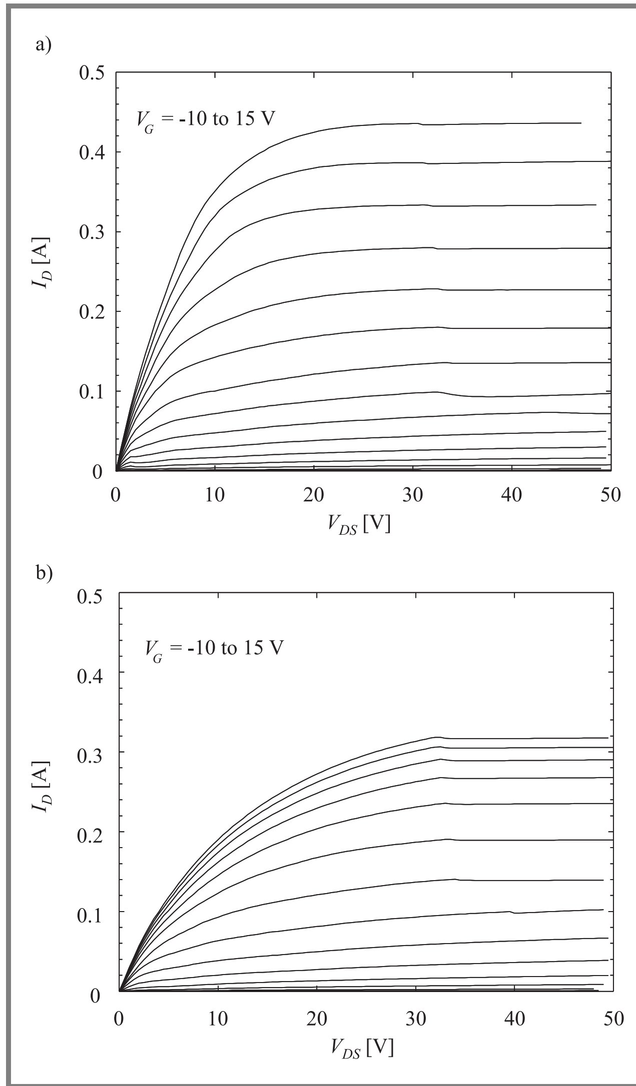


Fig. 4. Output characteristics of a device containing 102 hexagonal cells with 2 μm channel length and P implanted source (a) and with N implanted source (b).

The summary of the results obtained from MOSFET with P and N implanted source is given in Table 2.

Vertical 3C-SiC MOSFET devices were fabricated with the area up to 3×3 mm^2. The blocking capability of MOSFET devices was typically 100 V at leakage currents below 1 mA. The specific on-resistance obtained from 0.0025 cm^2 and 0.01 cm^2 devices with P doped source was 17 $m\Omega cm^2$ and 30 $m\Omega cm^2$ for the channel length of 2 μm (cell pitch 28 μm) and 4 μm (cell pitch 33 μm), respectively. For the devices with N doped source, the corresponding values of the specific on-resistance were 24 $m\Omega cm^2$ and 46 $m\Omega cm^2$.

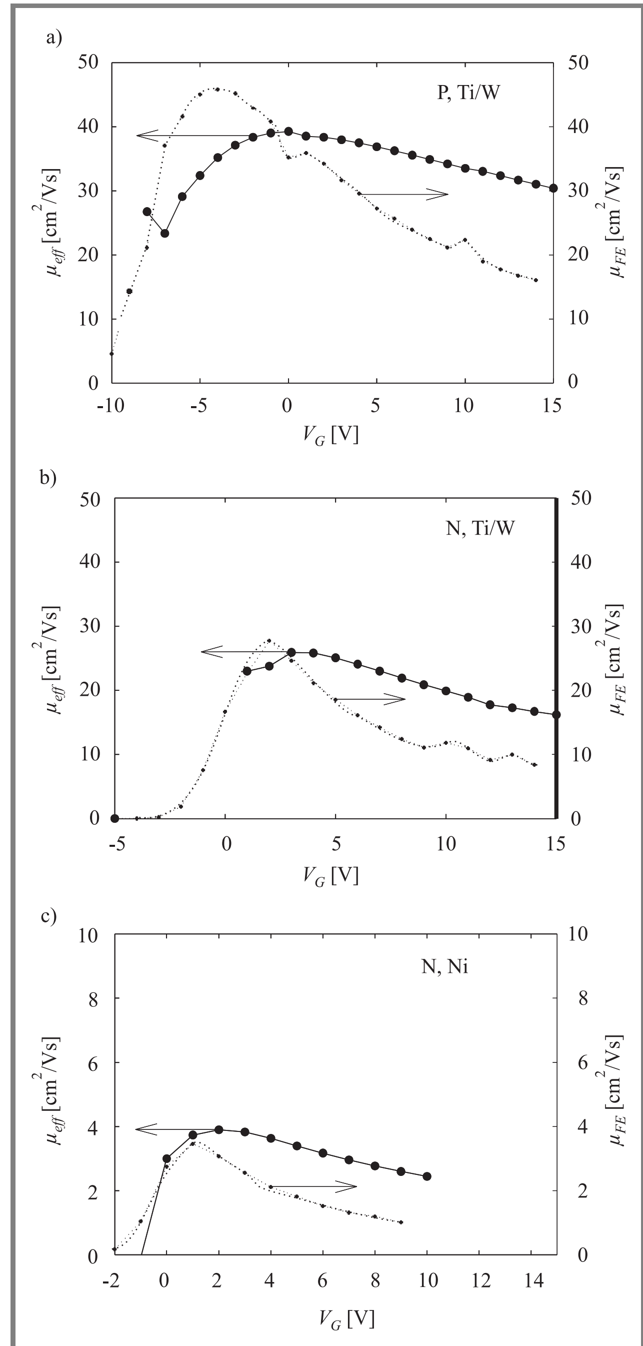


Fig. 5. Channel mobility extracted from the output characteristics of single hexagonal cell devices with P implanted source and Ti/W contact (a), with N implanted source and Ti/W contact (b) and with N implanted source and nickel silicide contact (c). Channel length $L_{ch} = 2$ μm.

The drain and the leakage currents scale linearly with the device size up to the maximum investigated size of 1×1 mm^2, as in the case of lateral devices. The linear scaling of the device performance with the device

area confirms the potential of the 3C-SiC material for the fabrication of large area MOSFET devices. The blocking capability deteriorates with increasing number of cells due to increasing leakage. Large differences in leakage current were observed between devices positioned in areas containing stacking faults and in stacking fault-free areas.

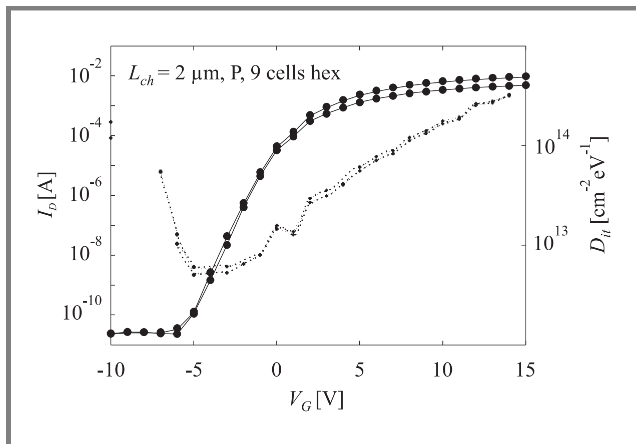


Fig. 6. Sub-threshold characteristics and calculated interface state density for a device with P implanted source at V_{DS} of 0.1 and 0.2 V.

A correlation of high leakage current and high number of extended crystal defects (mainly stacking faults) has been observed. To make full use of the performance 3C-SiC MOSFETs can offer the number of crystalline defects has to be reduced.

The values of specific on-resistance are comparable to the best values demonstrated for 4H-SiC vertical DMOSFETs.

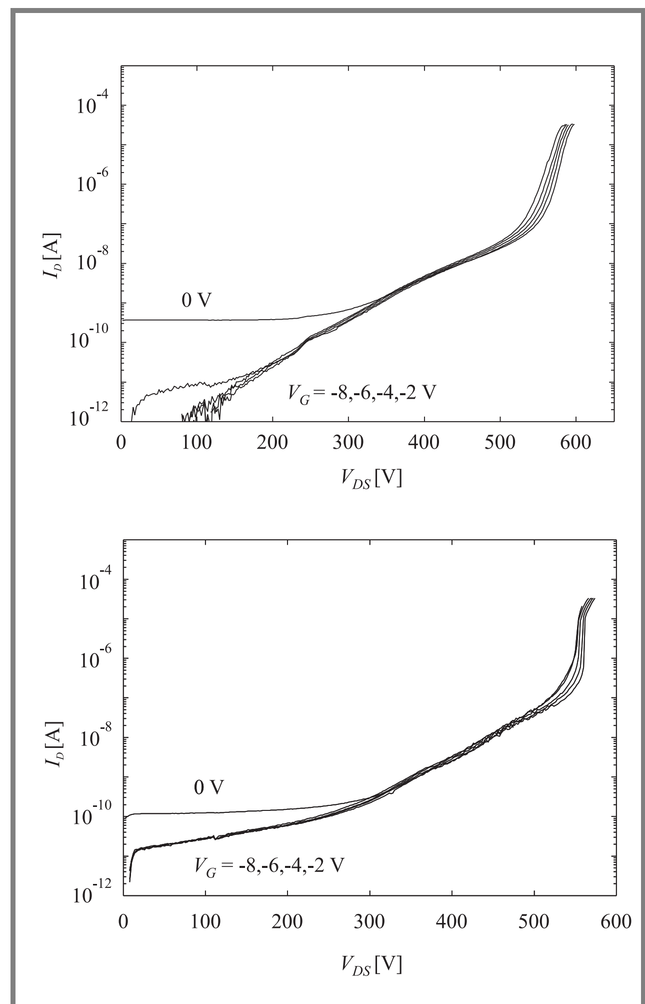


Fig. 7. Good blocking characteristics obtained on 1 cell MOSFETs made on the latest SBE material.

4. Recent results

A significant improvement of the reverse blocking capability have been obtained in devices made on the material obtained recently using the so-called switch-back epitaxy (SBE) [13]. Blocking capability between 550 and 600 V has been achieved at a leakage current of 1 μA, as shown in Fig. 7. The value of the maximum blocking voltage agrees well with the expected value of the breakdown voltage corresponding to the doping and thickness of the epitaxial drift layer.

Another optimistic finding is that the temperature dependence of the channel mobility is similar to that of the bulk material as can be seen in Fig. 8. The expected mobility dependence for the bulk material is shown as a dashed line in the figure. This is an interesting finding considering that the temperature dependence of channel mobility in hexagonal 6H- and 4H-SiC polytypes is often explained in terms of hopping conduction involving deep interface traps and that mobility increases with temperature [14].

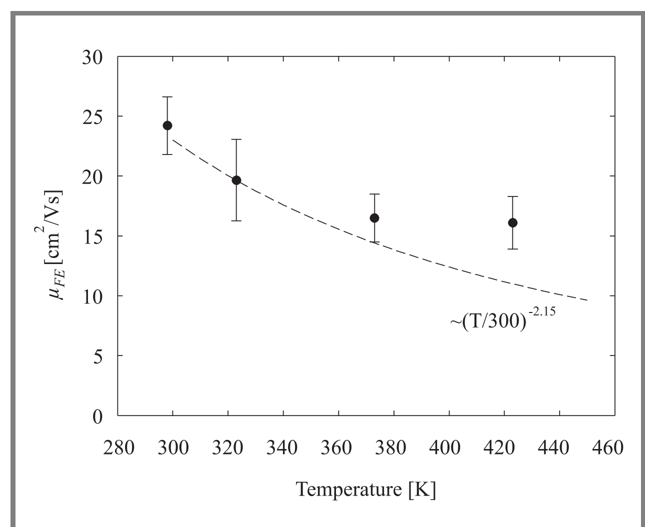


Fig. 8. Temperature dependence of the channel mobility extracted from the latest devices. The bulk mobility temperature dependence dominated by the phonon scattering is shown as a dashed line (the shown dependence holds for the 4H- and 6H- bulk material).

5. Comparison with 4H-SiC MOSFETs

The main differences in the expected performance of the 3C-SiC MOSFETs and 4H-SiC MOSFETs are illustrated in Figs. 9 and 10 [7].

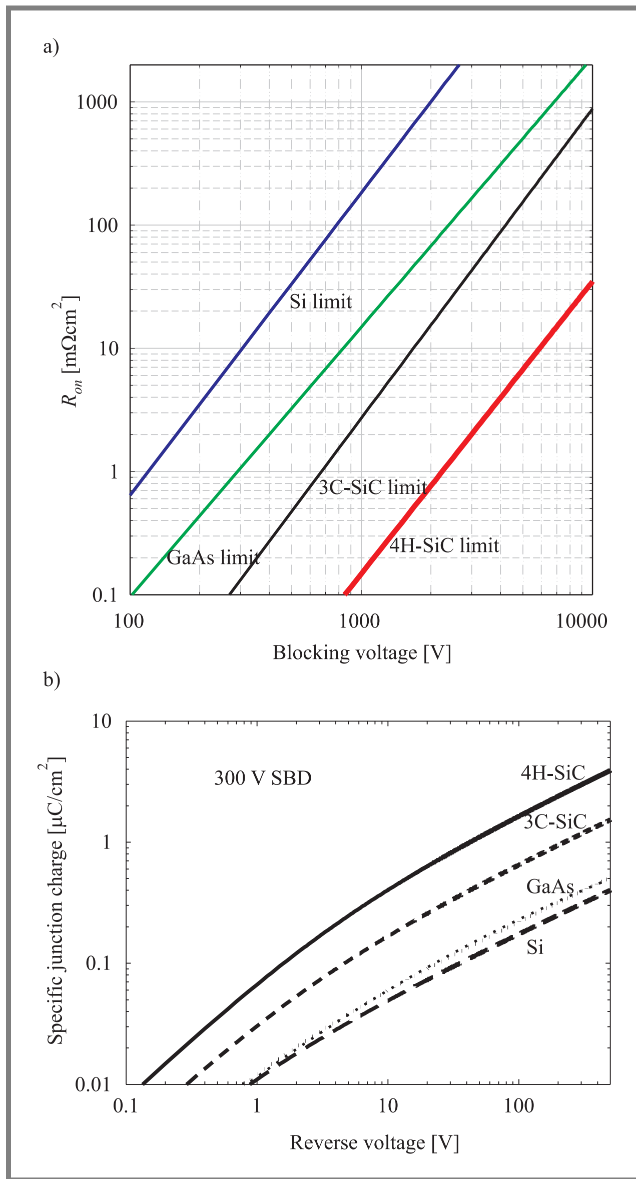


Fig. 9. The specific on-resistance, R_{on}, due to the drift region required for different blocking voltages for selected materials. This is the theoretical limit of R_{on} for the MOSFETs with homogeneously doped drift region in these materials (a). The specific value of the junction capacitance for the case of the homogeneously doped drift region for selected materials (b).

The window of possible improvement of specific resistance R_{on} in relation to silicon is smaller in the case of 3C-SiC due to the lower value of the critical electric field as shown in Fig. 9a. At the same time the lower doping required for a given blocking voltage in the case of 3C-SiC results in lower specific junction capacitance and faster devices as illustrated in Fig. 9b.

In Fig. 10 an estimation of the ultimate performance of DMOSFET devices for different values of the channel mo-

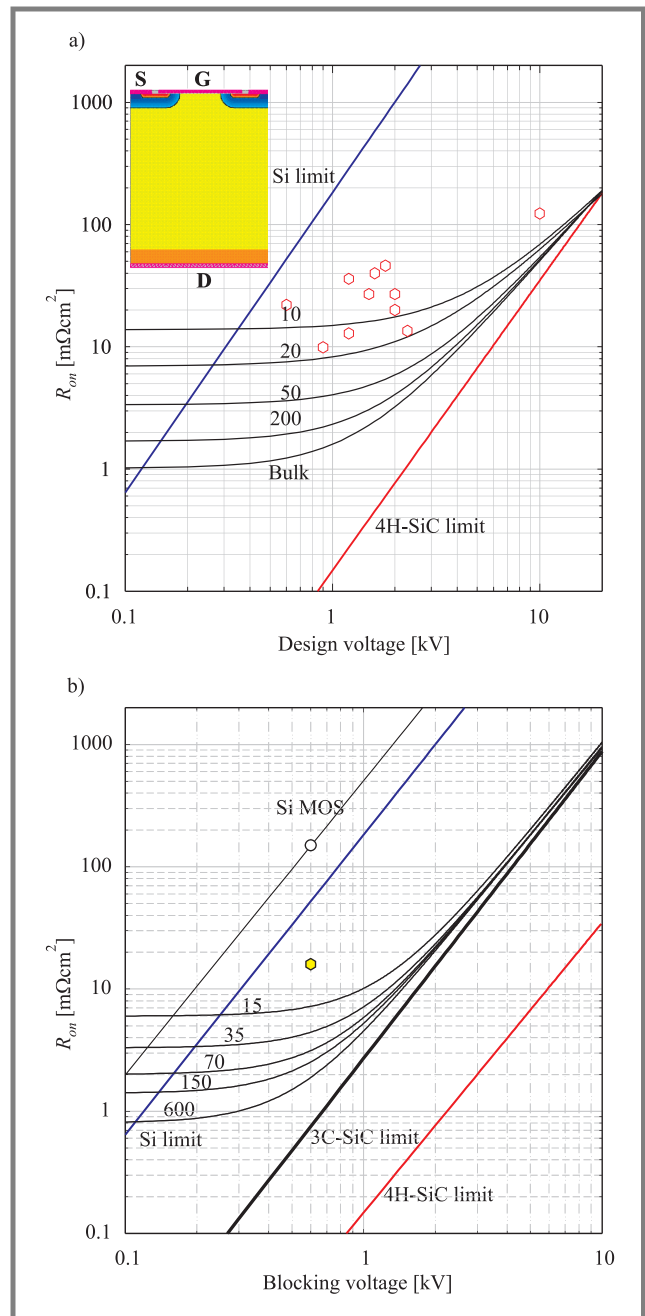


Fig. 10. A comparison of calculated and experimental data for 4H-SiC (a) and the 3C-SiC (b) DMOSFETs, respectively. The calculated data are for the ideal structure with 10 μm cell pitch and 1 μm channel length (the 1 μm line-width lithography) and for different values of channel mobility. The experimental data point for 3C-SiC corresponds to the structure with 28 μm cell pitch and 2 μm channel length ($R_{on} = 17$ mΩcm^2). The experimental values for 4H-SiC are values published by others.

bility is shown based on simulations together with the values of R_{on} demonstrated experimentally in the case of both 4H- (Fig. 10a) and 3C-SiC (Fig. 10b) DMOSFETs. It can

clearly be seen that the voltage range of 3C-SiC DMOS-FETs is limited to about 1200 V. The simulated structure is shown in an inset in Fig. 10a.

6. Summary and conclusions

The potential of 3C-SiC for 600 V MOSFET devices has been verified. The full blocking capability of 550–600 V has been obtained with new SBE material in the case of small devices. The leakage current was at the same time reduced by several orders of magnitude compared to the conventional material. The current capability has been shown to scale linearly with the device area. High enough values of the channel mobility (30–40 cm^2/Vs) have been obtained on large-area devices. The temperature dependence of channel mobility similar to that of the bulk material has been obtained in spite of high density of interface states. The device performance is dominated by the material quality and specifically by the density and distribution of the dislocations. Further improvement is necessary concerning the density of the dislocations and interface states.

References

[1] H. Nagasawa, K. Yagi, T. Kawahara, N. Hatta, G. Pensl, W. J. Choyke, T. Yamada, K. M. Itoh, and A. Schöner, "Low-defect 3C-SiC grown on undulant-Si (001) substrates", in *Silicon Carbide*, W. J. Choyke, H. Matsunami, G. Pensl, Eds. Berlin: Springer, 2003, pp. 207–228.

[2] R. Schörner, P. Friedrichs, and D. Peters, "Detailed investigation of n-channel enhancement 6H-SiC MOSFETs", *IEEE Trans. Electron Dev.*, vol. 46, pp. 533–541, 1999.

[3] M. Krieger, G. Pensl, M. Bakowski, A. Schöner, H. Nagasawa, and M. Abe, "Hall effect in the channel of 3C-SiC MOSFETs", *Mater. Sci. Forum*, vol. 483–485, pp. 441–444, 2005.

[4] J. Wan, M. A. Capano, M. R. Melloch, and J. A. Cooper, "N-channel 3C-SiC MOSFETs on silicon substrate", *IEEE Electron Dev. Lett.*, vol. 23, pp. 482–484, 2002.

[5] T. Ohshima, K. K. Lee, Y. Ishida, K. Kojima, Y. Tanaka, T. Takahashi, M. Yoshikawa, H. Okumura, K. Arai, and T. Kamiya," The electrical characteristics of metal-oxide-semiconductor field effect transistors fabricated on cubic silicon carbide ", *Jpn. J. Appl. Phys.*, vol. 42, pp. L625–L627, 2003.

[6] K. K. Lee, Y. Ishida, T. Ohshima, K. Kojima, Y. Tanaka, T. Takahashi, H. Okumura, K. Arai, and T. Kamiya, "N-channel MOSFETs fabricated on homoepitaxy-grown 3C-SiC films", *IEEE Electron Dev. Lett.*, vol. 24, pp. 466–468, 2003.

[7] M. Bakowski, "Status and prospects of SiC power devices", *IEE J. Trans. Ind. Appl.*, vol. 126, pp. 391–399, 2006.

[8] A. Schöner, M. Bakowski, P. Ericsson, H. Strömberg, H. Nagasawa, and M. Abe, "Realisation of large area 3C-SiC MOSFETs", *Mater. Sci. Forum*, vol. 483–485, pp. 801–804, 2005.

[9] M. Abe, H. Nakagawa, P. Ericsson, H. Strömberg, M. Bakowski, and A. Schöner, "High current capability of 3C-SiC vertical DMOS-FETs", *Microelectron. Eng.*, vol. 83, pp. 24–26, 2006.

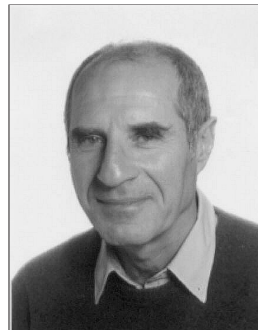
[10] A. Schöner, M. Bakowski, P. Ericsson, H. Strömberg, H. Nagasawa, and M. Abe, "Realisation of large area vertical 3C-SiC MOSFET devices", *Mater. Sci. Forum*, vol. 527–529, pp. 1273–1276, 2006.

[11] F. Ciobanu, G. Pensl, V. Afanas'ev, and A. Schöner, "Low density of interface states in n-type 4H-SiC MOS capacitors achieved by nitrogen implantation", *Mater. Sci. Forum*, vol. 483–485, pp. 693–696, 2005.

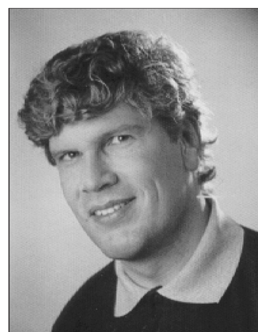
[12] V. Afanas'ev, F. Ciobanu, S. Dimitrijev, G. Pensl, and A. Stesmans, "SiC/SiO_2 interface states: properties and models", *Mater. Sci. Forum*, vol. 483–485, pp. 563–568, 2005.

[13] K. Yagi, T. Kawahara, N. Hatta, and H. Nagasawa, "Reducing planar defects in 3C-SiC", *Mater. Sci. Forum*, vol. 527–529, pp. 291–294, 2006.

[14] D. M. Brown, M. Ghezzo, J. Kretchmer, E. Downey, J. Pimbley, and J. Palmour, "SiC MOS interface characteristics", *IEEE Trans. Electron Dev.*, vol. 41, pp. 618–620, 1994.



Mietek Bakowski was born in Bydgoszcz, Poland, in 1946. He completed M.Sc. studies at the Faculty of Electronics, Warsaw University of Technology, in 1969. He received his Ph.D. and the Assistant Professor competence from Chalmers University of Technology, Gothenburg, Sweden, in 1974 and in 1981, respectively, and is presently a Senior Scientist at Acreo AB, Kista, Sweden. He has worked with the development of silicon bipolar and BiMOS power devices and since 1994 with the design, simulation and electrical evaluation of SiC power devices. In 2000–2003 he has been appointed Adjunct Professor at Royal Institute of Technology, KTH, Kista, Sweden.
e-mail: mietek.bakowski@acereo.se
Acreo AB
Electrum 236
SE-164 40 Kista-Stockholm, Sweden



Adolf Schöner was graduated in physics at the Institute of Applied Physics, University Erlangen-Nürnberg, Germany. After receiving his Ph.D. in 1994, he joined the Industrial Microelectronics Center (IMC) in Kista, Sweden, to continue working with material and processing issues of the material silicon carbide (SiC), specializing in electrical characterization of SiC epitaxial layers and devices. At Acreo AB he is now responsible for the development of electronic devices based on wide band gap materials and the epitaxial SiC-material development.
e-mail: adolf.schoner@acreo.se
Acreo AB
Electrum 236
SE-164 40 Kista-Stockholm, Sweden

Per Ericsson was graduated in 1992 and obtained a Ph.D. in solid state electronics in 1997 from Chalmers University of Technology, Gothenburg, Sweden. He worked with LD MOS RF-power device design and process integration issues at Ericsson Microelectronics between 1998 and 2000. He is now employed by Acreo AB in the Electronic Devices Group. His work mainly involves process development and device design for high power and high frequency devices based on wide band gap materials such as SiC and GaN.

e-mail: per.ericsson@acreo.se
Acreo AB
Electrum 236
SE-164 40 Kista-Stockholm, Sweden

Helena Strömberg focuses her interest on solid state electronics and has 14 years industrial experience from Ericsson Co. working with high power RF transistors and capacitors. Since 2002 she has been with Acreo AB working with process development, qualification of new processes and processing materials and fabrication of the SiC and GaN devices. Specifically, she was involved in processing of MOSFET, MESFET and HEMT transistors and SBD and PiN rectifiers.

e-mail: helena.stromberg@acreo.se
Acreo AB
Electrum 236
SE-164 40 Kista-Stockholm, Sweden

Hiroyuki Nagasawa received the B.E., M.E., and Ph.D. degrees of engineering from Tokai University, Japan, in 1985, 1987, and 1997, respectively. After receiving M.E. degree, he worked at Toshiba Co., as a semiconductor process engineer. From 1989, he was involved in development of X-ray lithography technologies in Hoya Co., and was mainly concerned with thin-film formation of 3C-SiC. In 2002, he joined in Hoya Advanced Semiconductor Technologies (HAST) Co. Ltd. as a chief-technology-officer (CTO). From 2006, he is conducting developments of SiC material and devices at Hoya Co. Ltd. as Senior General Manager.

e-mail: Hiroyuki_Nagasawa@sngw.rdc.hoya.co.jp
Hoya Co.
SiC Development Center
1-17-16 Tanashioda
Sagamihara 229-1125, Japan

Masayuki Abe received the B.E., M.E., and Ph.D. degrees in electrical engineering from Osaka University, Japan, in 1967, 1969 and 1973, respectively. In 1973, he joined Fujitsu Laboratories Ltd., Japan, where he was engaged in developing high-radiance LEDs, microwave HEMTs, high-speed HEMT LSIs for supercomputer, and low-temperature poly-Si TFTs for LCD flat-panels. Since 1998, he has been engaged in developing hexagonal GaN-HEMTs, infrared image-sensor at KRI Inc., and also 3C-SiC power MOSFETs and cubic GaN-HEMTs at Hoya Co. He is currently General Manager, Hoya Co., and President, HEMTCORE and 3D-bio Co. Ltd. Doctor M. Abe is a Fellow of IEEE.

e-mail: Masayuki_Abe@sngw.rdc.hoya.co.jp
Hoya Co.
SiC Development Center
1-17-16 Tanashioda
Sagamihara 229-1125, Japan

Invited paper

Properties and benefits of fluorine in silicon and silicon-germanium devices

Peter Ashburn and Huda A. W. El Mubarek

Abstract— This paper reviews the behaviour of fluorine in silicon and silicon-germanium devices. Fluorine is shown to have many beneficial effects in polysilicon emitter bipolar transistors, including higher values of gain, lower emitter resistance, lower 1/f noise and more ideal base characteristics. These results are explained by passivation of trapping states at the polysilicon/silicon interface and accelerated break-up of the interfacial oxide layer. Fluorine is also shown to be extremely effective at suppressing the diffusion of boron, completely suppressing boron transient enhanced diffusion and significantly reducing boron thermal diffusion. The boron thermal diffusion suppression correlates with the appearance of a fluorine peak on the SIMS profile at approximately half the projected range of the fluorine implant, which is attributed to vacancy-fluorine clusters. When applied to bipolar technology, fluorine implantation leads to a record f_T of 110 GHz in a silicon bipolar transistor.

Keywords— bipolar transistor, boron diffusion, fluorine, passivation of interface states, polysilicon emitter.

1. Introduction

Over the past few years there has been considerable interest in the behaviour of fluorine in silicon and silicon-germanium devices. This interest initially arose from the use of a BF_2^+ implant for shallow junction formation [1–3]. In bipolar technologies, a BF_2^+ implant was also used to create p-n-p polysilicon emitters and this led to a study of the behaviour of fluorine in polysilicon emitters [4]. This work showed that fluorine influenced a number of key bipolar transistor parameters, including gain [4], base current [5] and $1/f$ noise [6, 7]. More recently, research has focussed on the effect of fluorine on boron diffusion and it has been shown that fluorine suppresses boron transient enhanced diffusion [8, 9] and increases boron activation in silicon [8]. It has also been shown that fluorine decreases boron thermal diffusion in both silicon [10] and silicon-germanium [11, 12].

Recently fluorine implantation has been applied to MOS transistors to reduce boron diffusion in critical areas of the source and drain [13, 14]. Liu *et al.* [13] used a $1 \cdot 10^{15}$ cm^{-2} fluorine implant to create a super halo for both 50 nm n- and p-channel transistors. The fluorine-assisted halo process resulted in reduced junction capacitance and an improved $I_{on} - I_{off}$ trade-off. Fukutome *et al.* [14] used a $5 \cdot 10^{14} - 2 \cdot 10^{15}$ cm^{-2} fluorine implant prior to the p-channel extension implant to minimise the diffusion of boron in the extension. The fluorine implant led to dramatically improved threshold voltage roll-off characteristics without any degradation of drive current in sub-50 nm p-channel MOSFETs. Scanning tunnelling microscopy was used to show that the improvement was due to a reduction of the overlap length, for example from 13 to 7 nm in 40 nm gate length p-channel MOSFETs. Finally, Kham *et al.* [15] applied fluorine implantation to silicon bipolar technology to reduce the diffusion of the base and as a result have delivered a silicon bipolar transistor with a record f_T of 110 GHz.

In this paper, we review the properties and benefits of fluorine in silicon and silicon-germanium devices. We concentrate on fluorine behaviour in bipolar technology, beginning with a study of fluorine in polysilicon emitters and progressing to an investigation of the effects of fluorine on boron diffusion in silicon and silicon-germanium.

2. Effect of fluorine in polysilicon emitters

Fluorine exhibits a variety of interesting behaviour when implanted into the polysilicon emitter of a bipolar transistor. Figure 1 illustrates the effect of fluorine on the common emitter current gain, when 50 keV, $1 \cdot 10^{16}$ cm^{-2} fluorine is implanted into half of each wafer and annealed for 15 min at 850°C (Fig. 1a) or 120 min at 850°C (Fig. 1b) in nitrogen. For the short anneal, Fig. 1a shows that the fluorine implant leads to a higher gain, whereas for the long anneal, Fig. 1b shows that it leads to a lower gain. Thus fluorine gives radically different behaviour for low and high thermal budget anneals.

Figure 2 shows that the effect of fluorine on gain is due to a change in base current rather than a change in collector current. It also shows that for short drive-in times fluorine gives a lower base current (and hence a higher gain), while for long drive-in times it gives a higher base current than the control devices. The same trend is seen for devices given an HF etch (Fig. 2a) and an RCA clean (Fig. 2b) prior to polysilicon deposition, though the cross-over of the two curves occurs at longer times for the RCA clean than the HF etch.

The explanation for the lower values of base current at short anneal times is passivation of dangling bonds at the polysilicon/silicon interface by the fluorine. This is illustrated in Fig. 3, where dangling bonds are passivated by the formation of Si-O-F and Si-F complexes [16]. With the passivation of these dangling bonds, recombination of minority carriers at the interface is reduced and conse-

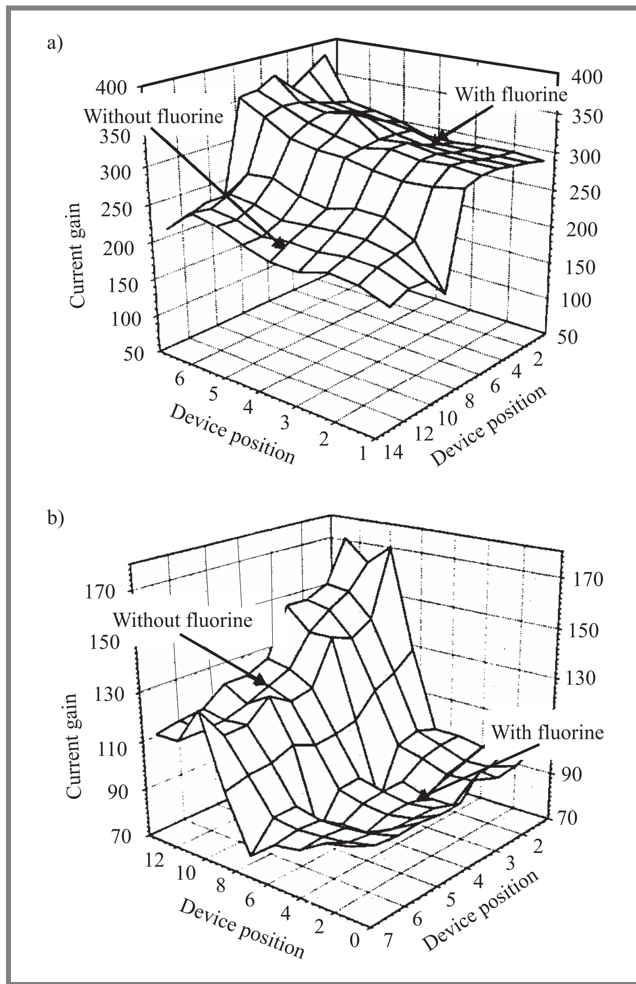


Fig. 1. The 3D plot of polysilicon emitter bipolar transistor gain measured across a wafer. The wafer had a 50 keV, $1 \cdot 10^{16}$ cm^{-2} F^{+} implant into half of each wafer and the wafers were given the following anneals: (a) 15 min at 850°C; (b) 120 min at 850°C. *After Moiseiwitsch et al. [4] © IEEE.*

quently the base current decreases. The explanation for the higher values of base current at long anneal times is the effect of fluorine in enhancing the break-up of the native oxide layer at the polysilicon/silicon interface [17]. This is shown in Fig. 4, which shows cross-section transmission electron microscopy (TEM) micrographs of polysilicon emitters without and with a fluorine implant. Figure 4a shows the situation for the control device without a fluorine implant and it can be seen that no regrowth of the polysilicon has occurred. In contrast for the fluorine implanted device in Fig. 4b, the interfacial oxide layer has broken up and the bottom part of the polysilicon layer has epitaxially regrown. Detailed experiments [18] have shown that fluorine increases both the interfacial oxide break-up and the polysilicon regrowth rate.

Fluorine has additional benefits in polysilicon emitter bipolar transistors, including reduced $1/f$ noise [6] and improved base current ideality [5]. The improved $1/f$ noise occurs because the fluorine breaks up the interfacial oxide layer, which is a source of noise in polysilicon emitters. The improved base current ideality is shown in Fig. 5a, where it can be seen that the ideality factor improves from 1.51 to 1.26 as a result of the fluorine implant. This improvement arises because the fluorine passivates interface states at the oxide/silicon interface where the emitter/base depletion region intersects the surface oxide, as shown in Fig. 5b. To reach the oxide/silicon interface, the fluorine has diffused through the polysilicon emitter and the screen oxide during the emitter anneal. This is possible because fluorine diffuses extremely rapidly in polysilicon [19] and silicon dioxide.

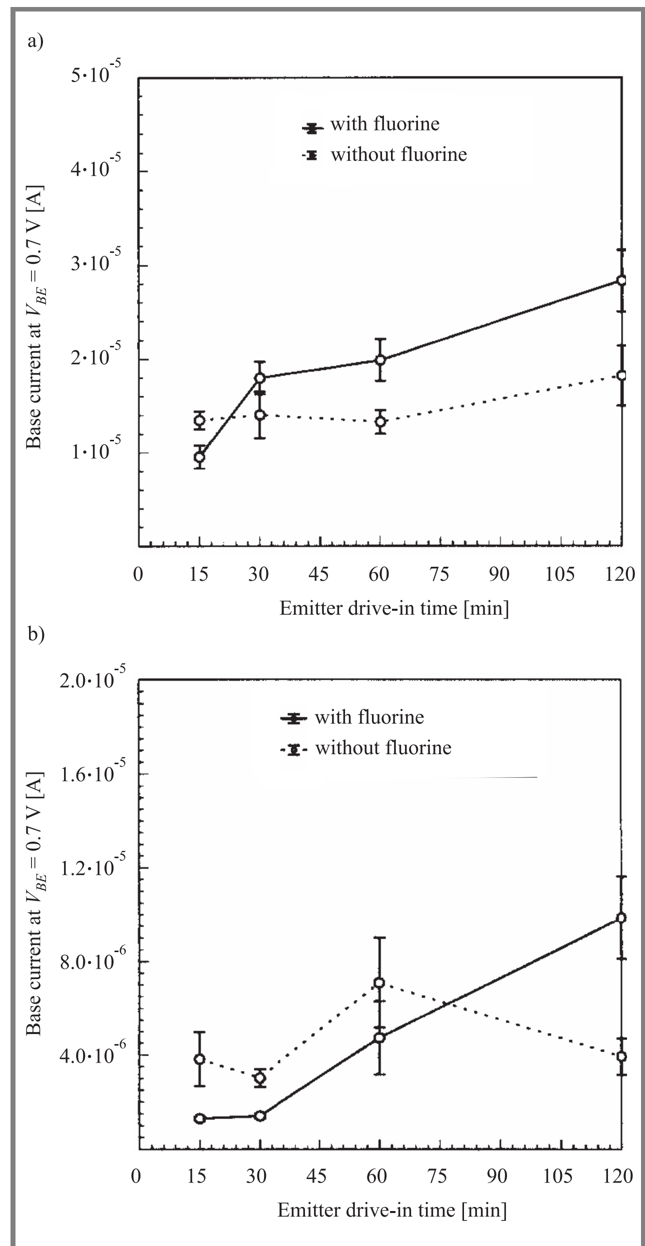


Fig. 2. Common emitter current gain as a function of emitter anneal time at 850°C for wafers given different wet chemical treatments prior to polysilicon deposition: (a) HF etch; (b) RCA clean. *After Moiseiwitsch et al. [4] © IEEE.*

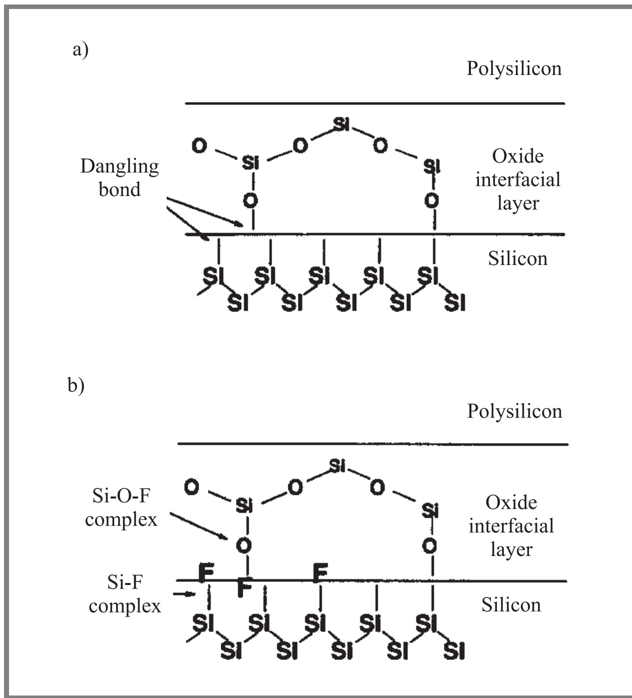


Fig. 3. Schematic illustration of the polysilicon/silicon interface showing the passivation of dangling bonds by fluorine as a result of the formation of Si-O-F and Si-F complexes: (a) interface in the absence of fluorine; (b) interface in the presence of fluorine. *After Moiseiwitsch et al. [4] © IEEE.*

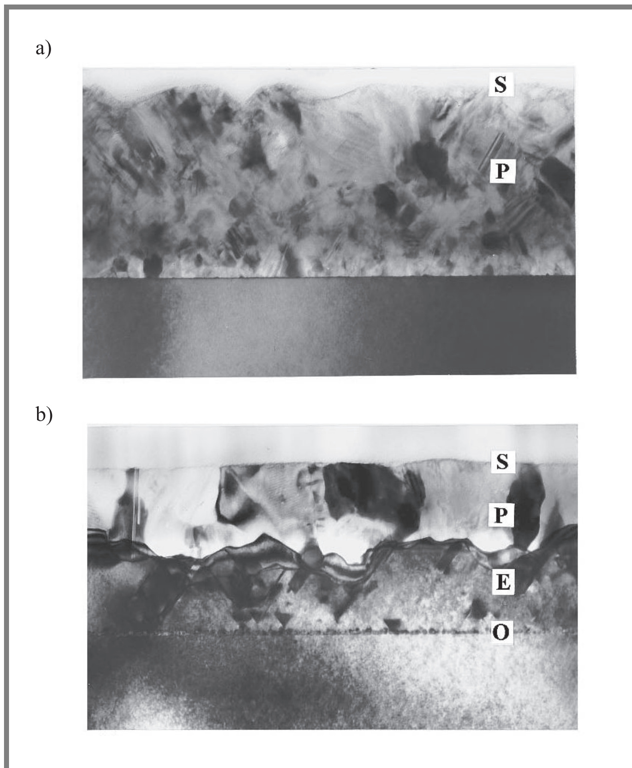


Fig. 4. Cross-section TEM micrographs of polysilicon emitters without and with a fluorine implant: (a) no fluorine implant and anneal of 480 min at 850°C; (b) 50 keV, $1 \cdot 10^{16}$ cm^{-2} F implant and anneal of 240 min at 850°C. *After Moiseiwitsch et al. [17] © AIP.*

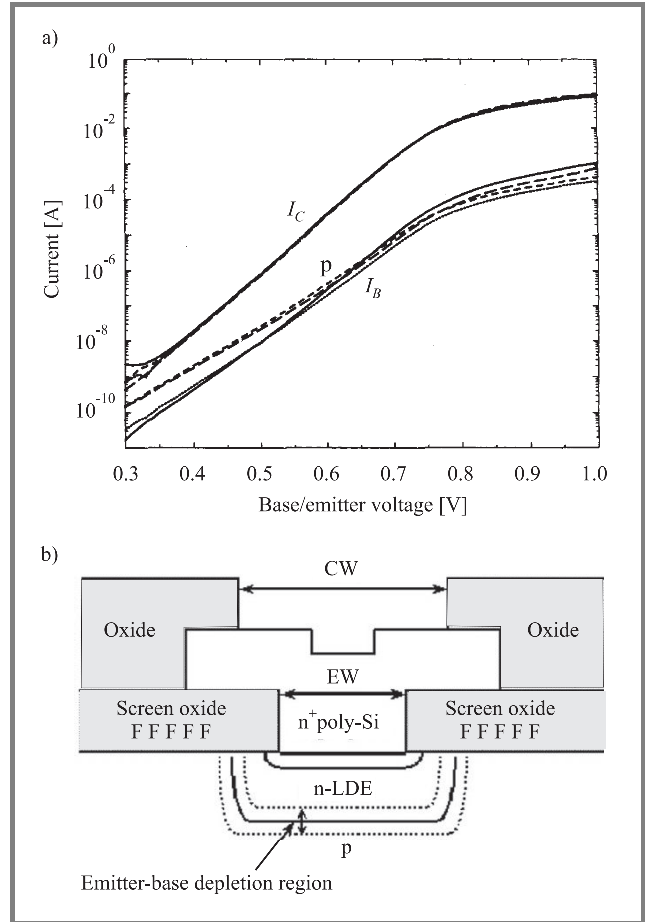


Fig. 5. Gummel plots of fluorine implanted and control polysilicon emitter bipolar transistors (a). The solid lines show fluorine-implanted transistors and the dashed lines show transistors without a fluorine implant. Schematic cross-section of the polysilicon emitter (b) showing the location of passivating fluorine at the interface between the screen oxide and the silicon wafer. *After Moiseiwitsch et al. [4] © IEEE.*

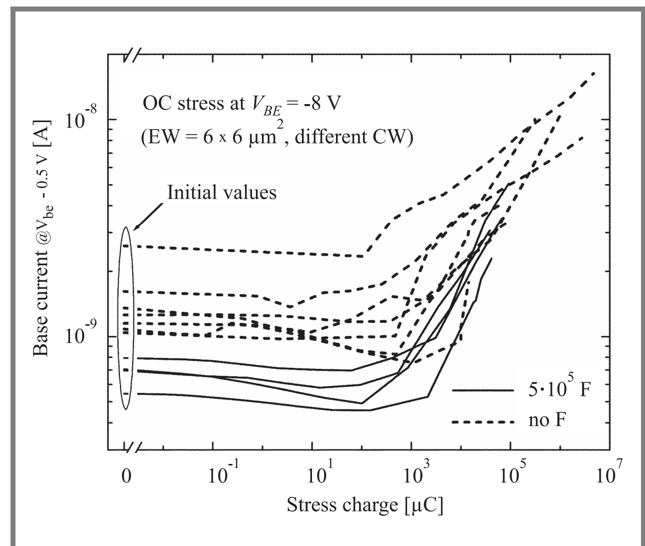


Fig. 6. Base current (at 0.5 V) as a function of stress charge in polysilicon emitter bipolar transistors with and without a fluorine implant. *After Sheng et al. [20] © IEEE.*

Hot carrier stressing experiments have been carried out on polysilicon bipolar transistors incorporating fluorine and results are shown in Fig. 6. The initial values of base current in all fluorine implanted devices are lower than those of the control devices. In all cases, the base current begins to increase at a stress charge of $10^2 - 10^4$ μC, whether the device is fluorine implanted or not, and then degrades at approximately the same rate with increasing stress charge. Detailed characterisation [20] has shown that benefits achieved by fluorine implantation are robust. This is, there is no evidence that defects passivated by fluorine are reactivated during stressing or that fluorine implantation introduces any additional defects that are activated during stressing.

3. Boron diffusion control using fluorine

Figure 7 shows the effect of fluorine on boron transient enhanced diffusion and boron thermal diffusion in silicon. A boron marker layer was used to monitor boron diffusion and transient enhanced diffusion was created using the damage from a P^+ implant at the same depth as the fluorine implant [9]. The results in Fig. 7a show that the P^+ implant induces considerable transient enhanced boron diffusion as a result of point defects created by the P^+ implant. However, the profile for the sample implanted with P^+ and F^+ indicates that the fluorine has dramatically suppressed the transient enhanced boron diffusion. Furthermore, Fig. 7b shows that the sample implanted with (185 keV, $2.3 \cdot 10^{15}$ cm^{-2}) F^+ shows less diffusion than the unimplanted sample. This is a surprising result, which indicates that the fluorine implant has significantly decreased boron thermal diffusion.

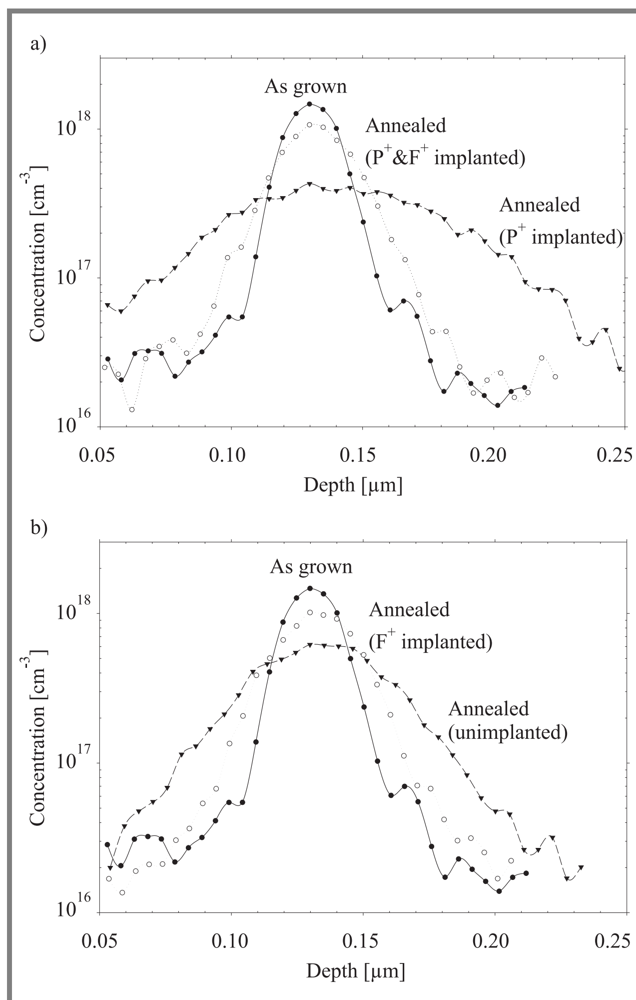


Fig. 7. Boron SIMS profiles of a buried marker layer after anneal for 30 s at 1000°C in nitrogen: (a) samples implanted with P^+ only and with P^+ and F^+; (b) an unimplanted sample and a sample implanted with F^+ only. *After Mubarek et al. [9] © AIP.*

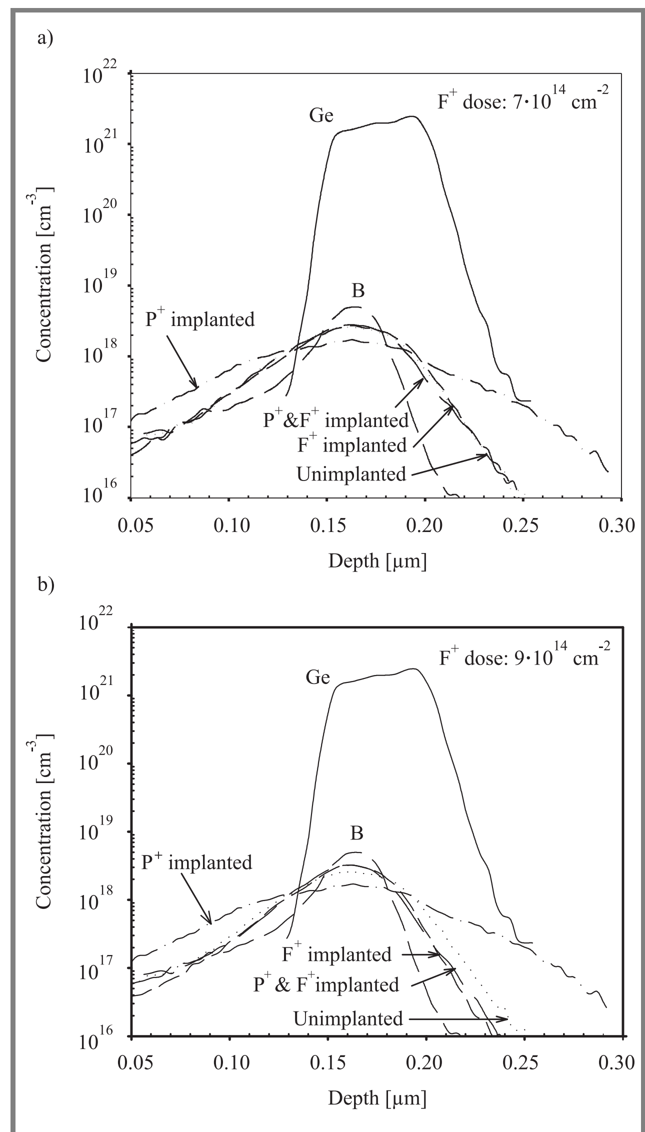


Fig. 8. Boron profiles after anneal for 30 s at 1000°C in nitrogen for samples implanted with P^+ and F^+, P^+ only, F^+ only and for an unimplanted sample. Results are shown for F^+ implantation doses of (a) $7 \cdot 10^{14}$ cm^{-2} and (b) $9 \cdot 10^{14}$ cm^{-2}. *After Mubarek et al. [12] © IEEE.*

Further insight into the effect of fluorine on boron diffusion can be obtained by investigating the effect of

the fluorine implantation dose on the diffusion suppression [12]. Figure 8 shows boron secondary ions mass spectroscopy (SIMS) profiles for silicon-germanium samples implanted with P^+ and F^+, P^+ only, F^+ only and for an unimplanted sample. The results in Fig. 8a show that at a F^+ implantation dose of $7 \cdot 10^{14}$ cm^{-2}, the boron profiles for the two F^+ implanted samples are identical to that of the unimplanted sample, indicating that the F^+ implant has completely suppressed boron transient enhanced diffusion. In contrast, Fig. 8b shows that at a F^+ implantation dose of $9 \cdot 10^{14}$ cm^{-2}, the boron profiles for the two F^+ implanted samples show less diffusion than that of the unimplanted sample. This result indicates that the F^+ implant suppresses boron thermal diffusion at this implantation dose. There is therefore a critical F^+ implantation dose, above which boron thermal diffusion is suppressed and below which only boron transient enhanced diffusion is suppressed.

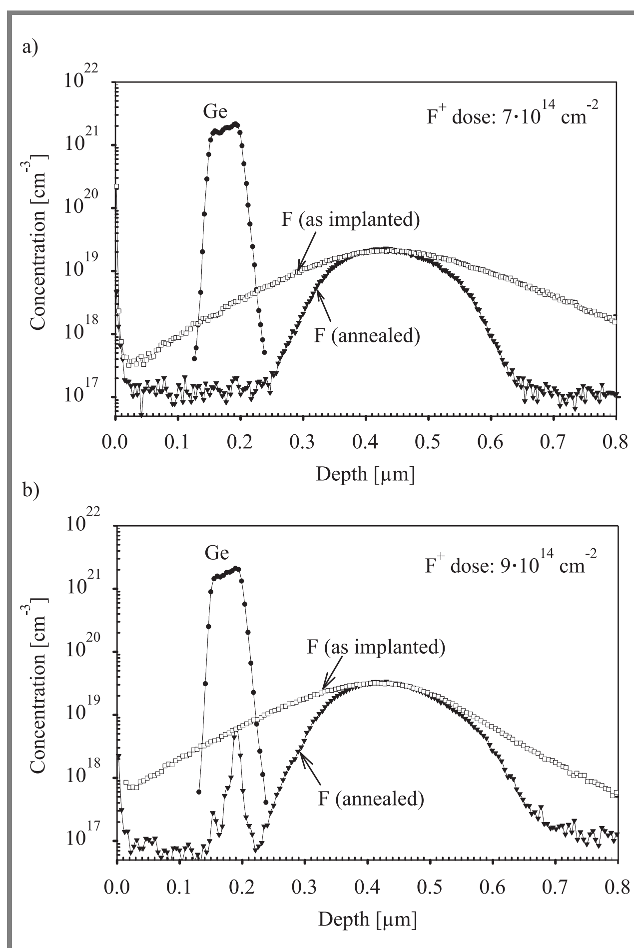


Fig. 9. Fluorine profiles after anneal for 30 s at 1000°C in nitrogen for samples implanted with P^+ and F^+, P^+ only, F^+ only and for an unimplanted sample. Results are shown for F^+ implantation doses of (a) $7 \cdot 10^{14}$ cm^{-2} and (b) $9 \cdot 10^{14}$ cm^{-2}. *After Mubarek et al. [12] © IEEE.*

Insight into the mechanism of boron thermal diffusion suppression can be obtained from the fluorine SIMS profiles of samples implanted with different F^+ doses, as shown in Fig. 9. For a F^+ implantation dose of $7 \cdot 10^{14}$ cm^{-2}, Fig. 9a shows the presence of a broad peak around the range of the fluorine implant, but little fluorine is present in the silicon-germanium layer. In contrast, for a F^+ implantation dose of $9 \cdot 10^{14}$ cm^{-2}, Fig. 9b shows the presence of an additional shallow fluorine peak in the silicon-germanium layer. There is therefore a correlation between the appearance of the fluorine peak in the silicon-germanium layer and the reduction in boron thermal diffusion.

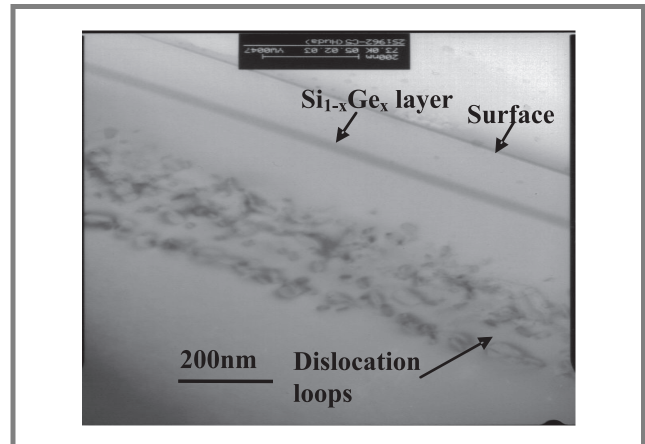


Fig. 10. Cross-section TEM micrograph of a sample implanted with $2.3 \cdot 10^{15}$ cm^{-2} F^+ and annealed at 1000°C for 30 s in nitrogen. *After Mubarek et al. [12] © IEEE.*

Figure 10 shows a cross-section TEM micrograph of the sample implanted with $2.3 \cdot 10^{15}$ cm^{-2} F^+. A band of dislocation loops can be seen extending from a depth of about 0.3 to 0.5 μm, but no defects are visible in the silicon-germanium ($Si_{1-x}Ge_x$) layer. The band of dislocation loops correlates with the broad fluorine peak in Fig. 9, indicating that this peak is due to fluorine trapped at the dislocation loops. The lack of any defect contrast in the $Si_{1-x}Ge_x$ layer indicates that any defects present must be too small to resolve by TEM. The shallow fluorine peak is located in the vacancy-rich region of the fluorine implant damage profile and this leads us to conclude that it is due to fluorine trapped in vacancy-fluorine clusters [12]. This conclusion has been confirmed by point defect injection studies on fluorine implanted samples [21]. The reduction of boron thermal diffusion above the critical fluorine dose is then explained by the action of the vacancy-fluorine clusters in suppressing the interstitial concentration in the silicon-germanium layer. Since boron diffusion is mediated by interstitials, this suppressed interstitial concentration gives reduced boron diffusion.

4. Application of fluorine diffusion suppression technique in bipolar transistors

The above fluorine diffusion suppression technique has been applied to a double polysilicon silicon bipolar tech-

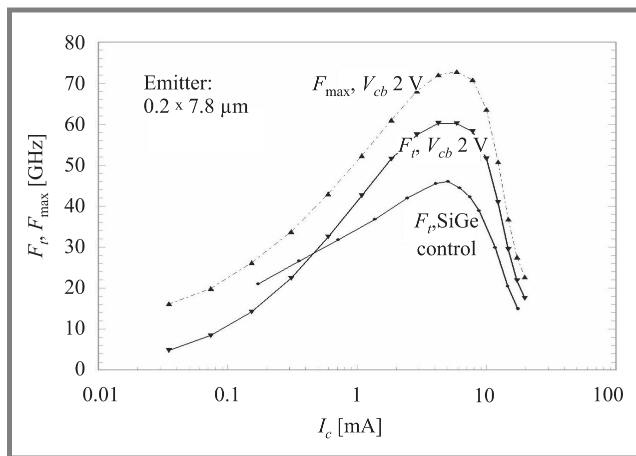


Fig. 11. Effect of a 150 keV, $5 \cdot 10^{14}$ cm^{-2} F$^+$ implant on the f_T and f_{max} of a double polysilicon silicon bipolar transistor. *After Kham et al [15] © IEEE.*

nology in collaboration with ST Microelectronics, Catania, Italy [15]. Figure 11 shows the effect of a 150 keV, $5 \cdot 10^{14}$ cm^{-2} F$^+$ implant on the f_T of a silicon bipolar transistor and it can be seen that fluorine increases the maximum f_T from 46 to 60 GHz. A further improvement in performance can be obtained by scaling the basewidth and optimising the collector doping, while keeping the fluorine

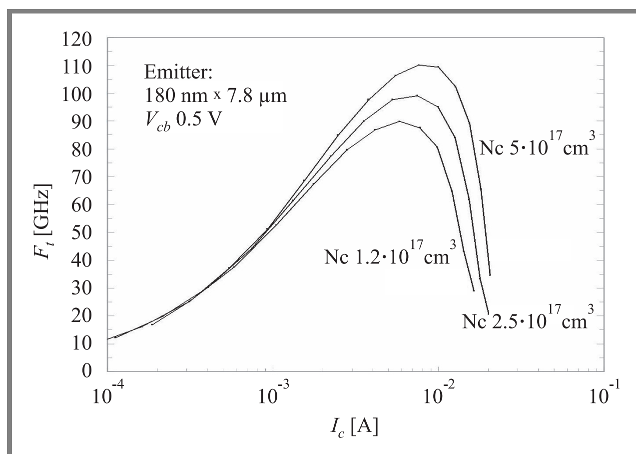


Fig. 12. Values of f_T as a function of collector current for double polysilicon silicon bipolar transistors implanted with 150 keV, $5 \cdot 10^{14}$ cm^{-2} F$^+$ and with different collector profiles. *After Kham et al [15] © IEEE.*

implant conditions the same. Results are shown in Fig. 12, where it can be seen that values of peak f_T of 90, 100 and 110 GHz are obtained for collector junction concentrations of 1.2, 2.5 and $5.0 \cdot 10^{17}$ cm^{-2}, respectively. The value of BV$_{ceo}$ at $\sim$ 5 mA collector current varied slightly with collector implant dose, with values around 2.5 V, such that all three variants had Johnson numbers of $\sim$ 250 GHz-V. These values of f_T are the highest ever reported for silicon bipolar transistors.

5. Conclusions

A review has been undertaken of the properties and benefits of fluorine in silicon and silicon-germanium devices. Fluorine has been found to exhibit a variety of interesting effects, many of which are beneficial to device performance. When fluorine is implanted into polysilicon emitters, benefits can include higher gain, lower emitter resistance, reduced $1/f$ noise and improved base current ideality. Two conflicting mechanisms have been identified, the first being passivation of interface states by fluorine segregated at the polysilicon/silicon interface and the second being accelerated break up of the interfacial oxide layer. When fluorine is implanted into the collector, a dramatic suppression of boron diffusion is obtained. A critical fluorine dose has been identified, below which fluorine suppresses boron transient enhanced diffusion and above which boron thermal diffusion is also suppressed. This suppression of boron thermal diffusion correlates with the appearance of a fluorine SIMS peak in the silicon-germanium layer that has been attributed to vacancy-fluorine clusters. The reduction of boron thermal diffusion has been explained by the effect of the vacancy-fluorine clusters in suppressing the interstitial concentration in the silicon-germanium layer. When applied to silicon bipolar transistors, fluorine has delivered a record f_T of 110 GHz.

Acknowledgements

The authors would like to acknowledge EPSRC and the EU SINANO project for funding this research and RAEng/EPSRC for funding a Fellowship.

References

[1] R. G. Wilson, "Boron, fluorine and carrier profiles for B and BF$_2$ implants into crystalline and amorphous Si", *J. Appl. Phys.*, vol. 54, pp. 6879–6889, 1983.

[2] D. F. Downey, J. W. Chow, E. Ishida, and K. S. Jones, "Effect of fluorine on the diffusion of boron in ion implanted silicon", *Appl. Phys. Lett.*, vol. 73, p. 1263, 1998.

[3] Y.-J. Park and J.-J. Kim, "Fluorine implantation effect on boron diffusion in silicon", *J. Appl. Phys.*, vol. 85, p. 803, 1999.

[4] N. E. Moiseiwitsch and P. Ashburn, "The benefits of fluorine in pnp polysilicon emitter bipolar transistors", *IEEE Trans. Electron Dev.*, vol. ED-41, pp. 1249–1255, 1994.

[5] J. F. W. Schiz and P. Ashburn, "Improved base current ideality in polysilicon emitter bipolar transistors due to fast fluorine diffusion through oxide", *Electron. Lett.*, vol. 35, pp. 752–753, 1999.

[6] N. Siabi-Shahrivar, W. Redman-White, P. Ashburn, and H. A. Kemhadjian, "Reduction of $1/f$ noise in polysilicon emitter bipolar transistors", *Solid State Electron.*, vol. 38, pp. 389–400, 1995.

[7] N. Lukyanchikova, N. Garbar, M. Petrichuk, J. F. W. Schiz, and P. Ashburn, "The influence of BF$_2$ and F implants on the $1/f$ noise in SiGe HBTs with a self-aligned link base", *IEEE Trans. Electron Dev.*, vol. 48, no. 12, pp. 2808–2815, 2001.

[8] K. Ohyu, T. Itoga, and N. Natsuaki, "Advantages of fluorine introduction in boron implanted shallow pn junction formation", *Jpn. J. Appl. Phys.*, vol. 29, p. 457, 1990.

[9] H. A. W. El Mubarek and P. Ashburn, "Reduction of boron thermal diffusion and elimination of boron transient enhanced diffusion in silicon by high energy fluorine implantation", *Appl. Phys. Lett.*, vol. 83, no. 20, pp. 4134–4136, 2003.

[10] H. A. W. El Mubarek, M. Karunaratne, J. M. Bonar, G. D. Dilliway, Y. Wang, R. Price, J. Zhang, P. L. F. Hemment, A. F. Willoughby, P. Ward, and P. Ashburn, "Effect of fluorine implantation dose on boron thermal diffusion in silicon", *J. Appl. Phys.*, vol. 96, no. 8, pp. 4114–4121, 2004.

[11] H. A. W. El Mubarek and P. Ashburn, "Reduction of boron thermal diffusion and elimination of boron transient enhanced diffusion in silicon-germanium by high energy fluorine implantation", *IEEE Electron Dev. Lett.*, vol. 25, pp. 535–537, 2004.

[12] H. A. W. El Mubarek, M. Karunaratne, J. M. Bonar, G. D. Dilliway, Y. Wang, P. L. F. Hemment, A. F. Willoughby, and P. Ashburn, "Effect of fluorine implantation dose on boron transient enhanced diffusion and boron thermal diffusion in $Si_{1-x}Ge_x$", *IEEE Trans. Electron Dev.*, vol. 52, no. 4, pp. 518–526, 2005.

[13] K. Liu, J. Wu, J. Chen, and A. Jain, "Fluorine assisted super halo for sub-50 nm transistors", *IEEE Electron Dev. Lett.*, vol. 24, pp. 180–182, 2003.

[14] H. Fukutome, Y. Momiyama, H. Nakao, T. Aoyama, and H. Arimoto, "Fluorine implantation impact in extension region on the electrical performance of sub-50 nm p-MOSFETs", *Tech. Dig. Int. Electron Dev. Meet.*, Washington, USA, 2003, pp. 485–488.

[15] M. N. Kham, H. A. W. El Mubarek, J. M. Bonar, P. Ashburn, P. Ward, L. Fiore, R. Petralia, C. Alemanni, and A. Messina, "110 GHz f_T silicon bipolar transistors implemented using fluorine implantation for boron diffusion suppression", *IEEE Trans. Electron Dev.*, vol. 53, no. 3, pp. 545–552, 2006.

[16] Y. Ono, M. Tabe, and Y. Sakakibara, "Segregation and defect termination of fluorine at $Si0_2$/Si", *Interf. Appl. Phys. Lett.*, vol. 62, p. 375, 1993.

[17] N. E. Moiseiwitsch, C. Marsh, P. Ashburn, and G. R. Booker, "Epitaxial regrowth of n+ polycrystalline silicon at 850°C induced by fluorine implantation", *Appl. Phys. Lett.*, vol. 66, pp. 1918–1920, 1995.

[18] C. D. Marsh, N. E. Moiseiwitsch, G. R. Booker, and P. Ashburn, "Behaviour and effects of fluorine in annealed n+ polycrystalline silicon layers on silicon wafers", *J. Appl. Phys.*, vol. 87, pp. 7567–7578, 2000.

[19] G. R. Nash, J. F. W. Schiz, C. D. Marsh, P. Ashburn, and G. R. Booker, "Activation energy for fluorine transport in amorphous silicon", *Appl. Phys. Lett.*, vol. 75, pp. 3671–3673, 1999.

[20] S. R. Sheng, W. R. McKinnon, S. P. McAlister, C. Storey, J. S. Hamel, and P. Ashburn, "Hot carrier stressing of NPN polysilicon emitter bipolar transistors incorporating fluorine", *IEEE Trans. Electron Dev.*, vol. 50, no. 4, pp. 1141–1144, 2003.

[21] M. N. Kham, H. A. W. El Mubarek, J. M. Bonar, and P. Ashburn, "Study of fluorine behaviour in silicon by selective point defect injection", *Appl. Phys. Lett.*, vol. 87, p. 11902, 2005.

Peter Ashburn was born in Rotherham, England, in 1950. He received the B.Sc. degree in electrical and electronic engineering in 1971, and the Ph.D. degree in 1974, both from the University of Leeds. His dissertation topic was an experimental and theoretical study of radiation damage in silicon p-n junctions. In 1974, he joined the technical staff of Philips Research Laboratories and worked initially on ion implanted integrated circuit bipolar transistors, and then on electron lithography for sub-micron integrated circuits. In 1978 he joined the academic staff of the Department of Electronics & Computer Science of the University of Southampton as a lecturer, and currently is the holder of a Personal Chair in Microelectronics. Since taking up a post at University of Southampton, Professor Ashburn has worked on polysilicon emitter bipolar transistors, high-speed bipolar and BiCMOS technologies, gate delay expressions for bipolar circuits and the effects of fluorine in bipolar transistors. His current research interests include SiGe HBTs on SOI, SiGeC devices and ultimate DMOS devices. HE has authored and co-authored 170 papers in technical literature, given invited papers on polysilicon emitters and SiGe heterojunction bipolar transistors and has authored two books on bipolar transitors and SiGe heterojunction bipolar transistors.

e-mail: pa@ecs.soton.ac.uk
School of Electronics & Computer Science
University of Southampton
Southhampton, SO17 1BJ, UK

Huda A. W. El Mubarek has received a first clas honours B.Eng. degree in electronic engineering from the University of Southampton, United Kingdom, in 1999. Since then she has been working on several research areas including fluorine effect on boron diffusion, SiGe and SiGeC heterojunction bipolar transistors (HBTs) on bulk, silicon on insulator (SOI) and silicide silicon on insulator (SSOI) substrates and has over 20 publications. During this period she has both completed her Ph.D. (2004) and worked for 3 years as a research Assistant at the University of Southampton in an EPSRC funded project. During her Ph.D. she has developed a technique using fluorine implantation to suppress boron diffusion both in Si and SiGe and has several publications in this area. She has recently been awarded a prestigious joint Royal Academy of Engineering and EPSRC 5 years research fellowship at the University of Southampton. Her fellowship research topic is point defect engineering for suppression of dopant diffusion.

e-mail: hawem@ecs.soton.ac.uk
School of Electronics & Computer Science
University of Southampton
Southhampton, SO17 1BJ, UK

Invited paper

Low frequency noise in Si and Si/SiGe/Si PMOSFETs

Stephen M. Thomas, Martin J. Prest, Dominic J. F. Fulgoni, Adam R. Bacon, Tim J. Grasby, David R. Leadley, Evan H. C. Parker, and Terence E. Whall

Abstract— Measurements of $1/f$ noise in Si and $Si_{0.64}Ge_{0.36}$ PMOSFETs have been compared with theoretical models of carrier tunnelling into the oxide. Reduced noise is observed in the heterostructure device as compared to the Si control. We suggest that this is primarily associated with an energy dependent density of oxide trap states and a displacement of the Fermi level at the SiO_2 interface in the heterostructure relative to Si. The present study also emphasizes the important role of transconductance enhancement in the dynamic threshold mode in lowering the input referred voltage noise.

Keywords— electronic noise, silicon germanium heterostructures, MOSFET, dynamic threshold mode.

1. Introduction

Low frequency noise limits the performance of analogue CMOS circuits and could ultimately degrade the noise margin in digital CMOS circuits. Si/SiGe/Si PMOS devices offer low noise solutions with enhanced maximum voltage gain for analogue applications; however, a consensus is yet to emerge on the detailed mechanism of noise reduction. In this paper, we summarize our previous measurements on this system which have led us to suggest that the origin of the noise reduction lies in the energy dependent density of oxide trap states and the suppression of carrier number fluctuations [1]. New data is presented in support of this contention and the important role of transconductance enhancement in the dynamic threshold (DT) mode (gate connected to body) as opposed to body tied (BT) mode (source connected to body) operation is illustrated.

2. Experimental results

Figure 1 shows the devices on which measurements were made, which were fabricated in a 0.5 μm process [2]. The carrier mobility in these devices, Fig. 2, depends on silicon cap thickness, which strongly suggests that alloy scattering is not a dominant mobility limiting process and this conclusion is supported by our other work [3].

The normalised current noise power spectral density (PSD) S_I/I^2 of a SiGe device is shown in Fig. 3, together with that of a surface channel Si control having the same vertical doping profile. The noise in the SiGe device is clearly much reduced as compared to the control – by an order of magnitude in the low frequency region where $1/f$ noise dominates. The corresponding PSDs for resistance fluctua-

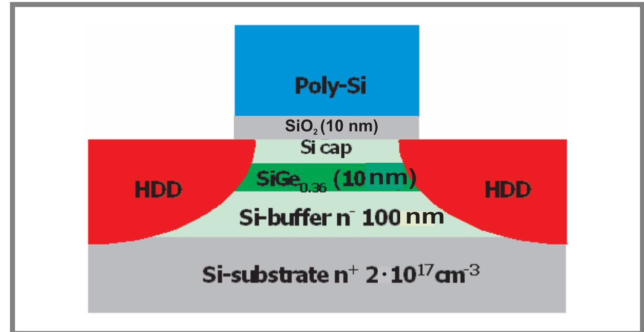


Fig. 1. Schematic diagram of PMOSFET devices used. The Si control is identical apart from the alloy layer.

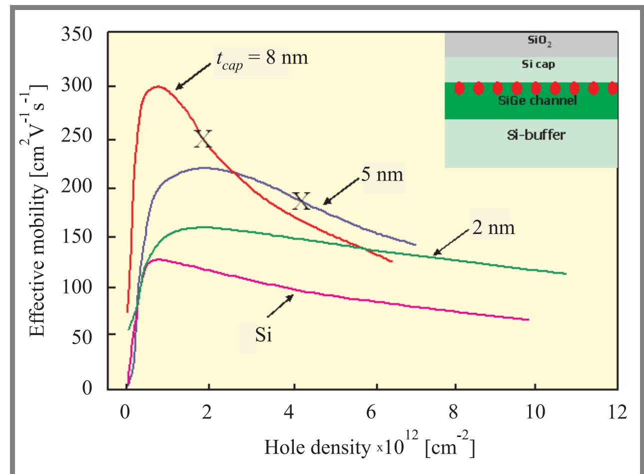


Fig. 2. Room temperature hole mobility for three different cap thicknesses, demonstrating clearly that alloy scattering is not a dominant mobility limiting mechanism [2].

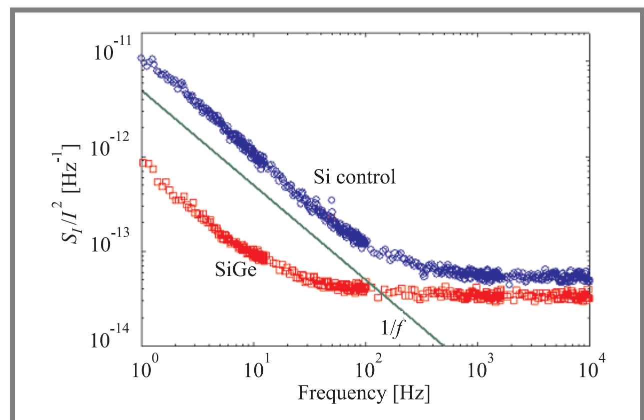


Fig. 3. Current noise power spectral density, demonstrating that SiGe devices display lower noise than Si control ($V_{GT} = -3.5$ V, $L = 40$ μm, $W = 40$ μm, $T = 300$ K, $V_{DS} = -50$ mV) [1].

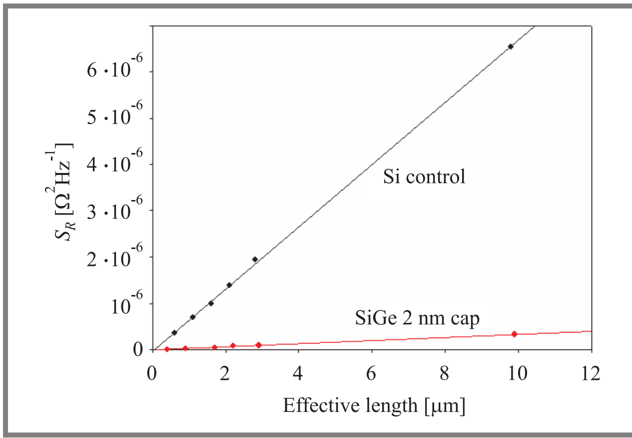


Fig. 4. Resistance fluctuations at 10 Hz as a function of gate length ($V_{GT} = -3.5$ V, $W = 40$ μm, $V_{DS} = -50$ mV). Extrapolation to zero gate length demonstrates that source and drain do not contribute noise. Noise is attributed to channel only [1].

tions S_R are plotted as a function of gate length L in Fig. 4, where a linear variation with L and extrapolation through the origin indicate that the noise may be attributed solely to the channel [4].

3. Comparison with noise models

We will now examine whether our noise data can be interpreted in terms of the commonly applied number fluctuation theory combined with correlated mobility fluctuations. There are three possible expressions that could be applied: for carrier number fluctuations (CNF) only [5, 6]

$$\frac{S_{I_D}}{I_D^2} = \left(\frac{R_N}{N_s}\right)^2 \frac{N_{ox}(E_F)\,kT}{WL\gamma f}. \quad (1)$$

If the associated oxide charge variations give rise to correlated mobility fluctuations (CMF1) [5, 6] then the normalised current noise PSD is

$$\frac{S_{I_D}}{I_D^2} = \left(\frac{R_N}{N_s} + \alpha\mu\right)^2 \frac{N_{ox}(E_F)\,kT}{WL\gamma f}. \quad (2)$$

Alternatively, the carrier number fluctuations can lead to fluctuations in total mobility (CMF2) [7]

$$\left(\frac{\Delta I_D}{I_D}\right)^2 = \left(\frac{\Delta N}{N}\right)^2 \left(1 + \frac{N}{\mu}\left(\frac{d\mu}{dN}\right)\right)^2, \quad (3)$$

where α is a factor associated with correlated mobility fluctuations, N_S is the sheet density of carriers in the inversion layer and N their total number. The term $R_N = C_{inv}/(C_{ox} + C_D + C_{inv})$ is associated with carrier number fluctuations, C_{inv}, C_{ox} and C_D are the inversion, oxide and depletion capacitance, respectively; μ is the effective mobility, $N_{ox}(E_F)$ the volume density of oxide traps per unit energy at the Fermi level, W the device width, and γ is the attenuation coefficient of the carrier wavefunction into the oxide.

We have tried to fit our data for the silicon control, using Eqs. (1), (2) or (3) and making the commonly used assumption [5] that the density of oxide trap states N_{ox} is independent of energy. This is shown in Fig. 5, where the normalised current PSD is plotted against current. The fact that none of these equations fits the data is attributed to the failure of this assumption.

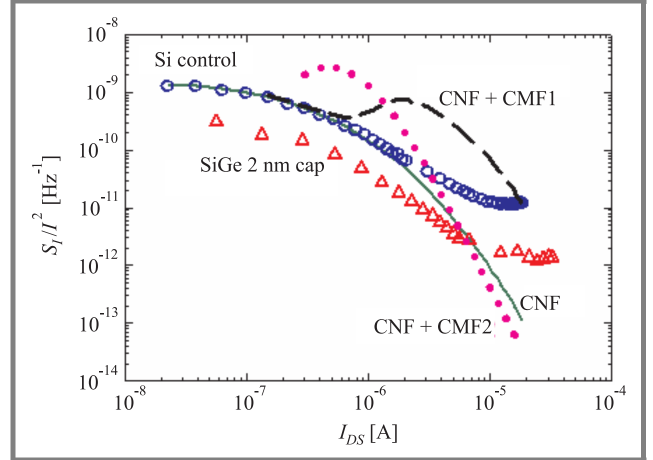


Fig. 5. Attempts to fit the experimental noise data from the Si control with models based on carrier number fluctuations and carrier mobility fluctuations as described in the text, assuming an energy independent density of oxide traps [1].

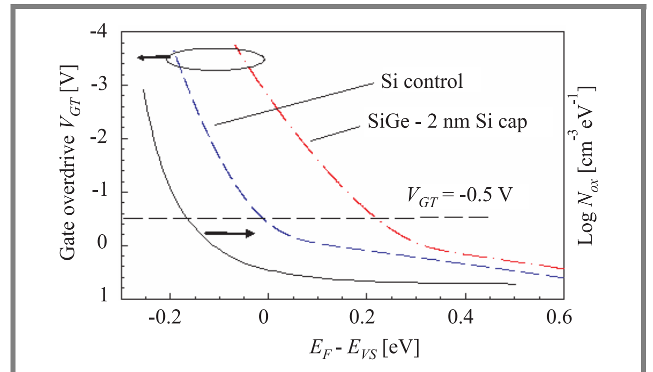


Fig. 6. Showing the displacement of the Fermi energy from the valence band edge as gate overdrive V_{GT} varies in the Si and SiGe devices, together with a schematic view of the variation in density of oxide traps across the band gap. For the same gate overdrive (dashed line shows case at $V_{GT} = -0.5$ V), the Fermi level in the SiGe device lies closer to mid-gap, where N_{ox} has a lower value, than in the Si control. Hence carrier tunnelling at E_F is into a reduced density of final states and the $1/f$ noise is therefore suppressed in the SiGe.

Chroboczek and Ghibaudo [5] have used equation Eq. (2) with N_{ox} constant, and on this basis attribute the suppression of noise in the buried channel to relatively weak scattering from oxide charge fluctuations, i.e., α small. While this is undoubtedly the case, we look for a more complete description of the behaviour. It is assumed that N_{ox} varies with energy roughly as shown schematically in Fig. 6. Since, for the same gate overdrive, the Fermi level in

the SiGe device lies closer to mid-gap than for the Si control, carriers at the Fermi level tunnel into a smaller number of oxide states for the alloy. Hence, the $1/f$ noise is suppressed in the alloy channel. Thus, there is the possibility that both mechanisms of noise suppression contribute. Thinking purely in terms of number fluctuations, S_I/I^2 should vary as $(g_m/I)^2$ if N_{ox} is constant [5]. This model is compared with the experimental results in Fig. 7,

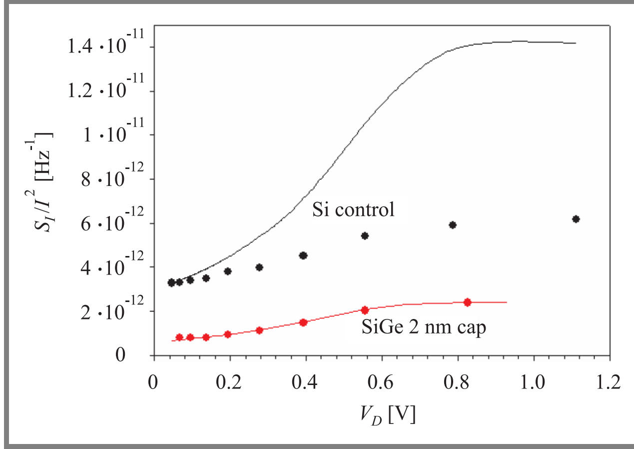


Fig. 7. Comparison of current noise PSD at 1 Hz (points) with a CNF model where $S_I/I^2 \propto (g_m/I)^2$ only. The discrepancy may be interpreted as N_{ox} varying with energy as in Fig. 6 ($V_{GT} = -3.5$ V, $W = 40$ μm, $L = 40$ μm).

where S_I/I^2 is plotted against drain voltage V_{DS}, and found to produce good agreement for the SiGe, but not for the Si devices. The effect of increasing the drain voltage is to raise the Fermi level closer to mid-gap, making the increase in noise with V_{DS} smaller than it would be if N_{ox} were independent of energy. If we accept Fig. 6 as a rough guide, then it is to be expected that the deviation from the theoretical $(g_m/I)^2$ curve is likely to be greater in Si, for which the variation of E_F spans the range where N_{ox} is rapidly varying, than in SiGe where N_{ox} only varies slowly over the energy range of interest. It is tentatively proposed that the observed behaviour provides further support for our model.

4. Dynamic threshold mode

Further improvements in performance are provided by DTMOS operation, with the gate (as opposed to the source) contact connected to the transistor body. As seen in Figs. 8, 9 and 10, the subthreshold swing, transconductance and maximum voltage gain all improve in both the SiGe and Si devices. The current noise in the SiGe and Si devices is independent of the mode of operation, as shown in Fig. 11. However, it is the input referred voltage noise $S_V = S_I/g_m^2$, shown in Fig. 12, which is important for circuit applications. Whereas this is equal to the flatband voltage fluctuation in the BT mode, and is therefore ultimately independent of g_m, in the DT mode of operation

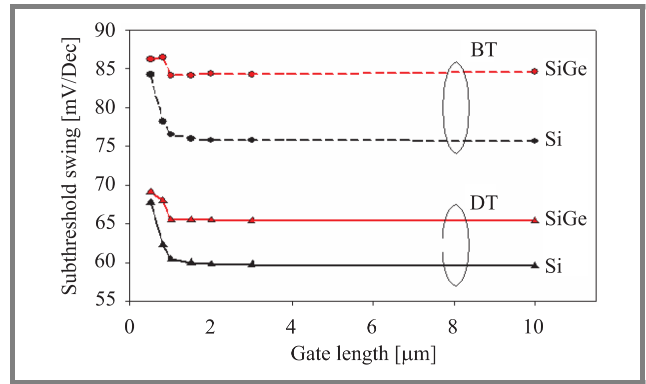


Fig. 8. Subthreshold swing in body tied and dynamic threshold modes as a function of gate length.

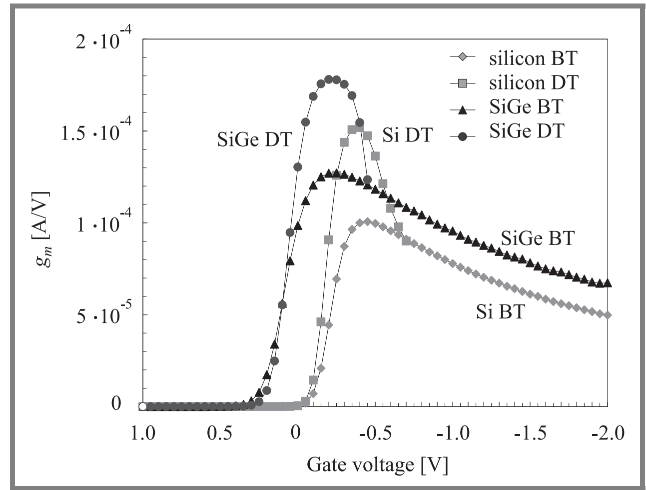


Fig. 9. Transconductance (g_m) as a function of gate voltage, showing particular enhancement in DT operation ($L = 3$ μm, $W = 160$ μm, $V_{DS} = -50$ mV).

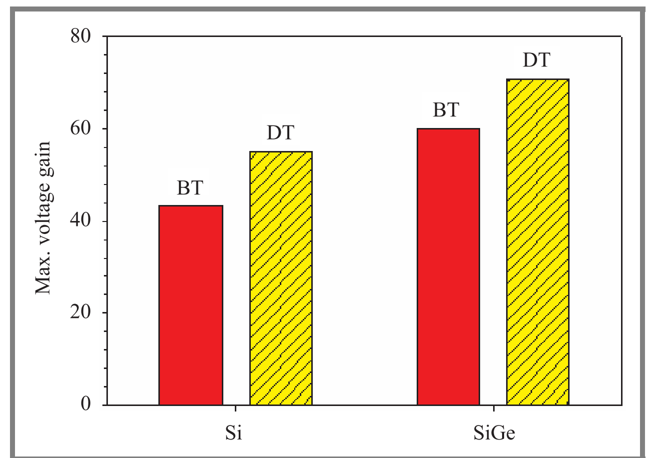


Fig. 10. Maximum voltage gain improved both by alloy channel and operation in DT mode.

Haendler *et al.* [8] predict a transconductance enhancement factor $(1 + \eta)$ with $\eta = dV_T/dV_G$, where V_T is the threshold voltage and V_G is the gate voltage. We can extract values of γ from the experimental data in two

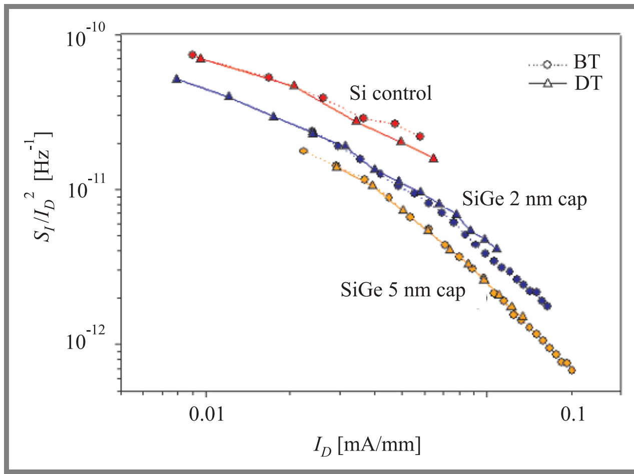


Fig. 11. Normalised current noise power spectral density at 1 Hz showing no variation between BT or DT mode for any of the devices studied ($V_{DS} = -50$ mV, $L = 40$ μm, $W = 40$ μm).

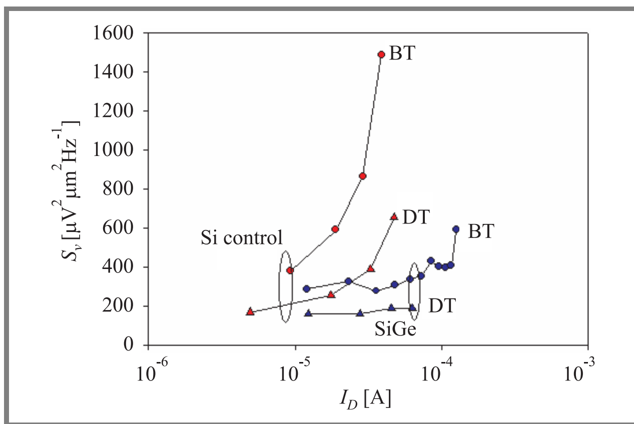


Fig. 12. Input referred voltage noise at 1 Hz, $S_v = S_I/g_m^2$. The enhanced g_m reduces S_v ($L = 3$ μm, $W = 160$ μm, $V_{DS} = -50$ mV).

ways: from the ratio of peak transconductance (Fig. 9), $\eta = 0.52$ for the Si device and 0.34 for SiGe, whereas from the threshold voltage shift with gate voltage, $\eta = 0.26$ and 0.30, respectively. A possible explanation for this discrepancy in the Si case lies in the original paper on DT action by Assaderaghi *et al.* [9], in which it is noted that the effective field is lowered by DTMOS action, which could lead to enhanced mobility and impact beneficially on transconductance. As pointed out by Takagi *et al.* [10], a SiGe device will have a larger value of η for the same threshold voltage, which points the way to further improvements in noise performance for a suitably designed DT mode device.

5. Conclusions

Measurements of $1/f$ noise in Si and Si/SiGe/Si MOSFETs have been compared with theoretical models of carrier number fluctuations due to tunnelling into the gate oxide. Analysis of the data suggests that the reduced noise in the heterostructure device as compared to Si is primarily associated with an energy dependent density of oxide trap states and a displacement of the Fermi level at the SiO_2/Si interface in the heterostucture relative to the Si control. High transconductance associated with dynamic threshold mode operation will further lower the input referred voltage noise, offering important benefits for circuit operation.

References

[1] M. J. Prest *et al.* "Low-frequency noise mechanisms in Si and pseudomorphic SiGe p-channel field-effect transistors", *Appl. Phys. Lett.*, vol. 85, pp. 6019–6021, 2004.

[2] M. J. Palmer *et al.*, "Effective mobilities in pseudomorphic Si/SiGe/Si p-channel metal-oxide-semiconductor field-effect transistors with thin silicon capping layer", *Appl. Phys. Lett.*, vol. 78, pp. 1424–1426, 2001.

[3] T. E. Whall and E. H. C. Parker, "Si/SiGe/Si pMOS performance – alloy scattering and other considerations", *Thin Solid Films*, vol. 368, pp. 297–301, 2000.

[4] X. Li and L. K. J. Vandamme, "1/f noise in series resistance of LDD MOSTs", *Solid State Electron.*, vol. 35, pp. 1471–1475, 1992.

[5] J. A. Chroboczek and G. Ghibaudo, "Low frequency noise in SiGe-base heterojunction bipolar transistors and SiGe-channel metal oxide semiconductor field effect transistors", *Microelectron. Reliab.*, vol. 40, pp. 1897–1903, 2000.

[6] A. Pacelli *et al.*, "Quantum effects on the extraction of MOS oxide traps by $1/f$ noise measurements", *IEEE Trans. Electron Dev.*, vol. 46, pp. 1029–1035, 1999.

[7] S. Martin, G. P. Li, H. Guan, and S. D'Souza, "A BSIM3-based flat-band voltage perturbation model for RTS and $1/f$ noise", *IEEE Electron Dev. Lett.*, vol. 21, pp. 30–33, 2000.

[8] S. Haendler, J. Jomaah, G. Ghibaudo, and F. Balestra, "Improved analysis of low frequency noise in dynamic threshold MOS/SOI transistors", *Microelectron. Reliab.*, vol. 41, pp. 855–860, 2001.

[9] F. Assaderaghi, D. Sinitsky, S. A. Parke, J. Bokor, P. K. Ko, and C. Hu, "Dynamic threshold-voltage MOSFET (DTMOS) for ultra-low voltage VLSI", *IEEE Trans. Electron Dev.*, vol. 44, pp. 414–422, 1997.

[10] T. Takagi, A. Inoue, Y. Hara, Y. Kanzawa, and M. Kubo, "A novel high performance SiGe channel heterostructure dynamic threshold pMOSFET (HDTMOS)", *IEEE Electron Device Lett.*, vol. 22, pp. 206–208, 2001.

Stephen M. Thomas completed a B.Eng. (Hons.) in electronic engineering at the University of Southampton (2005) and a research. M.Sc. degree at the University of Warwick (2006). Currently a Ph.D. student at Warwick, he is researching $1/f$ noise in strained silicon and Si/SiGe MOSFETs.

e-mail: Stephen.M.Thomas@warwick.ac.uk
University of Warwick
Coventry, CV4 7AL, UK

Tim J. Grasby was awarded a Ph.D. from the University of Warwick, UK, in 2000, where he is now the Chief Grower. He is an expert in epitaxial growth of silicon related materials by MBE and LP-CVD and holds several patents in the area of low defect SiGe virtual substrates.

e-mail: T.J.Grasby@warwick.ac.uk
University of Warwick
Coventry, CV4 7AL, UK

Terence E. Whall was awarded an honours B.Sc. from City University (London 1965) and a Ph.D. from Sussex University (1970). His major field of study has been the electrical properties of solids, including Kondo/spin glass alloys, molecular solids and transition metal oxides. Professor Whall holds a personal chair in physics at the University of Warwick, UK, where he heads the device physics activity looking at Si and Si/SiGe MOSFETs. He is the co-author of more than 200 research publications and has given 20 invited papers at international conferences.

e-mail: T.E.Whall@warwick.ac.uk
University of Warwick
Coventry, CV4 7AL, UK

David R. Leadley is an Associate Professor at the University of Warwick, UK, where he has worked in the NanoSilicon Group since 1995. Previously he spent 10 years studying magnetotransport in GaAs heterojunctions at the University of Oxford, with particular emphasis on magnetophonon interactions and the fractional quantum hall effect. Doctor Leadley has approx. 100 publications and conference proceedings to his name.

e-mail: D.R.Leadley@warwick.ac.uk
University of Warwick
Coventry, CV4 7AL, UK

Evan H. C. Parker is Professor of semiconductor physics at the University of Warwick, UK, and leader of the NanoSilicon Group since 1986. He has made many pioneering developments in MBE growth technology applied to Si related materials, including boron delta layers in silicon, high hole mobility pseudomorphic SiGe layers, limited area growth and compositionally graded SiGe virtual substrates. He has co-authored over 250 papers and holds a number of patents relating to epitaxial growth.

e-mail: EHCParker@warwick.ac.uk
University of Warwick
Coventry, CV4 7AL, UK

Martin J. Prest was awarded a Ph.D. from the University of Warwick, UK, in 2001, where he studied low-frequency noise and low-temperature transport in SiGe MOSFETs. His research interests then moved to strained silicon MOSFETs until 2004, since when he has been a research fellow in the Emerging Device Technology Group at the University of Birmingham. His latest work concerns fabrication of silicon MEMS actuators for application in tuning of superconducting microwave filters.

e-mail: m.j.prest@bham.ac.uk
Emerging Device Technology Research Centre
Department of Electronic, Electrical and Computer Engineering
School of Engineering
The University of Birmingham
Birmingham, B15 2TT, UK

Invited paper

On-wafer wideband characterization: a powerful tool for improving the IC technologies

Dimitri Lederer and Jean-Pierre Raskin

Abstract— In the present paper, the interest of wideband characterization for the development of integrated technologies is highlighted through several advanced devices, such as 120 nm partially depleted (PD) silicon-on-insulator (SOI) MOSFETs, 120 nm dynamic threshold (DT) voltage – SOI MOSFETs, 50 nm FinFETs as well as long-channel planar double gate (DG) MOSFETs.

Keywords—silicon-on-insulator, MOSFET, wideband characterization, microwave frequency, extraction techniques, small-signal equivalent circuit.

1. Introduction

Usually, at the early stage of advanced transistor development only the static behavior of novel devices is considered. Indeed, the I_{on}/I_{off} ratio, the subthreshold slope (S), the threshold voltage roll-off and the drain induced barrier lowering (DIBL) are the primary figures of merit that are extracted after each process run and which provide insights into the process quality.

The static analysis of the built transistors is then extended to the measurements of the gate transconductance (G_m) in saturation as well as the output conductance (G_d) or early voltage (V_{EA}). A typical feature of a node development scheme is that the dynamic behavior of advanced devices is most of the time considered only at the end of the fabrication process developments. Moreover, this dynamic analysis is usually limited to the frequency band of available vectorial network analyzer (VNA) leading to an unexplored frequency band beginning somewhere between a few Hz and 1 GHz.

However, a full frequency band analysis is precious for separating physical phenomena taking place in advanced MOS devices and characterized by clearly distinct time constants. Indeed, thermal and floating body effects typically appear from DC up to a few kHz in partially depleted (PD) SOI devices. In the MHz range the efficiency of the body contact in body-contacted (BC) PD SOI MOSFETs starts to degrade and untied carriers (i.e., minority carriers in the substrate or majority carriers in a floating body) can no longer follow the AC excitation. Finally, in the GHz range the relaxation time related to majority carriers is no more negligible and most of the parasitic capacitances and resistances specific to the 3D physical structure mainly affect the dynamic behavior of active devices.

Electrical characterization between DC and 110 GHz of advanced SOI MOS devices will be presented in this paper. Several direct characterization techniques have been developed for extracting small and large signal electrical models. Beside measurement facilities, our laboratory has several commercially available simulation software. We have already successfully simulated DC and RF behavior of MOSFETs such as fully and partially depleted, body-contacted and dynamic threshold MOSFETs in SOI technology. The 3D module of Atlas was used for simulating gate-all-around MOS as well as FinFETs. Combining the experimental characterization techniques and the simulation facilities, we have developed several macro-models based on a complete extrinsic small-signal equivalent circuit and an improved CAD model for the intrinsic device for those types of SOI MOSFETs.

2. Gate induced floating body effects in ultrathin oxide PD SOI MOSFET

Tunneling through gate oxides about 2-nm-thick is one of the major challenges faced by today's bulk-Si and SOI CMOS technologies. Gate tunneling does not only increase device leakage and power dissipation, but also leads to charging and discharging of PD SOI MOSFET body region causing floating body and device history-dependent effects [1, 2]. For n-type PD SOI MOSFETs, gate tunneling injects holes into the floating body thus increasing its voltage. This affects the device DC-characteristics and induces the so-called gate induced floating body effects (GIFBE) recently disclosed in [3, 4]. The impact of GIFBE on device DC transconductance was carefully examined in [3, 4], in which it was shown that gate tunneling induces a sharp second peak in the G_m-gate voltage curve of the studied devices. It was also demonstrated that the amplitude as well as the location of this peak are dependent on the measurement conditions and the device history.

Recently, we proposed a method based on wideband small-signal frequency measurements to characterize the dynamics of GIFBE [5]. This study was performed on n-channel 120 nm PD SOI MOSFETs with a silicon film and a buried oxide thickness of respectively 150 and 400 nm and a gate oxide with the thickness of approximately 2 nm. The DC measurements were performed with an HP4145 at

a drain bias of 50 mV and using a delay time of 2 s between successive DC points. The results of the DC measurements clearly display GIFBE (Fig. 1), since a second peak appears in the G_m versus V_g curve. This peak is associated with a DC body voltage increase caused by the injection of holes, which are generated by electron valence band tunneling through the gate oxide [3, 4]. The experimental set up used for the AC measurements is depicted in the inset of Fig. 1. A small amplitude (20 mV) AC signal at the gate electrode was superimposed on the DC bias using an HP3563 system analyzer. The AC variations of the drain current were recorded by measuring the potential drop across a resistance and an operational amplifier was also used to fix the DC drain bias at 50 mV. The AC measurements were performed between 0.1 Hz and 10 kHz for various DC bias conditions. The AC values are displayed in Fig. 1 for different frequencies.

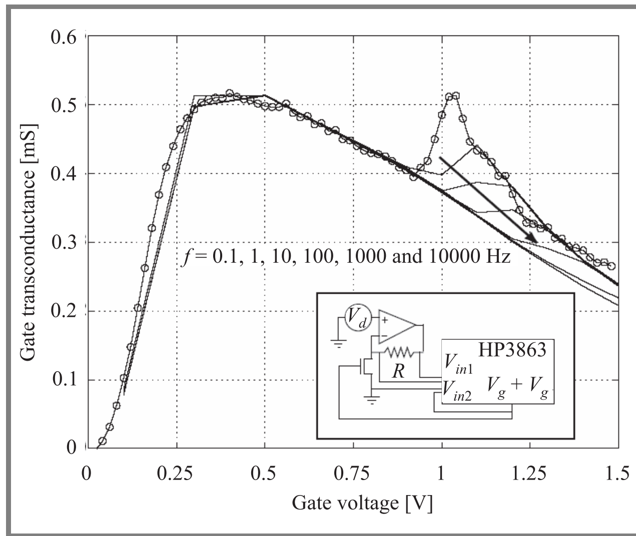


Fig. 1. Total gate transconductance G_m versus gate voltage for a floating body PD SOI MOSFET with $L = 2$ μm, $W = 60$ μm and $V_d = 50$ mV.

These AC variations of G_m are expected to have negative impact on low-frequency analog circuits requiring high gain and high accuracy. For example, in the case of operational amplifiers, they may cause gain reduction at very low frequencies and produce circuit instabilities due to reduced settling time constants.

As shown in details in [5] GIFBE are characterized by a very low frequency pole that is associated with the high impedance seen by the floating body toward external nodes. These effects can be relatively well reproduced with the BSIMSOI [11] model and with a simple equivalent AC circuit that includes the internal body node as well as gate tunneling (Fig. 2).

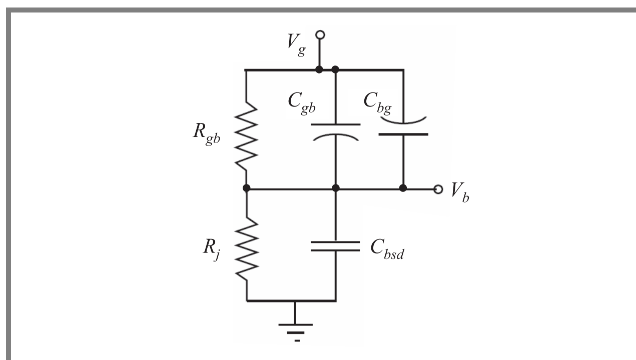


Fig. 2. Equivalent circuit of internal body node dynamic behavior due to AC gate excitation through ultrathin gate oxide for a floating body PD SOI MOSFET.

3. High frequency degradation of body-contacted PD SOI MOSFET output conductance

Partially depleted SOI technology suffers from non linearities in MOSFET output conductance introduced by the so called "kink effect". Under DC or low frequency conditions, this inconvenience has now been successfully overcome by several alternative solutions, such as fully depleted (FD), body tied (BT) or dynamic threshold (DT) MOS devices (Fig. 3) [6]. However, with the aggressive downscaling of channel length and SOI film thickness the efficiency of the body contact is reduced due to an increase of body resistance. In [7, 8], we analyzed the efficiency of the body contact from an output conductance point of view by comparing the G_d values measured on floating body (FB), BT, DT and FD devices of the 0.18 and 0.24 μm SOI technology node. The measurements were performed in DC and in the 100 kHz – 4 GHz frequency range.

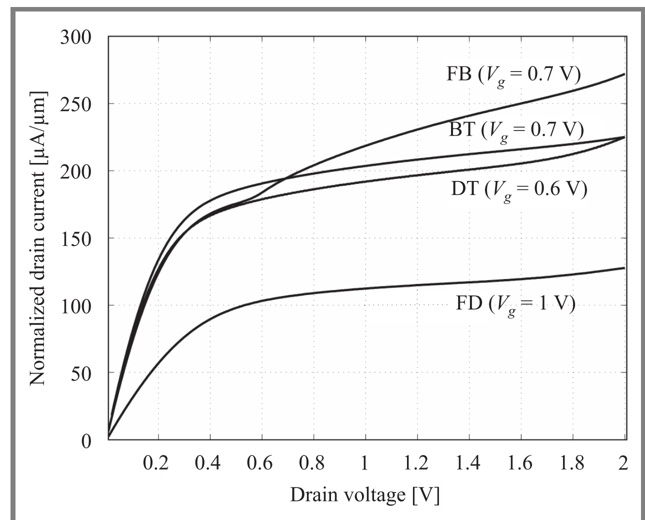


Fig. 3. Measured drain current I_d normalized by the total drawn gate width for FB, BT, DT PD SOI MOSFETs with $L = 0.18$ μm and FD SOI MOSFET $L = 0.24$ μm.

At low frequencies, Fig. 4 highlights the significant improvement obtained on G_d by connecting the body to the source or the gate, since both DT and BT structures exhibit a value of G_d more than two times lower than that of FB devices. However, it seems that this positive effect is lost at higher frequencies, since both DT and BT devices suffer from a 150% G_d degradation between DC and high frequency (4 GHz) levels.

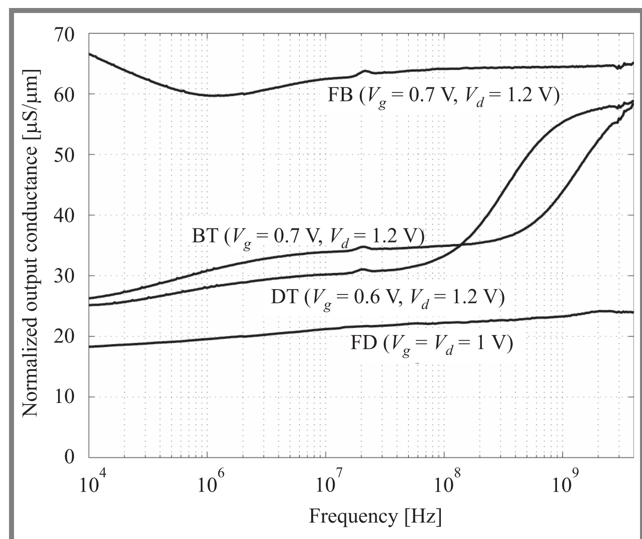


Fig. 4. Normalized output conductance measured as a function of the frequency for the different devices under analysis.

To explain and discuss these observations we proposed in [7, 8] small signal modeling of the PD devices seen from the drain terminal. The model includes the body region and its accesses to external nodes. It also accounts for AC impact ionization effects and AC charging of body potential. The model clearly points to the non zero value of the body resistance (R_{be}) as the origin of the G_d degradation. Reducing R_{be} by technological means would then provide an efficient way to reject this parasitic G_d increase to higher frequencies. In the next section an original method based on 3-port RF measurements [9] to accurately extract the body resistance in body-accessed PD SOI MOSFETs is presented.

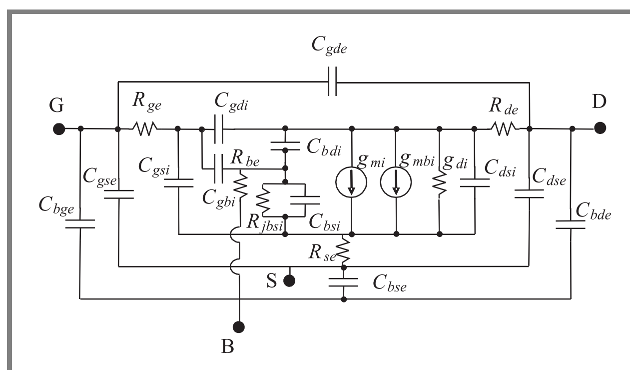


Fig. 5. Complete small signal equivalent circuit of the 3-port devices, including the external body node.

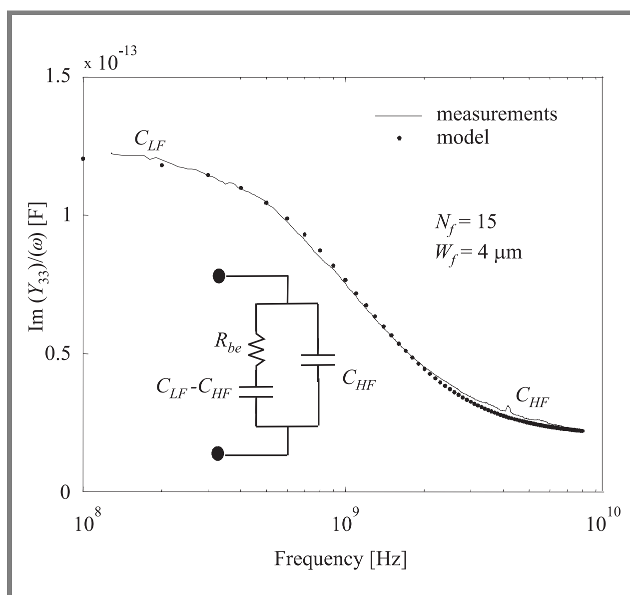


Fig. 6. Measured and simulated $\text{Im}(Y_{33})/(\omega)$ data as a function of frequency for $V_g = V_b = 0.6$ V and $V_d = 1.2$ V, inset: R-C model used.

4. Extraction of the body contact resistance

Accurate characterization of R_{be} is crucial to assess the efficiency of the body contact in a given technology. It can also be used to assess the validity of compact models such as BSIMSOI. The proposed method is based on the measurement of S-parameters over a wide frequency band under three-port configuration: the two classical ports are for the gate and the drain and the third port is connected to the body of the device.

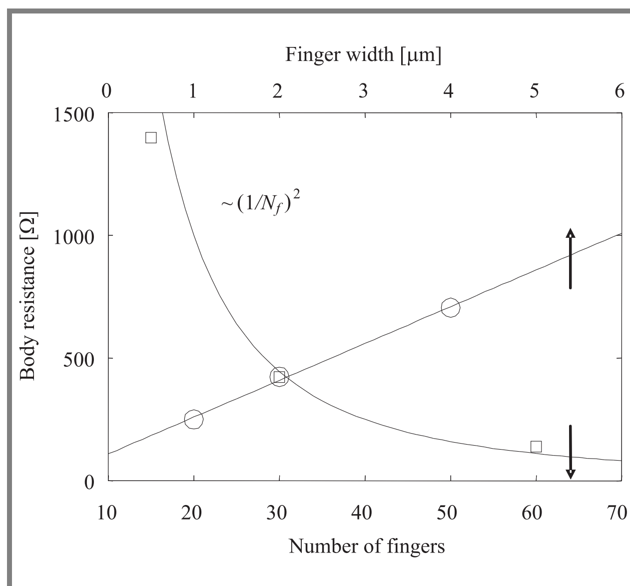


Fig. 7. Extracted values of R_{be} with respect to W_f and N_f.

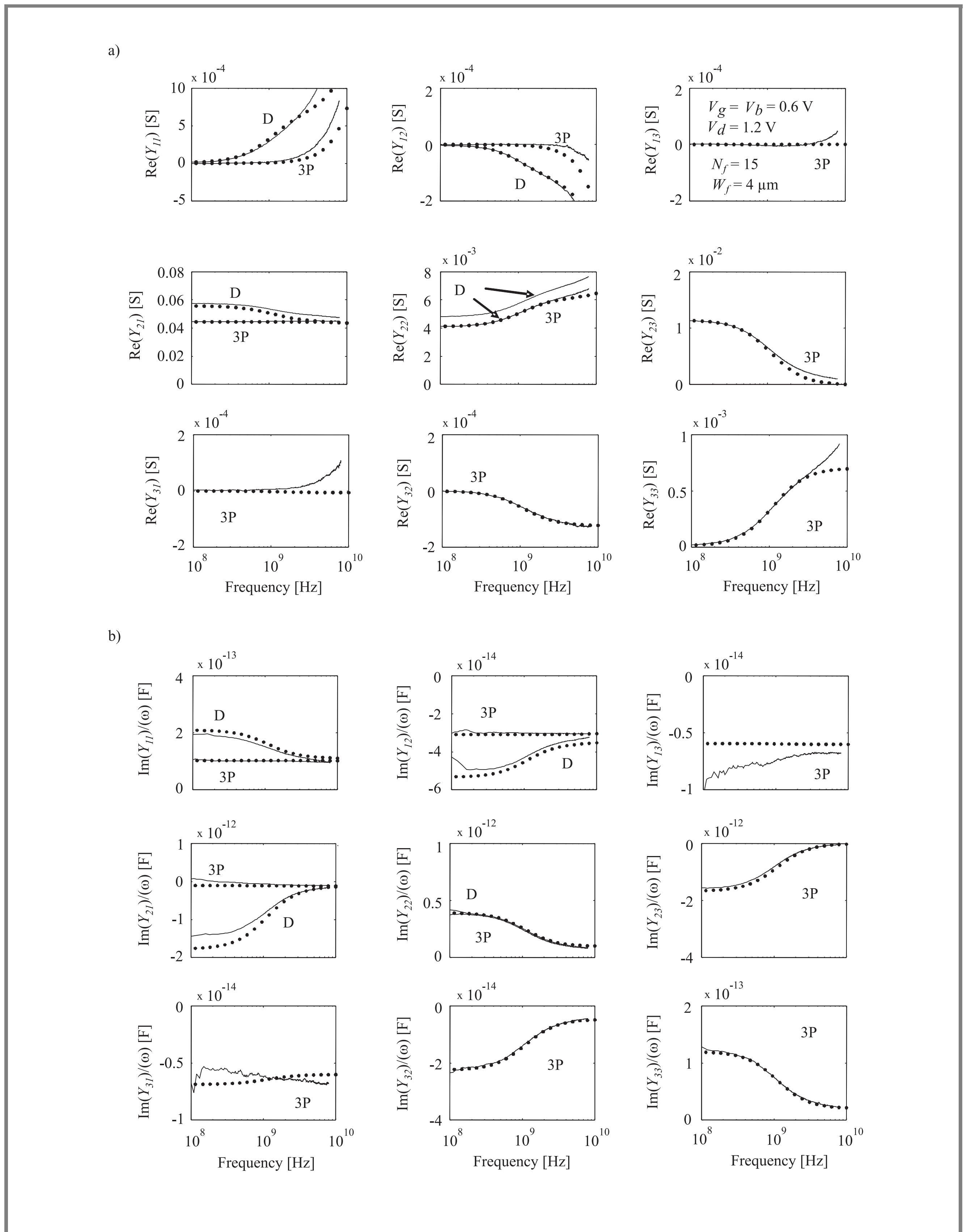


Fig. 8. Measured (straight lines) and simulated (dots) (a) Re(Y_{ij}) and (b) Im(Y_{ij})/(ω) for the 3-port device (3P) as well as for the DTMOSFET with body connected to gate.

The measured devices originate from a 0.13 μm SOI technology from ST Microelectronics, Grenoble. Different geometries (varying number of fingers (N_f) or finger width (W_f)) and body connections (either to the gate (DT) or to a third RF access) were considered. In all cases the body was accessed at both sides of the active area in order to reduce R_{be}. The S-parameters were measured with a multiport Rhode&Schwartz VNA up to 8 GHz and were de-embedded with a 3-port open subtraction method.

The modeling of the RF PD SOI devices was based on the small signal equivalent circuit presented in Fig. 5. It can be seen that both intrinsic (subscript i) and extrinsic (subscript e) elements were considered between each pair of electrodes, including the body (B). By extrinsic, the authors mean all parasitics that could not be removed during the de-embedding step.

The modeling of the body node was therefore achieved by considering all extrinsisic capacitances (C_{bge}, C_{bse} and C_{bde}), its access resistance (R_{be}), its intrinsic body-gate (C_{bgi}) as well as body-source (C_{bsi}) and body-drain (C_{bdi}) junction capacitances, and its intrinsic body-source junction resistance (R_{jbsi}). The body-drain junction resistance was neglected due to reverse biasing of this junction when devices operate in saturation. The extraction of R_{be} was performed by analyzing the output admittance seen from the body node terminal (Y_{33}). Indeed, as shown in Fig. 6 the value of $\mathrm{Im}(Y_{33})/(2\pi f)$ clearly exhibits a pole-zero pair dependence, which is typical of the simple R-C network included in Fig. 6. In this circuit, the high frequency (C_{HF}) value of $\mathrm{Im}(Y_{33}/(2\pi f)$ is simply the sum of all extrinsic capacitances seen from the body terminal (i.e., $C_{HF} = C_{be} = C_{bse} + C_{bde} + C_{bge}$) while its low frequency value (C_{LF}) is the sum of all intrinsic and extrinsic capacitances seen from the body terminal (i.e., $C_{LF} = C_{bsi} + C_{bdi} + C_{bgi} + C_{be}$). In these conditions, it is immediate to see that the pole expression is given by

$$f_p = \frac{1}{2\pi R_{be}(C_{LF} - C_{HF})} = \frac{1}{2\pi R_{be}(C_{bsi} + C_{bgi} + C_{bdi})} \quad (1)$$

and is therefore strongly dependent on the value of R_{be}.

By fitting the R-C network to the measurement results this value could then be extracted. Figure 7 shows the value of R_{be} obtained for devices with varying N_f (with constant $N_f W_f$ product) and W_f (with constant N_f) values. It is seen that Eq. (1) R_{be} decreases approximately as $(1/N_f)^2$ (similarly to R_{ge} [10]) while Eq. (2) a linear dependence on W_f is observed. These two trends agree with predictions made by the scalable BSIMSOI model [11], further supporting the validity of the extraction method.

The extrapolated value of R_{be} obtained for $W_f = 0$ ($\sim$150 Ω) therefore provides a good estimation of the parasitic resistance associated with the body interconnects (R_{bout}) outside the active region. The figure shows that its contribution is not negligible with regards to the overall value of R_{be}. Indeed, normalized values of R_{be} and R_{bout} were found to be close to 21 kΩ/μm $\cdot$ finger and 2.25 k$\Omega \cdot$ finger, respectively, assuming only one contact per finger. The body node characterization was further achieved by extracting (C_{bge}, C_{bgi}) and (C_{bde}, C_{bdi}) from the high frequency and low frequency parts of $\mathrm{Im}(Y_{31})$ and $\mathrm{Im}(Y_{32})$, respectively. The body-source capacitances were then obtained from C_{LF} and C_{HF} and the back gate transconductance (g_{mbi}) was given by $\mathrm{Re}(Y_{23})$. The value of R_{jbsi} could theoretically be extracted from $\mathrm{Re}(Y_{33}) = (R_{be} + R_{jbsi})^{-1}$ at low frequencies but was too high (> 100 kΩ) to be accurately measured. It was therefore assumed to be infinitely high in the model. The rest of the device parameters were obtained with the Cold-FET method [10], while R_{se}, R_{de} and R_{ge} were obtained with a method depicted in [12]. Figures 8a and 8b show that an accurate modeling of both real and imaginary parts of the Y parameters is obtained up to ~ 4 GHz for a 4 μm-wide device with $N_f = 15$, $V_g = 0.6$ V, $V_d = 1.2$ V, $V_b = 0.6$ V. The model was then further tested by connecting the body to the gate terminal, forming a two-port network and compared with measurement results obtained on a DTMOSFET with the same geometry. It is seen in Figs. 8a and 8b that a good agreement is obtained between measured and simulated data, despite a small output conductance difference observed for the DTMOSFET, which could be due to a subtle bias shift.

5. Extraction of parasitic capacitances and resistances of FinFET

The dynamic performance of FinFETs was investigated on 50 nm-long RF n-doped devices with 2 gate fingers of 5, 15, 25 and 30 μm-width. The fin width and the fin spacing were set to 55 and 100 nm, respectively. It is expected that these devices operate in fully depleted regime for such fin width, which was further confirmed by DC measurements. The S-parameters of the RF devices were measured in saturation ($V_d = 1.2$ V and $V_g = V_g(G_{m\max})$) up to 110 GHz

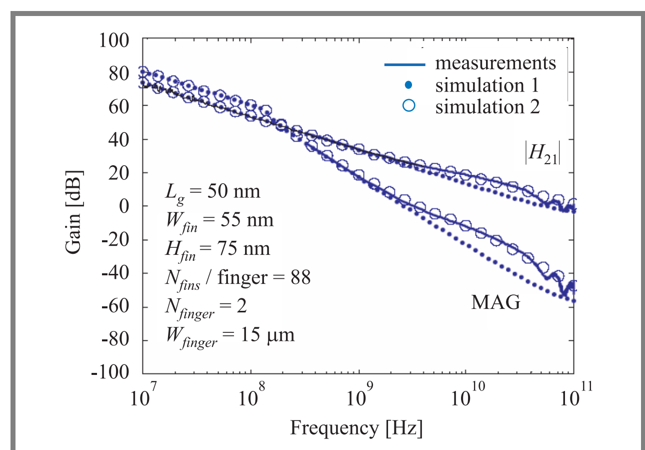


Fig. 9. Measured and simulated $|H_{21}|$ and maximum available gain (MAG) for a RF FinFET. The models considered were a classical equivalent circuit for FET (simulation 1, dots) and that of Fig. 10 (simulation 2, circles).

and the pad parasitics were removed from the raw data with an open subtraction method. In Fig. 9, the measured and modeled current gains are presented. The curve denoted as "simulation 1" was obtained with a classical equivalent circuit for FET (including a simple RC network to model the transistor input impedance) and "simulation 2" for the model shown in Fig. 10 in which a distributed parasitic network (R_{g1}, C_{gs1}, C_{gd1} and R_{g2}, C_{gs2}, C_{gd2}) at the transistor input is considered. It can be seen in Fig. 9 that this improved model (circles) can closely reproduce the frequency behavior of the FinFET gain curves over the whole frequency band. The physical origin of this distributed RC at the input of the FinFET is related to the non-optimized gate silicidation (high gate resistance) and higher gate capacitance due to polysilicon residues along the silicon fins [13]. These lines of residual polysilicon are due to an incomplete polysilicon etch in the buried oxide (BOX) recess when the polysilicon gate is patterned by resist trimming. These technological problems were solved and cutoff frequencies higher than 100 GHz have been recently measured for 60 nm FinFETs [14].

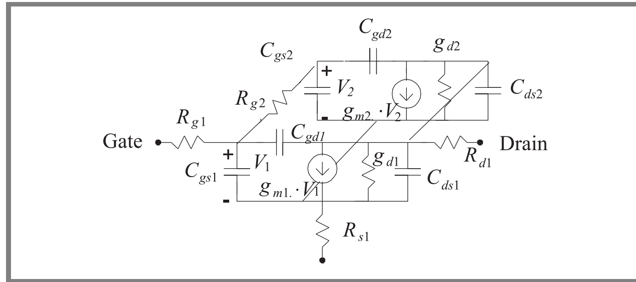


Fig. 10. Small signal equivalent circuit used for modeling the RF FinFET devices.

6. Backgate resistance extraction of planar double gate SOI MOSFET

In the clean room facilities of Universite catholique de Louvain (UCL) we have built and measured long-channel (20 down to 1 μm) planar double gate (DG) SOI MOSFETs. The fabrication process of these DG devices is based on the transfer of a high quality thin silicon film above a pre-etched cavity in an oxide layer [15]. As shown in Fig. 11 the G_m DC value of the DG device is indeed twice higher than that of the single gate (SG) device.

As expected by the high sheet resistance of the unsilicided top polysilicon gate, a severe drop of G_m occurs above a few GHz. However, we also observe another kink in the G_m curve of the DG transistor at around a few kHz. After the extraction of a complete equivalent circuit over this wide frequency band, it was demonstrated that this G_m drop is related to the higher resistance of the back gate. This dissymmetry between the front and back gate polysilicon resistivity can be explained by the poor diffusion of doping atoms into the polysilicon filling the cavity. In-situ doping of the polysilicon is then required to avoid this high backgate parasitic resistance.

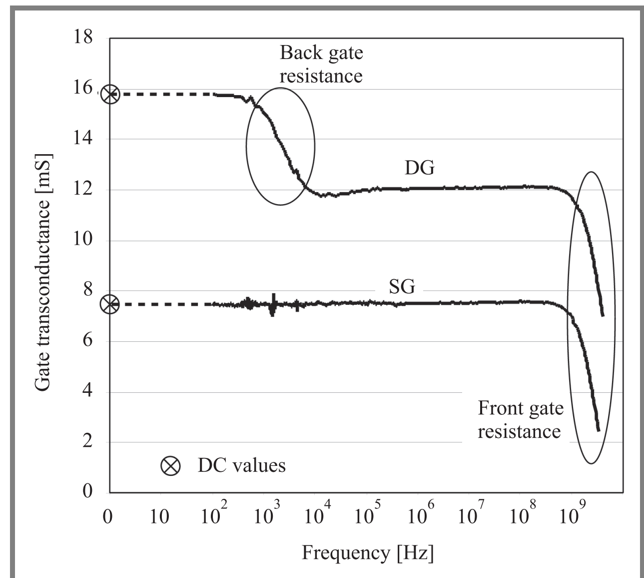


Fig. 11. Gate transconductance G_m of the measured SG and DG devices versus frequency.

7. Original de-embedding technique for high input impedance MOS devices

Usually, the high frequency performance of transistors is extracted through on-wafer S-parameter measurements performed with a vector network analyzer. First, off-wafer calibration is undertaken at probe-tips using classical cali-

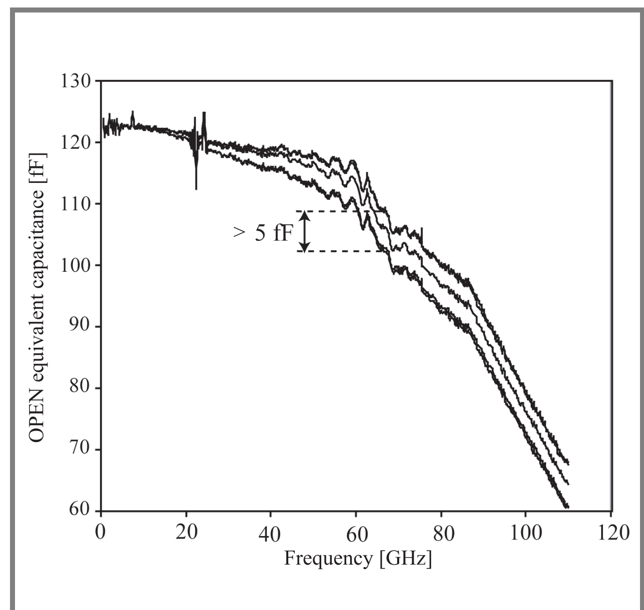


Fig. 12. OPEN capacitance versus frequency for various probing contacts.

bration techniques (LRM, LRRM, SOLT) on an alumina substrate. Then, an on-wafer de-embedding procedure is required to obtain the S-parameters of the active device. This is generally done through the use of dedicated on-wafer RF test structures (OPEN, SHORT, THRU, LINES) associated with the device under test (DUT). First of all, these RF test structures consume a non-negligible area on the wafer. Secondly, for extracting the intrinsic performance of the measured DUT, we have to probe several tests structures (i.e., OPEN, SHORT, THRU, LINES) and these multiple probings could lead to uncertainties directly related to the non-repeatability of the contact from device to device. This is illustrated in Fig. 12, which presents the equivalent capacitance of the same OPEN obtained for 5 different probing contacts. As we may observe, a variation of up to 5 fF of this capacitance can be obtained from one probing contact to another. Similar variations of the de-embedded RF structure Y-parameters were observed from one die to another on the same wafer, due to a classical spread of the technological parameters.

At the same time, MOS devices are aggressively scaled down to improve their RF performance. This contributes to a decrease of their input intrinsic capacitance, which may therefore become very small in comparison with the parasitic capacitance associated with the probing and access pads.

In order to analyze the impact of the OPEN capacitance variations on the de-embedded RF performance of advanced MOS devices, we performed some simulations using Agilent ADS. We considered a device composed of 30 gate fingers of 60 nm channel length and 500 nm gate width each, resulting in a input intrinsic capacitance of around 12 fF.

Considering a variation of the OPEN equivalent capacitance (C) from -2 fF to 2 fF, the extracted transient frequency f_T varies from 195 GHz ($C = 2$ fF) and 218 GHz ($C = 0$ fF) to 250 GHz ($C = -2$ fF). This corresponds to a variation of almost $\pm 15\%$ of the extracted f_T.

Such dispersion may be avoided when using our new de-embedding technique, in which only one probing contact is needed for performing the de-embedding of the measured DUT. This new and recently proposed [16] de-embedding technique allows us to extract the high frequency performance of a DUT without any associated RF test structure. The method is based on the behavior of the field effect transistors under ColdFET bias conditions [17, 18]. When the device is biased with V_{gs} below threshold ($V_{gs} \ll V_{th}$) and $V_{ds} = 0$ V, its intrinsic part may be neglected and the general equivalent circuit can then be simplified to the one shown in Fig. 13.

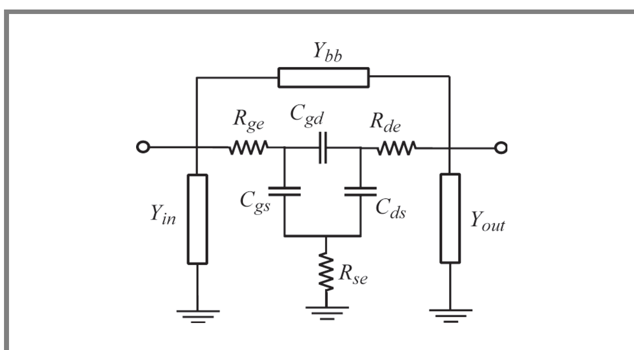


Fig. 13. Small signal equivalent circuit of the device in ColdFET bias conditions.

In addition, if we make the assumption that the transistor is symmetrical ($C_{gs} \approx C_{gd}$), the Y_{in}, Y_{out} and Y_{bb} admittances equivalent to the RF access structure, can be extracted from the measured Y_{COLD} parameters of the device in ColdFET bias conditions:

$$Y_{in} = Y_{COLD11} + 2.Y_{COLD12} + Y_{bb}\,, \tag{2}$$

$$Y_{out} = Y_{COLD22} + Y_{COLD12}\,, \tag{3}$$

$$Y_{bb} = Y_{COLD22} - Y_{COLD11} - Y_{COLD12}\,. \tag{4}$$

The RF access structure can then be de-embedded from the measured Y_{MOS} admittance matrix of the device at the bias point of interest, for instance in saturation regime ($V_{gs} > V_{th}$, $V_{ds} = V_{dd}$) with:

$$Y_{COR11} = Y_{MOS11} - Y_{in} - Y_{bb}\,, \tag{5}$$

$$Y_{COR12} = Y_{MOS12} + Y_{bb}\,, \tag{6}$$

$$Y_{COR21} = Y_{MOS21} + Y_{bb}\,, \tag{7}$$

$$Y_{COR22} = Y_{MOS22} - Y_{out} - Y_{bb}\,. \tag{8}$$

Figure 14 presents the current gain and maximum available gain obtained for a 130 nm-channel length SOI MOSFET device with 30 gate fingers of 4 and 1 μm-width, respectively. In addition, the results obtained considering a classical OPEN de-embedding procedure are plotted. A very good agreement can be observed between these two de-embedding techniques for both the current and maximum available gain.

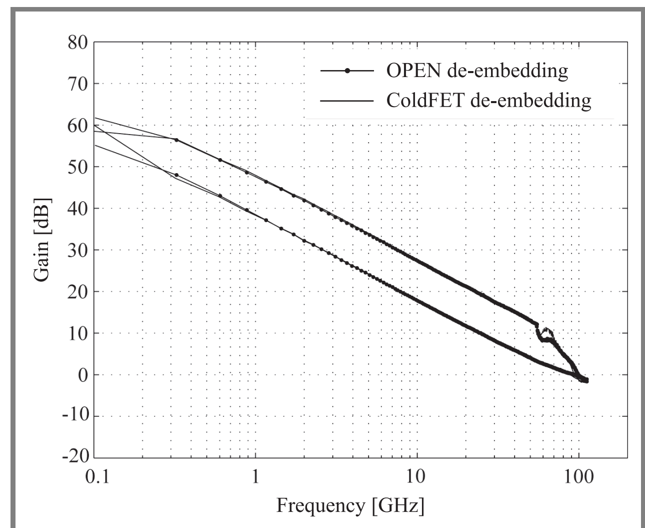


Fig. 14. Comparison of the current gain H_{21} and the maximum available gain using a classical OPEN de-embedding technique and the new ColdFET technique (NMOSFET $30 \times 4 \times 0.13$ μm^2, $V_{gs} = 0.59$ V, $V_{ds} = 1.2$ V).

To conclude, one should notice that the ColdFET de-embedding results were obtained without any other structure than the device under test and with the same probing contact for depletion and saturation regimes. This technique may save up to 50% of the wafer area (one dedicated OPEN per device is usually considered for RF test structures). Furthermore, with the downscaling process of advanced devices dimensions, the accuracy of the extracted RF performance may be affected by the dispersion and contact quality on the wafer. Our new technique allows us to break through these problems due to only one probing contact to extract the RF performance of the DUT.

8. Conclusion

From these examples, it is quite obvious that a wideband electrical characterization has to be considered at the early stage of technology development. The direct extraction of physical parameters such as parasitic capacitances, resistances, relaxation of carriers, body contact, etc., that cannot be extracted under static bias conditions is of great importance in the improvement cycle of any advanced technology. The results presented here also indicate that as transistors dimensions are continuously shrinking, it is also crucial to develop new measurement and characterization techniques in order to maintain high accuracy of the extracted parameters of advanced devices.

References

[1] P. Su, S. K. H. Fung, W. Liu, and C. Hu, "Studying the impact of gate tunneling on dynamic behaviors of partially-depleted SOI CMOS using BSIMPD", in *Proc. Int. Symp. Qual. Electron. Des.*, San Jose, USA, 2002, pp. 75–76.

[2] R. V. Joshi, C. T. Chuang, S. K. H. Fung, F. Assaderaghi, M. Sherony, I. Yang, and G. Shahidi, "Effects of gate-to-body tunneling current on PD/SOI CMOS SRAM", in *Proc. VLSI Tech. Dig.*, Kyoto, Japan, 2001, pp. 75–76.

[3] J. Pretet, T. Matsumoto, T. Poiroux, S. Cristoloveanu, R. Gwoziecki, C. E. Raynaud, A. Roveda, and H. Brut, "New mechanism of body charging in partially depleted SOI-MOSFETs with ultra-thin gate oxides", in *Proc. ESSDERC*, Florence, Italy, 2002, pp. 515–518.

[4] A. Mercha, J. M. Rafi, E. Simoen, E. Augendre, and C. Claeys, "Linear kink effect induced by electron valence band tunneling in ultrathin gate oxide bulk and SOI MOSFETs", *IEEE Trans. Electron Dev.*, vol. 50, no. 7, pp. 1675–1682, 2003.

[5] D. Lederer, D. Flandre, and J.-P. Raskin, "AC behavior gate transconductance for ultra thin gate oxide PD SOI MOS", *IEEE Electron Dev. Lett.*, vol. 25, no. 2, pp. 104–106, 2004.

[6] Y.-C. Tseng *et al.*, "AC floating body effects and the resultant analog circuit issues in submicron floating body and body grounded SOI MOSFETs", *IEEE Trans. Electron Dev.*, vol. 46, no. 8, pp. 1685–1692, 1999.

[7] D. Lederer, D. Flandre, and J.-P. Raskin, "High frequency degradation of body-contacted PD SOI MOSFET output conductance", *Semicond. Sci. Technol.*, no. 20, pp. 469–472, 2005.

[8] M. Dehan and J.-P. Raskin, "Dynamic threshold voltage MOS in partially depleted SOI technology: a wide frequency band analysis", *Elsev. Sci., Perg., Solid-State Electron.*, vol. 49, pp. 67–72, 2005.

[9] D. Lederer, O. Rozeau, and J.-P. Raskin, "Wideband characterization of body-accessed PD SOI MOSFETs with multiport measurements", in *Proc. IEEE Int. SOI Conf.*, Honolulu, Hawaii, USA, 2005, pp. 65–66.

[10] J.-P. Raskin, R. Gillon, J. Chen, D. Vanhoenacker, and J.-P. Colinge, "Accurate SOI MOSFET charaterisation at microwave frequencies for device performance optimisation and analogue modelling", *IEEE Trans. Electron Dev.*, vol. ED-45, no. 5, pp. 1017–1025, 1998.

[11] BSIMSOI project, http://www-device.eecs.berkeley.edu/ bsimsoi

[12] A. Bracale, V. Ferlet-Cavrois, N. F. D. Pasquet, J.-L. Gautier, J.-L. Pelloie, and J. du Port de Poncharra, "A new approach for SOI devices small-signal parameters extraction", *Anal. Integr. Circ. Sig. Proc.*, vol. 25, no. 2, pp. 157–169, 2000.

[13] D. Lederer, V. Kilchytska, T. Rudenko, N. Collaert, D. Flandre, A. Dixit, K. De Meyer, and J.-P. Raskin, "FinFET analog characterization from DC to 110 GHz", *Elsev. Sci. Perg. Solid-State Electron.*, vol. 49, pp. 1488–1496, 2005.

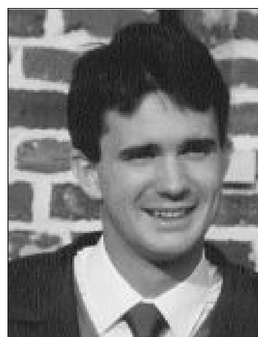
[14] D. Lederer, B. Parvais, A. Mercha, N. Collaert, M. Jurczak, J.-P. Raskin, and S. Decoutere, "Dependence of FinFET RF performance on fin width", in *Proc. 6th Top. Meet. Silic. Monol. Integr. Circ. RF Syst.*, San Diego, USA, 2006, pp. 8–11.

[15] T. M. Chung, B. Olbrechts, D. Flandre, U. Södervall, S. Bengtsson, and J.-P. Raskin, "Planar double-gate SOI MOS devices by wafer bonding over pre-patterned cavities", in *Proc. EUROSOI Worksh.*, Grenoble, France, 2006, pp. 111–112.

[16] G. Pailloncy and J.-P. Raskin, "New de-embedding technique based on Cold-FET measurement", in *36th Eur. Microw. Week (36th EuMW) Eur. Microw. Conf. (EuMC)*, Manchester, UK, 2006.

[17] R. Anholt and S. Swirhum, "Measurement and analysis of GaAs MESFET parasitic capacitances", *IEEE Trans. Microw. Theory Techn.*, vol. 39, no. 7, pp. 1247–1251, 1991.

[18] G. Dambrine, A. Cappy, F. Heliodore, and E. Playez, "A new method for determining the FET small-signal equivalent circuit", *IEEE Trans. Microw. Theory Techn.*, vol. MTT-36, no. 7, pp. 1151–1159, 1988.



Dimitri Lederer received the M.S.E.E. from Universite catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 2000. He joined the Microwave Laboratory of UCL in 2001, where he has worked as a teaching Assistant in the fields of electromagnetics, transmission line theory and RF active circuit design. He completed a Ph.D. (2006) on the fabrication and characterization of high resistivity SOI substrates with increased resistivity for system-on-a-chip and monolithic microwave integrated circuit applications. His thesis research also focused on the characterization of advanced RF SOI MOSFETs, such as dynamic threshold MOSFET (DTMOSFET) and RF FinFETs, his main fields of interest. He is the author and co-author of more than 50 papers in refereed journals and symposia proceedings.
e-mail: dimitri.lederer@tyndall.ie
Microwave Laboratory
Université catholique de Louvain (UCL)
Place du Levant, 3
B-1348 Louvain-la-Neuve, Belgium

Jean-Pierre Raskin was born in Aye, Belgium, in 1971. He received the industrial engineer degree from the Institut Supérieur Industriel d'Arlon, Belgium, in 1993, and the M.Sc. and Ph.D. degrees in applied sciences from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1994 and 1997, respectively. From 1994 to 1997, he was a research engineer at the Microwave Laboratory, Université catholique de Louvain, Belgium. He worked on the modeling, characterization and realization of MMIC's in silicon-on-insulator (SOI) technology for low-power, low-voltage applications. In 1998, he joined the EECS Department of The University of Michigan, Ann Arbor, USA. He has been involved in the development and characterization of micromachining fabrication techniques for microwave and millimeter-wave circuits and microelectromechanical transducers/amplifiers working in hard environments. Since January 2000, he is Associate Professor at the Microwave Laboratory of the Université catholique de Louvain, Louvain-la-Neuve, Belgium. His research interests are the modeling, wideband characterization and fabrication of advanced SOI MOSFETs as well as micro- and nanofabrication of MEMS/NEMS sensors and actuators. He is a IEEE Senior Member, EuMA Associate Member and Member of the Research Center in Micro and Nanoscopic Materials and Electronic Devices of the Université catholique de Louvain. He is author or co-author of more than 250 scientific articles.
e-mail: raskin@emic.ucl.ac.be
Microwave Laboratory
Université catholique de Louvain (UCL)
Place du Levant, 3
B-1348 Louvain-la-Neuve, Belgium

Invited paper

Evaluation of MOSFETs with crystalline high-*k* gate-dielectrics: device simulation and experimental data

Florian Zaunert, Ralf Endres, Yordan Stefanov, and Udo Schwalke

Abstract— The evaluation of the world's first MOSFETs with epitaxially-grown rare-earth high-*k* gate dielectrics is the main issue of this work. Electrical device characterization has been performed on MOSFETs with high-*k* gate oxides as well as their reference counterparts with silicon dioxide gate dielectric. In addition, by means of technology simulation with TSUPREM4, models of these devices are established. Current-voltage characteristics and parameter extraction on the simulated structures is conducted with the device simulator MEDICI. Measured and simulated device characteristics are presented and the impact of interface state and fixed charge densities is discussed. Device parameters of high-*k* devices fabricated with standard poly-silicon gate and replacement metal gate process are compared.

Keywords— crystalline high-k gate dielectric, rare-earth oxide, praseodymium oxide, gadolinium oxide, damascene metal gate, CMP, CMOS process, TSUPREM4, MEDICI, interface state density, carrier mobility, remote coulomb scattering.

1. Introduction

One of the key challenges of modern MOS technologies is engineering of the gate stack. Lateral dimensions have to be further reduced to the deca-nanometer region in order to continue with the superior performance improvement of CMOS circuits. However, the thickness of the gate dielectric is the main limiting factor at present. Scaling down for sub 32 nm technologies will require an equivalent oxide thickness (EOT) well below 1 nm.

In order to replace SiO_2, mainly amorphous and poly-crystalline high-*k* materials, such as HfO_2 and ZrO_2, have been investigated as alternative gate dielectrics. However, a major drawback of these materials is the need of a SiO_2 buffer layer between the silicon surface and the high-*k* dielectric, which increases the equivalent oxide thickness and eliminates the chance to achieve an EOT well below 1 nm. Therefore epitaxially-grown crystalline dielectrics with a lattice constant near to silicon have been proposed as alternatives. Praseodymium oxide (Pr_2O_3) was the first epitaxially grown rare-earth material to be investigated as gate dielectric [1, 2]. Fully functional MOSFETs have been fabricated for the first time by our group [3, 4] using conventional poly-silicon gate electrodes. Electrical properties and discussion of device properties can be found in [5–9]. Very recently, devices manufactured using a replacement gate process have been successfully fabricated with crystalline gadolinium oxide (Gd_2O_3) as dielectric and metal gate electrodes [10, 11]. Metal gate electrodes do not suffer from gate depletion, like poly-silicon electrodes, thus remote coulomb scattering (RCS) is eliminated and carrier mobility is expected to improve. However, metal gates seem to introduce other undesirable effects that nullify the advantages associated with the elimination of RCS. In [12] compressive strain in the channel induced by the metal gate and surface roughness is proposed to describe this behavior.

In this work, MOS device properties associated with the implementation of crystalline rare-earth oxides as gate dielectrics in MOS transistors are discussed. Measurements and simulated device characteristics of replacement gate MOSFETs with crystalline rare-earth high-*k* dielectric and metal gate are presented and compared to high-*k* devices manufactured with poly-silicon gate. This includes detailed measurements, including charge-pumping results [10] of prototype devices as well as comprehensive process and device simulation of these structures. By comparison of the fabricated devices with quasi-ideal simulated structures, a better understanding of the impact of high-*k* material on device properties is obtained. One key issue of this work is to investigate the cause of the low surface mobility of MOSFETs manufactured with high-*k* dielectrics and to study the impact of gate stack processing, i.e., conventional poly-Si gate versus metal gate damascene technology.

2. Device fabrication and structure

Metal gate damascene NMOSFETs have been fabricated with the replacement-gate technique [10, 11]. The gate stack contains a tungsten electrode and a dielectric composed of Gd_2O_3. A brief outline of the replacement gate process is given in Fig. 1. According to the EXIT-GATE (i.e., gate first) approach [13], dummy gate structures are formed on blank silicon wafers covered with a nitride/poly-Si/nitride sandwich. After source/drain implantation, a SiO_2 layer is deposited by menas of plasma enhanced chemical vapour deposition (PECVD) and the ion implantation is activated by a brief RTA-annealing step at 1000°C, which also stabilises the top nitride layer against the following chemical mechanical polishing (CMP) step. Then the CVD oxide is polished down to the top of the dummy gate using a CMP process. The dummy gate is removed completely by wet etching and all harsh pro-

cess steps (RIE, high-temperature anneals) are done. Now, the final gate stack with either high-k oxide or SiO_2 reference dielectric is manufactured. Molecular beam epitaxy (MBE) is used for growing epitaxially a thin Gd_2O_3 layer and subsequent in situ metal deposition (tungsten) is performed. Metal damascene Gd_2O_3 MOSFETs with two different EOTs (53 Å and 21 Å) have been fabricated. Finally, a standard back-end metallization process completes the MOSFET fabrication.

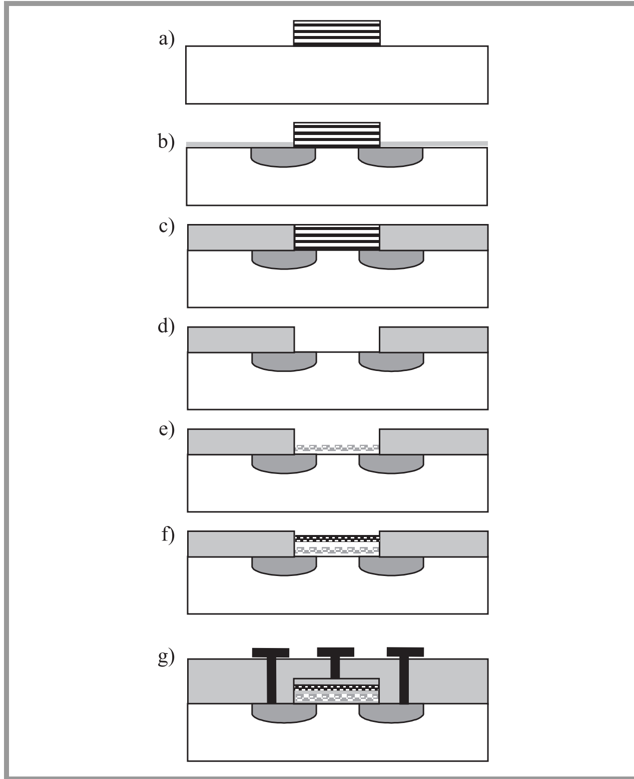


Fig. 1. Outline of the CMP-based metal gate damascene fabrication process: (a) dummy gate stack formation; (b) source/drain formation; (c) alignment oxide formation; (d) dummy gate stack removal; (e) gate dielectric growth; (f) metal gate formation; (g) back-end processing.

The SiO_2 reference devices were also manufactured in replacement-gate technology, having tungsten-titanium electrodes instead and a nitrided RTO-SiO_2 dielectric thickness of $t_{ox} = 50$ Å. Except for the gate dielectric, the reference devices are expected to be largely comparable to the Gd_2O_3-devices with an EOT of 53 Å, because the tungsten-titanium electrodes have almost the same work function of about 4.55 eV as pure tungsten [14]. Accordingly, there should be no significant difference in threshold voltage or other parameters generated by the gate electrodes. They main difference between both structures is expected to originate from the different gate dielectric, which may affect the silicon-insulator interface and the channel region close to this interface.

Previously, Pr_2O_3 high-k NMOSFETs have been fabricated using a conventional poly-silicon gate process technology [3] with a k-value of approx. 30, corresponding to an EOT of about 20 Å–25 Å. These devices are also used for comparison. Further details on the fabrication of the Pr_2O_3 high-k NMOSFETs can be found in [3] and [4].

3. Evaluation methodology

All devices examined in this work have nominal gate lengths of $L = 4\ \mu$m and a gate width of $W = 100\ \mu$m. Due to the relatively long gates, it is guaranteed, that short channel effects do not influence device properties. The measurements include standard characterization methods for determination of the carrier mobility, sub-threshold slope, threshold voltage and transconductance. Carrier mobility of MOSFETs is derived using two different methods. First, by calculating the effective mobility μ_{eff} from the channel conductance at low drain voltage ($V_{ds} = 50$ mV) using Eq. (1). Secondly, by operating the MOSFET in saturation condition ($V_{gs} = V_{ds}$) and using Eq. (2)

$$\mu_{eff} = \frac{L}{W} \frac{g_d}{C_{ins}\left(V_{gs} - V_T\right)} \quad \text{with } g_d = \frac{\partial I_d}{\partial V_{ds}}, \tag{1}$$

$$\mu_{sat} = \frac{L}{W} \frac{2 I_d}{C_{ins}\left(V_{gs} - V_T\right)^2} \tag{2}$$

with L and W channel length and width, channel conductance g_d and the area-independent gate capacitance C_{ins}.

The saturation mobility μ_{sat} is slightly lower than the effective mobility μ_{eff}, because of the presence of the vertical electric field which degrades the carrier mobility in the channel, when driving the MOSFET in saturation. Hence, the effective mobility derived from the channel conductance is usually preferred to describe device mobility, although even here some simplifications have to be made, especially in the estimation of the charge density in the channel [15–17]. Threshold voltage V_T is derived from the channel conductance curves, using Eq. (3)

$$V_T = V_{gs} - \frac{g_d}{K} \frac{L}{W} \tag{3}$$

with $K = \mu C_{ins}$.

The saturation method was used additionally, yielding slightly different values, as discussed above. With Eq. (4) one can calculate the threshold voltage from the saturation curve:

$$V_T = V_{gs} - \sqrt{\frac{2 I_d}{K} \frac{L}{W}}. \tag{4}$$

To compare the simulation data with the measured curves one-to-one, it is necessary to implement the same evaluation procedures, i.e., the simulated device results and experimental data are evaluated using the same methods as described above.

With the technology simulator TSUPREM4 the structures for high-k and reference MOSFETs have been implemented referring to the original process parameters. In Figs. 2 and 3 the simulation mesh and the structure of the devices, respectively, are shown. Subsequent to the process simulation,

the electrical evaluation is performed on these computer-generated structures with the device simulator MEDICI.

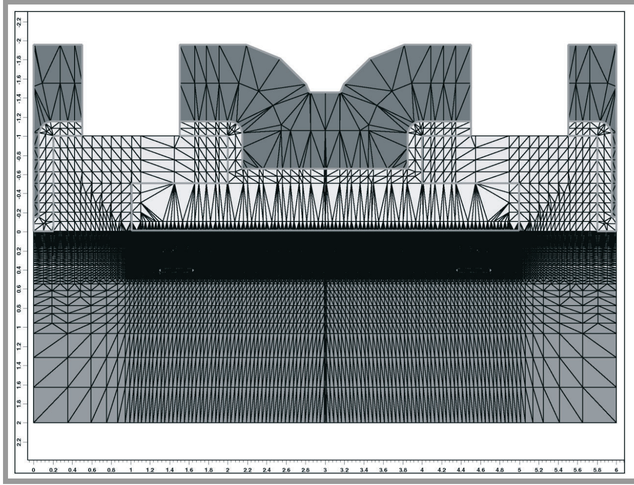


Fig. 2. Example of the simulation mesh of 4 μm n-channel MOSFET with damascene metal gate.

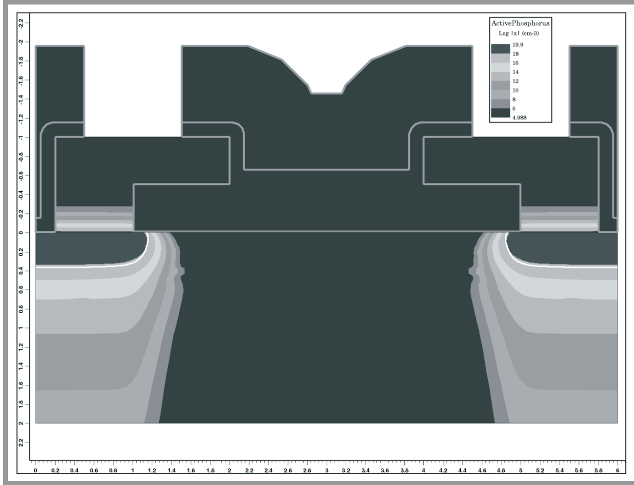


Fig. 3. Example of the simulated MOSFET cross-section with source/drain doping profiles (phosphorus).

It is found, that the selection of the mathematical mobility model is the key parameter, which has the strongest impact on the MOSFET characteristics. The main challenge is to combine several models that are needed for different regions of the semiconductor material. For the semiconductor bulk, a standard analytical model is used in this work, which is based on empirical data [18] and fitting parameters [19]:

$$\mu_{eff} = \mu_{\min} + \frac{\mu_{\max}\left(\frac{T}{300}\right)^{\nu} - \mu_{\min}}{1 + \left(\frac{T}{300}\right)^{\xi}\left(\frac{N(x,y)}{N_{ref}}\right)^{\alpha}}\,, \tag{5}$$

where $\mu_{\min}$ and $\mu_{\max}$ are the minimum and maximum carrier mobilities used by the model, $N(x,y)$ is the local impurity concentration, N_{ref} a given reference concentration and T the absolute temperature, again, α, ν and ξ are fitting parameters.

However, to describe the insulator-semiconductor interface region in detail, appropriate surface models are still required for taking into account scattering effects at impurity atoms and interface roughness:

$$\mu_{S,\perp} = G_{surf}\,\frac{\mu_{eff}\left(\frac{T}{300}\right)^{-EX_0}}{1 + \left(\frac{E_{eff\perp}}{E_{ref}}\right)^{EX}}\,, \tag{6}$$

where μ_{eff} is the calculated mobility of the analytical model, $E_{eff\perp}$ is the vertical electric field created by the gate voltage and E_{ref} a reference value, the factor G_{surf} is used to describe effects like surface roughness and crystal strain, which are not included analytically, EX and EX_0 are fitting parameters, the parameter used for modification of the surface mobility μ_S in this work was only G_{surf}.

Furthermore, a model is used to include carrier velocity saturation in the presence of high parallel electric fields. Equation (7) clarifies the correlation between parallel electric field and carrier mobility [18]

$$\mu_{S,\parallel} = \frac{\mu_{eff}}{\left(1 + \left(\frac{\mu_{eff}E_{\parallel}}{v_{sat}}\right)^{\beta}\right)^{\beta^{-1}}} \tag{7}$$

with $E_{\parallel}$ being the parallel field generated by the drain voltage, v_{sat} the saturation velocity [20], μ_{eff} the low field mobility and a fitting parameter β.

Since the sub-V_T slope is affected by the density of interface states, the first attempt was made to adjust the simulated curves to the measured ones, by varying the interface state density in the simulation. This is done by matching the simulated sub-V_T slope to the measured slope. With the sub-V_T slopes being identical, the corresponding interface state density is recorded and is cross-checked with experimental data from charge-pumping measurements.

With this value fixed, the next step in the calibration flow is the adjustment of the carrier mobility in the channel. The mobility, obtained by electrical measurements, is so low that it can only be explained by a high interface-state density. In order to obtain measured mobility values consistent with the simulation, interface roughness and surface scattering have to be included. Because no universal analytical model describing all the above mentioned effects consistently is available to us, the parameter G_{surf}, is used to adjust the surface quality to the experimental conditions.

The last step of the iteration process comprises the fine tuning of the threshold voltage. Especially for the high-k devices in this work, an increased threshold voltage due to the high interface-state density is observed. In addition to the interface-state density, the fixed charges in the dielectric and the gate electrode work function also have an impact on the threshold voltage. For a given metal the value of the work function can vary some tenths of an electronvolt

depending on the manner of depositing the electrode layer on the dielectric and the contact properties between metal and insulator. The alteration of parameters has to be done carefully, because a change in a single device parameter does not only affect one device property. For example, changing the interface state density has an effect on both threshold voltage and sub-threshold slope of the devices. Usually, several adjustment cycles are needed to fix all parameters with a reasonable accuracy.

4. Results

The results provided by the simulations with default parameters show much higher carrier mobility and better sub-threshold slopes than the experimental data of the fabricated devices. Obviously, these very first prototype devices do not feature ideal characteristics and possess substantial potential for improvement.

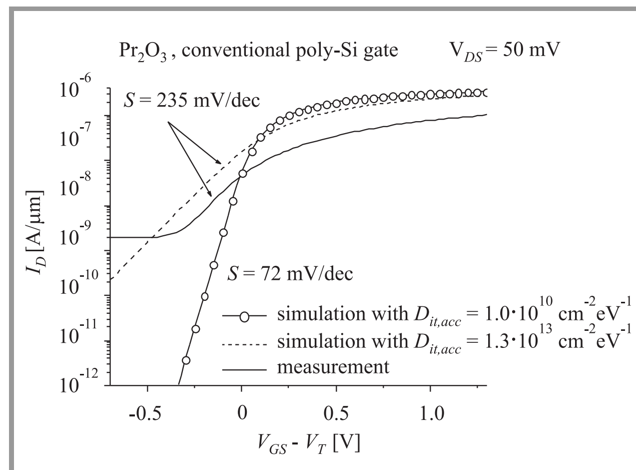


Fig. 4. Comparison of measured and simulated sub-threshold characteristic of MOSFET with Pr_2O_3 dielectric.

As can be seen in Fig. 4, MOSFETs with Pr_2O_3-dielectric exhibit a sub-threshold slope $S = 235$ mV/dec, which coincides with the simulation data, when assuming an acceptor interface-state density of about $D_{it,acc} = 1.3 \cdot 10^{13}$ cm^{-2}eV^{-1}, which is a rather high value. Without interface states the simulation yields a much better sub-threshold slope of $S = 72$ mV/dec. The interface-state density also influences the mobility of the device as derived from the channel conductance curves, as shown in Fig. 5. From device measurements a value of $\mu_{eff} = 40$ cm^2/Vs is obtained, which is too low to be explained only by interface states. Simulations with the above interface-state density lead to mobilities of $\mu_{eff} = 130$ cm^2/Vs. As mentioned previously, in order to achieve consistent mobility values between measurements and simulation, the G_{surf} parameter has to be adjusted.

Figure 6 compares sub-threshold slopes of the measured metal-gate SiO_2 reference devices with the simulation data for two different cases. Again, the slope of the simu-

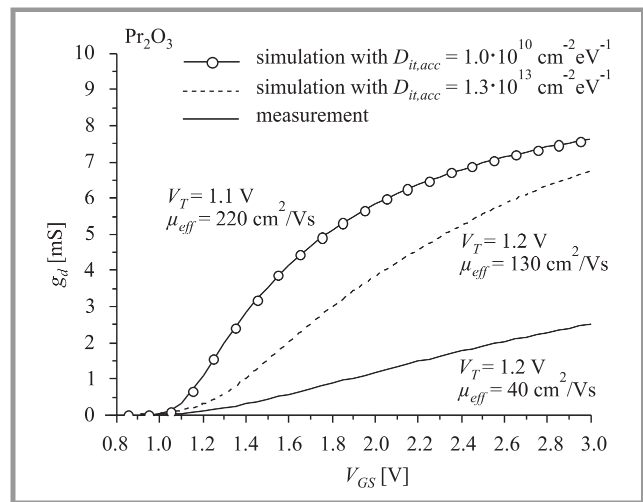


Fig. 5. Comparison of measured and simulated threshold voltage and mobility of Pr_2O_3-MOSFET.

lated device with ideal interface is much better than that of the measured one. The simulated reference devices exhibit a sub-threshold slope close to the theoretical limit of 60 mV/dec. Adjustment of the interface-state density of the simulated device, makes its sub-threshold slope become

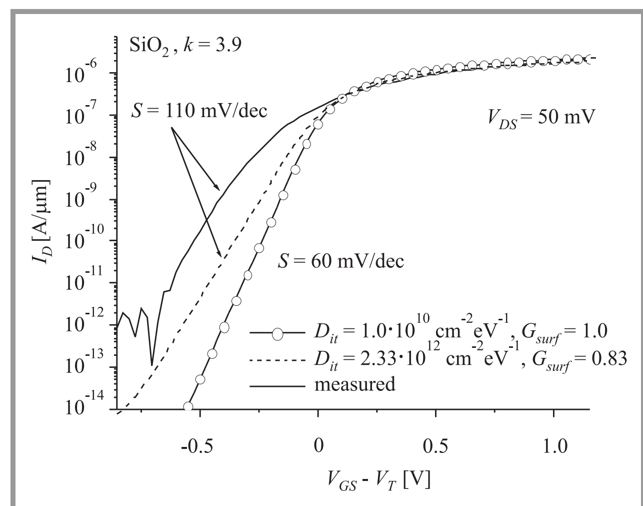


Fig. 6. Comparison of sub-threshold slopes of measurement data and simulated curves with ideal and degraded interfaces for the silicon dioxide reference devices.

very close to the measured one. However, a sub-threshold slope of 110 mV/dec corresponds to a high interface-trap density in the range of 10^{12} cm^{-2}eV^{-1}, which is unexpected for SiO_2. Currently, we investigate whether the metal gate blocks H-atoms from penetrating to the Si-SiO_2 interface to cure interface traps during the forming gas anneal. Interestingly, when looking at Fig. 7 a very similar situation can be found for the Gd_2O_3 high-k devices regarding interface trap density. The simulation identifies the theoretical feasible optimal value for the sub-threshold slope of 80 mV/dec, whereas the real devices

posses slopes of 150 mV/dec. To which extent forming gas anneal is beneficial also for Gd_2O_3 high-k devices is currently under investigation. Unfortunately, degraded leakage properties have been reported [21] after forming gas anneal, therefore this approach may not be feasible.

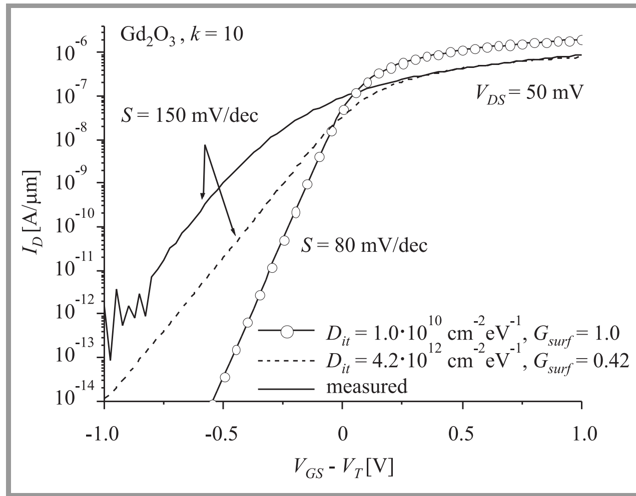


Fig. 7. Comparison of sub-threshold slopes of measurement data and simulated curves with ideal and degraded interfaces for the high-k devices.

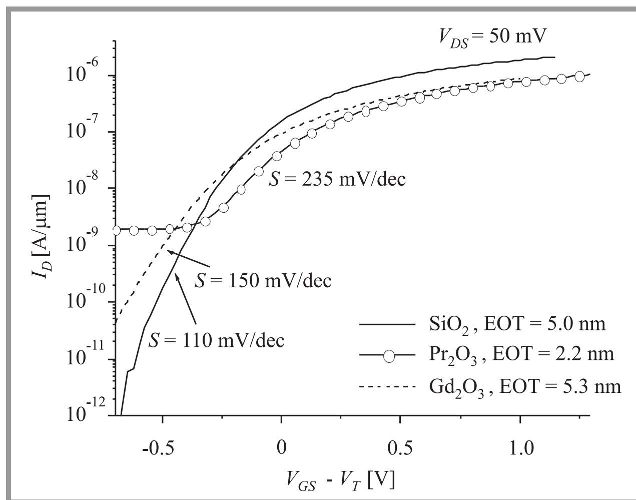


Fig. 8. Measured sub-threshold characteristics of fabricated Pr_2O_3-, Gd_2O_3- and SiO_2 devices.

On the other hand, when comparing the metal gate damascene Gd_2O_3 devices to the conventionally processed poly-silicon gate Pr_2O_3 NMOSFETs, a much better subthreshold slope (Fig. 8) and mobility value (Fig. 9) is observed. Obviously, for the replacement gate devices the interface quality is improved and/or less process damage has occurred than in the case of the devices manufactured with the conventional poly-Si CMOS process. This may result from the "gentle", CMP-based gate formation process of the metal gate devices, where the gate dielectric is not exposed to damaging process steps like reactive

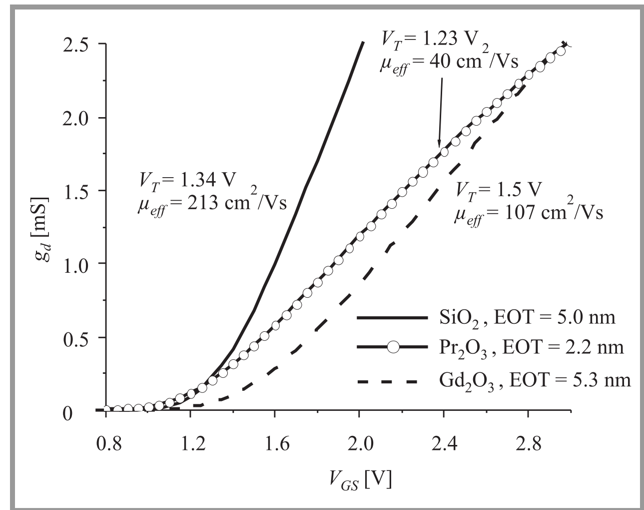


Fig. 9. Measured transconductance and extracted mobility values of fabricated Pr_2O_3-, Gd_2O_3- and SiO_2 devices.

ion etching (RIE) and high-temperature anneals as in the conventional CMOS process. From the adjusted simulated curves the density of interface acceptor states is determined to be $D_{it}(SiO_2) = 2.33 \cdot 10^{12}$ cm^{-2}eV^{-1} and $D_{it}(Gd_2O_3) = 4.2 \cdot 10^{12}$ cm^{-2}eV^{-1}. These values are consistent with the results obtained from energy resolved charge-pumping measurements [10].

Figure 10 compares the channel conductivity of measured and simulated Gd_2O_3 devices. Devices with ideal and degraded interfaces are simulated as can be seen in Fig. 10. Devices with ideal interfaces possess a 2.7-times

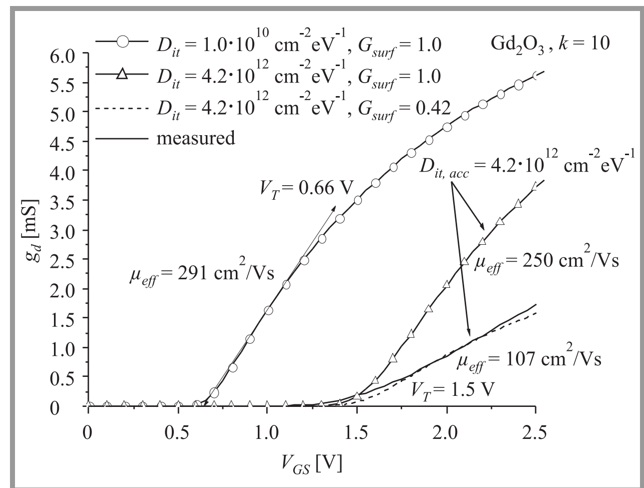


Fig. 10. Comparison of channel conductivities, threshold voltages and carrier mobility of measured data and simulated curves of Gd_2O_3-devices illustrating the effect of ideal and degraded interfaces by simulation.

higher carrier mobility than the measured ones. One factor, which degrading mobility is interface-state density, as mentioned above. However, even for interface-state density of $D_{it} = 4.2 \cdot 10^{12}$ cm^{-2}eV^{-1} a mobility of $\mu_{eff} = 250$ cm^2/Vs

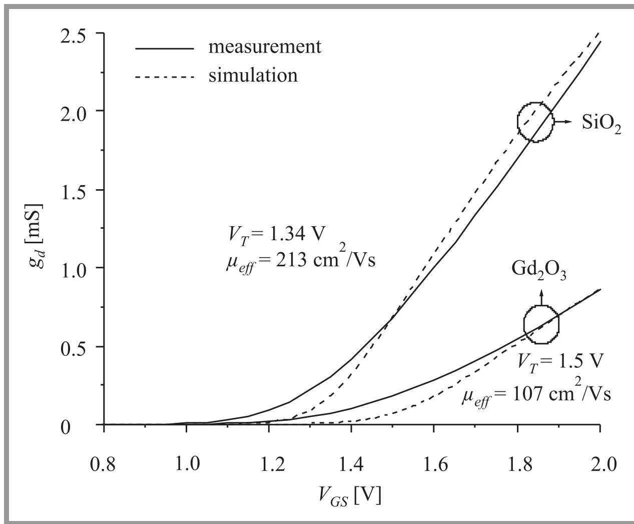


Fig. 11. Comparison of channel conductivities, threshold voltages and carrier mobility of SiO_2 reference and Gd_2O_3 high-k devices.

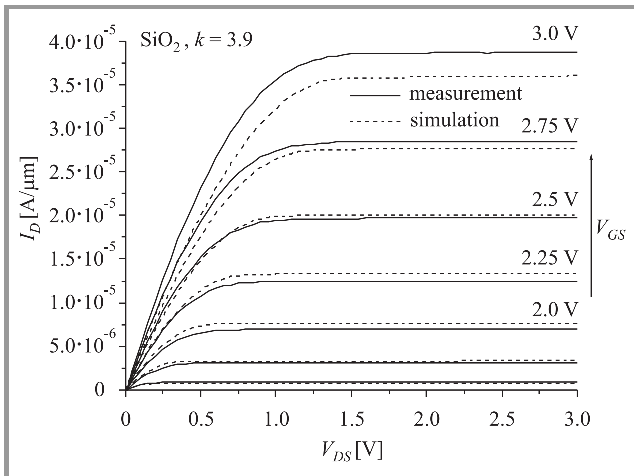


Fig. 12. Measured and simulated output characteristics of reference MOSFETs with SiO_2 gate dielectric.

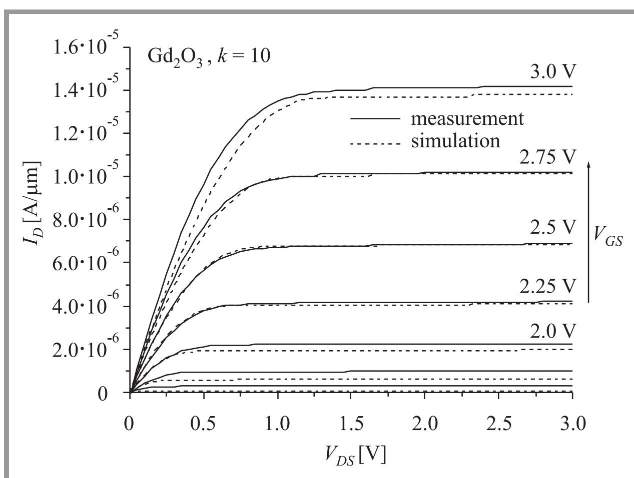


Fig. 13. Measured and simulated output characteristics of MOSFETs with Gd_2O_3 high-k gate dielectric.

is deduced from simulation with otherwise perfect Si-surface, i.e., $G_{surf} = 1$. To obtain the measured mobility value of $\mu_{eff} = 107$ cm^2/Vs, the surface quality needs to be degraded, additionally by lowering G_{surf} to 0.42. From Fig. 11 the threshold voltage and effective mobilities for the high-k devices are extracted from the slopes of the curves using Eqs. (1) and (3) to $V_T = 1.5$ V and $\mu_{eff} = 107$ cm^2/Vs, respectively. It is apparent that the mobility of the high-k devices is only half of that of the reference devices with SiO_2 dielectric of $\mu_{eff} = 213$ cm^2/Vs. Worse interface quality arousing from the manufacturing process may be the reason. One major advantage of the thermal oxidation of silicon is the fact that the Si-SiO_2 interface has never been in contact with the outside world, because of the silicon dioxide growing into the bulk material during the oxidation step. This is in contrast to the high-k devices, where the crystalline high-k material is grown on top of the original silicon surface by MBE. To improve the carrier mobility, the interface state density and surface quality, such as roughness or crystal strain [12], have to be improved.

After adjusting all important parameters output characteristic are simulated with the parameters from above and the plots are depicted in Figs. 12 and 13. Because of the considerable gate length, almost no channel length modulation is noticeable as expected. The agreement between simulation and measurement is very reasonable for the high-k devices. However, for the SiO_2 reference devices a deviation at higher gate voltages can be seen. A possible explanation for this behavior may be related to the surface mobility model used for simulation. The mobility model is limited to a dedicated range of electrical fields via Eq. (6). It can be adapted with the factor G_{surf}, when the field is altered over a large range and higher accuracy is required. This issue is subject to further investigations.

5. Conclusions

The first attempt is made to correlate experimental data from fully functional crystalline high-k MOSFETs with results obtained from process- and device-simulation. The measured device characteristics agrees resonably well with properly adjusted simulation results. The comparison reveals that increased values of the interface state density degrades sub-threshold slope and carrier mobility. However, even though carrier mobility depends slightly on the selected mobility model, interface states alone are insufficient to explain the low mobility values. Additional effects appear to be responsible, like enhanced surface roughness or strain. Nevertheless, damascene metal gate technology has proven to be superior to the conventional poly-Si gate CMOS processing, since the process steps potentially damaging to the high-k gate stack are omitted. However, when device performance is compared to SiO_2 reference MOSFETs, the need of further improvements of the crystalline high-k materials and processing is still obvious.

Acknowledgements

The authors acknowledge the financial support by the German Federal Ministry of Education and Research (BMBF) under the KrisMOS project (01M3142C), AMD Saxony LLC & Co. KG, Infineon Technologies AG and Freescale Halbleiter Deutschland GmbH. We also like to thank M. Czernohorsky, E. Bugiel, A. Fissel, and H. J. Osten of the University of Hannover for the MBE work.

References

[1] H. J. Osten, J. P. Liu, P. Gaworzewski, E. Bugiel, and P. Zaumseil, "High-*k* gate dielectrics with ultra-low leakage current based on praseodymium oxide", in *Tech. Dig. IEDM 2000*, San Francisco, USA, 2000, pp. 653–656.

[2] U. Schwalke, "Gate dielectrics: process integration issues and electrical properties", *J. Telecommun. Inform. Technol.*, no. 1, pp. 7–10, 2005.

[3] U. Schwalke, K. Boye, K. Haberle, R. Heller, G. Hess, G. Müller, T. Ruland, G. Tzschöckel, J. Osten, A. Fissel, and H.-J. Müssig, "Process integration of crystalline Pr_2O_3 high-*k* gate dielectrics", in *Proc. 32nd Eur. Solid State Dev. Res. Conf. ESSDERC*, Firenze, Italy, 2002, p. 407.

[4] H. D. B. Gottlob, M. C. Lemme, T. Mollenhauer, T. Wahlbrink, J. K. Efavi, H. Kurz, Y. Stefanov, K. Haberle, R. Komaragiri, T. Ruland, F. Zaunert, and U. Schwalke, "Introduction of crystalline high-*k* gate dielectrics in a CMOS process", in *Proc. SiO_2, Adv. Dielectr. Rel. Dev.*, Chamonix Mont-Blanc, France, 2004.

[5] U. Schwalke, Y. Stefanov, R. Komaragiri, and T. Ruland, "Electrical characterisation of crystalline praseodymium oxide high-*k* gate dielectric MOSFETs", in *Proc. 33rd Eur. Solid State Dev. Res. Conf. ESSDERC*, Estoril, Portugal, 2003, pp. 247–250.

[6] Y. Stefanov, R. Komaragiri, and U. Schwalke, "Device level and nanoscale electrical characterization of crystalline praseodymium oxide high-*k* gate dielectric MOSFETs", in *SEMATECH Int. Worksh. Electr. Charact. Reliab. High-k Dev.*, Austin, USA, 2004.

[7] Y. Stefanov, G. Hess, G. Tzschöckel, and U. Schwalke, "Global and local charge trapping effects in crystalline praseodymium oxide high-*k* gate dielectric MOSFETs", in *Electrochem. Soc. Int. Semicond. Technol. Conf. ECS-ISTC*, Shanghai, China, 2004.

[8] Y. Stefanov *et al.*, "Nanoscale electrical characterization of crystalline praseodymium oxide high-*k* gate dielectric MOSFETs with conductive atomic force microscopy", in *Proc. Seeing Nanosc. III*, Santa Barbara, USA, 2005.

[9] R. Komaragiri, F. Zaunert, and U. Schwalke, "Gate engineering for high-*k* dielectric and ultra-thin gate oxide CMOS technologies", in *Worksh. Semicond. Adv. Fut. Electron. SAFE 2004*, Veldhoven, The Netherlands, 2004.

[10] R. Endres, Y. Stefanov, and U. Schwalke, "Electrical characterization of crystalline Gd_2O_3 gate dielectric MOSFETs fabricated by damascene metal gate technology", in *Worksh. Dielectr. Microelectron. WoDiM*, Catania, Italy, 2006.

[11] H. D. B. Gottlob, T. Echtermeyer, T. Mollenhauer, M. Schmidt, J. Efavi, T. Wahlbrink, M. C. Lemme, H. Kurz, R. Endres, Y. Stefanov, U. Schwalke, M. Czernohorsky, E. Bugiel, A. Fissel, and H. J. Osten, "Approaches to CMOS integration of epitaxial gadolinium oxide high-*k* dielectrics", in *Eur. Solid State Dev. Res. Conf. ESSDERC*, Montreux, Switzerland, 2006.

[12] M. Cassé, L. Thevenod, B. Guillaumot, L. Tosti, F. Martin, J. Mitard, O. Weber, F. Andrieu, T. Ernst, G. Reimbold, T. Billon, M. Mouis, and F. Boulanger, "Carrier transport in HfO_2/metal gate MOSFETs: physical insight into critical parameters", *IEEE Trans. Electron Dev.*, vol. 53, no. 4, pp. 759–768, 2006.

[13] U. Schwalke, M. Kerber, K. Koller, and H. Jacobs, "EXTIGATE: the ultimate process architecture for submicron CMOS technologies", *IEEE Trans. Electron Dev.*, vol. 44, no. 11, pp. 2070–2077, 1997.

[14] Goodfellow Corporation, Material Index, www.goodfellow.com

[15] P. M. D. Chow and K. L. Wang, "A new ac technique for accurate determination of channel charge and mobility in very thin gate MOSFET's", *IEEE Trans. Electron Dev.*, vol. ED-33, pp. 1299–1304, 1986.

[16] C. L. Huang and G. S. Gildenblat, "Correction factor in the split C-V method for mobility measurements", *Solid-State Electron.*, vol. 36, pp. 611–615, 1993.

[17] C. L. Huang, J. V. Faricelli, and N. D. Arora, "A new technique for measuring MOSFET inversion layer mobility", *IEEE Trans. Electron Dev.*, vol. 40, pp. 1134–1139, 1993.

[18] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field", *Proc. IEEE*, vol. 55, pp. 2192–2193, 1967.

[19] S. Selberherr, "Process and device modeling for VLSI", *Microelectron. Reliab.*, vol. 24, no. 2, pp. 225–257, 1984.

[20] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley, 1981.

[21] H. J. Osten, private communication.

Florian Zaunert received the diploma degree in electrical engineering from Darmstadt University of Technology, Germany, in 2003. Currently he is working towards the Ph.D. degree at the Institute for Semiconductor Technology at the same university. His dissertation theme focuses on the study and characterization of novel gate stacks for MOS applications, especially with technology and device simulation. Before joining the institute, he experienced an industrial placement at Bosch Automotive Electronics in Reutlingen, Germany, division Development Integrated Circuits.
e-mail: zaunert@iht.tu-darmstadt.de
Institute for Semiconductor Technology
Darmstadt University of Technology
Schlossgartenstraße 8
64297 Darmstadt, Germany

Ralf Endres was born in Gerolstein, Germany. He received his diploma degree in electrical engineering from Darmstadt University of Technology in 2005 and is currently doing his Ph.D. study at the Institute for Semiconductor Technology, Darmstadt, Germany. His research interests include new materials in CMOS tech-

nology as alternative high-*k* gate dielectrics and metal gates with focus on their gentle process integration and electrical characterization.

e-mail: endres@iht.tu-darmstadt.de
Institute for Semiconductor Technology
Darmstadt University of Technology
Schlossgartenstraße 8
64297 Darmstadt, Germany

Yordan Stefanov was born in Sofia, Bulgaria. He received his diploma degree in electrical engineering at Sofia University in 2002. From 2002 to 2006 he worked as a Ph.D. student at the Institute for Semiconductor Technology at the Darmstadt University of Technology, Germany, where he is approaching his Ph.D. degree in microelectronics. Main research interests are complementary metal-oxide-semiconductor technology, the integration and characterization of novel gate dielectrics and chemical mechanical planarization.

e-mail: stefanov@iht.tu-darmstadt.de
Institute for Semiconductor Technology
Darmstadt University of Technology
Schlossgartenstraße 8
64297 Darmstadt, Germany

Udo Schwalke was awarded a Research Fellowship from the Alexander-von-Humboldt-Foundation for his Ph.D. thesis in 1984. During 1984–1986, he was appointed Caltech Research Fellow at the California Institute of Technology, USA. In 1987, he joined the Siemens AG, R&D Microelectronics in Munich, Germany. From 1990–1992 he was responsible for the 64 Mb DRAM device design at IBM/Siemens, USA. Subsequently, at Infineon Technologies AG (the former Siemens Semiconductor Group) he worked on novel process architectures and device reliability issues. Since August 2001, he is Professor in electrical engineering and the Managing Director of the Institute for Semiconductor Technology at Darmstadt University of Technology, Germany. Professor Schwalke has served as expert evaluator and reviewer for the European Commission and is a Member of the IEE, IEEE and ECS. He has authored or co-authored more than 80 technical papers and holds several patents. Currently his research interests are focused on nanoscale CMOS technologies and molecular nanoelectronics.

e-mail: schwalke@iht.tu-darmstadt.de
Institute for Semiconductor Technology
Darmstadt University of Technology
Schlossgartenstraße 8
64297 Darmstadt, Germany

Invited paper

Energy concepts involved in MOS characterization

Olof Engström, Tomasz Gutt, and Henryk M. Przewłocki

Abstract— Starting from a quantum statistical reasoning, it is demonstrated that entropy properties of silicon/silicon dioxide interface electron traps may have a strong influence on measured distributions of interface states, depending on measurement method used. For methods, where the Fermi-level is used as a probe to define an energy position, the scale is based on free energy. On the other hand, methods based on thermal activation of electrons give the distribution on an enthalpy scale. It is shown that measured interface state distributions are influenced by the distribution of entropy, and that common features of measured energy distributions may be influenced by entropy variations. These results are used to interpret experimental data on the energy distribution of electron capture cross sections with an exponential increase followed by a more or less constant value as the energy distance of the traps from the conduction band edge increases. Such a relation is shown to be consistent with a situation where the emission and capture processes of electrons obey the Meyer-Neldel rule.

Keywords—C-V technique, capture cross sections, interface states, Meyer-Neldel rule, MOS, thermally stimulated current.

1. Introduction

Techniques commonly used for investigating electrical interface properties of metal-oxide-semiconductor (MOS) structures can be divided into two different classes depending on the energy concepts involved. In the capacitance versus voltage (C-V) [1], charge pumping [2] and conductance techniques [1], the position of the Fermi-level at the oxide-semiconductor interface is used to define the energy scale when determining the position of an interface electron state in relation to the semiconductor band edges. Another class includes methods where the energy position is measured by thermal activation, as for example in a thermally stimulated current experiment [3]. It has been shown that the energy quantities in these two cases are to be interpreted as different thermodynamic potentials; in the first case as free energies, in the second case as enthalpies [4–7]. A third energy scale is governed by quantum mechanical energy level calculations, where only the electronic potentials of the defect systems are taken into account. This corresponds to the electronic eigenvalues of the interface states and gives rise to a third energy scale. In principle, these three different energy quantities are identical only at zero absolute temperature, $T = 0$ K.

When measuring the interface electron state densities, D_{it}, of MOS structures by Fermi-probe technique, often two characteristic features are obtained when D_{it} is plotted as a function of the Fermi-level position at the interface: a U-shaped over-all distribution and two peaks at about 0.3 eV from the conduction and valence bands, respectively [1]. The latter are generally associated with the Pb-center [8]. It has been suggested that the peaks may be influenced by the merging of states on a free energy scale, due to a varying entropy along the interface state distribution [6]. Another common feature is the energy dependence of capture cross sections. In a number of works, a distribution has been found where the capture cross section increases exponentially with increasing energy distance to the conduction band and rolls over to a more or less constant value for energies deeper than about 0.3 eV [9–15].

In the present paper, we study interface electron states by taking into account the influence of total entropy changes when electrons are emitted to and captured from the conduction band. This treatment will be used as a background to explain experimental electron capture data for MOS interfaces. By adding new experimental data to that existing in the literature, we demonstrate that the energy dependence of capture cross sections can be considered as a result of interface states obeying the Meyer-Neldel rule (MNR) [16].

2. Carrier statistics of interface states

2.1. Occupation probabilities in the grand canonical ensemble

The probability, $P(1)$, to capture one electron, for an interface trap capable to hold one electron, can be expressed in the grand canonical ensemble as [17]

$$P(1) = \frac{Z_1 \exp\left(\frac{1\mu}{\mathrm{k}T}\right)}{\sum_{r=0}^{1} Z_r \exp\left(\frac{r\mu}{\mathrm{k}T}\right)}, \tag{1}$$

where μ is the Fermi-level at the interface, k is Boltzmann's constant and T is absolute temperature. Z_r is the partition function of the canonical ensemble and is given by

$$Z_r = \sum_{i=1}^{I} \exp\left(-\frac{E_{i,r}}{\mathrm{k}T}\right) \equiv \exp\left(-\frac{G_r}{\mathrm{k}T}\right). \tag{2}$$

Here $E_{i,r}$ is the eigenenergy of the trap in state i with a maximum of I available states and with r electrons captured. The partition function can also be expressed by introducing free energy G_r as defined in Eq. (2). Using Eq. (2)

in Eq. (1) and introducing the definition of a "free energy level of the trap", G_T, as the free energy difference for $r = 1$ and $r = 0$,

$$G_T \equiv G_1 - G_0\,, \tag{3}$$

one finds

$$P(1) = \frac{1}{1 + \exp\left(\frac{G_T - \mu}{kT}\right)}\,. \tag{4}$$

From this equation, it can be seen that the occupation probability, $P(1)$, of the trap is determined by the relation between the free energy position, G_T, and the Fermi-level, μ. Especially, when these two quantities coincide, the probability of trap occupation is $1/2$. From this we can make the following important statement: *When the Fermi-level is used as a probe to select interface state energy positions, the energy scale of the distribution obtained is based on a free energy*. It has been argued that for deep bulk energy levels this quantity can be identified as the Gibbs free energy [18].

2.2. Thermal emission rates

At thermal equilibrium, the emission of electrons from an interface energy level, G_T, to the conduction band is balanced by a capture process governed by the average thermal velocity of electrons in the condution band, v_{th}, and the capture cross section, σ_n, of electrons according to the following equality:

$$e_n P(1) = v_{th}\sigma_n n\big(1 - P(1)\big)\,, \tag{5}$$

where e_n is the thermal emission rate for electrons from the trap level to the conduction band and n is the concentration

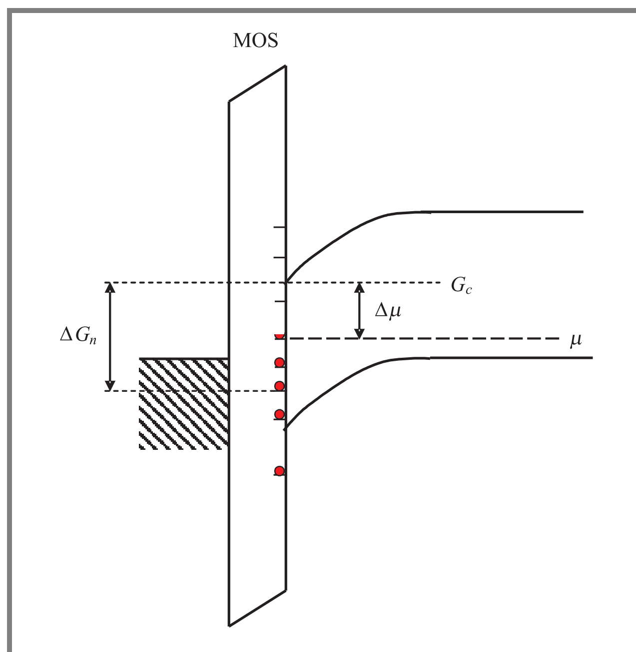


Fig. 1. Band model of the MOS structure on a free energy scale.

of electrons in the conduction band. The latter quantity follows the standard expression

$$n = N_c \exp\left(-\frac{G_c - \mu}{kT}\right), \tag{6}$$

where N_c is the effective density of states in the conduction band and G_c is the free energy of carriers in the conduction band (equal to the "conduction band edge position" at $T = 0$ K), respectively. Introducing the definition $\Delta G_n \equiv G_c - G_T$, as depicted in Fig. 1, we find from Eqs. (5) and (6):

$$e_n = v_{th}\sigma_n N_c \exp\left(-\frac{\Delta G_n}{kT}\right). \tag{7}$$

Taking into consideration that the electron emission from the interface trap system may be connected to a change

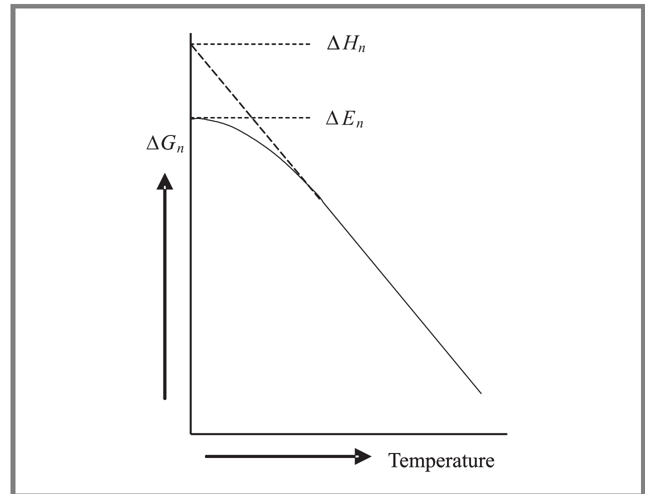


Fig. 2. Temperature dependence of the free energy position, ΔG_n of an interface trap often occurring in semiconductors [19]. For not too low temperatures, $\Delta G_n(T)$ approaches a linear function. As the entropy is $\Delta S_n = -\partial(\Delta G_n)/\partial T$, this quantity is temperature independent for the higher temperatures. At $T = 0$, the free energy and the eigen energy coincide. In an activation measurement, where the slope of an Arrhenius plot is ΔH_n, this quantity is the extrapolated value of ΔG_n in the temperature interval where the measurement is made. For simplicity, in the present treatment, we assume that ΔG_n is linear down to $T = 0$, which makes $\Delta G_n = \Delta H_n = \Delta E_n$ at this temperature.

in entropy, ΔS_n, due to electronic degeneracy and ionic vibration [4, 7, 17, 18], we use the thermodynamic relation:

$$\Delta H_n = \Delta G_n + \Delta S_n T\,, \tag{8}$$

where ΔH_n is an enthalpy. Thus, an alternative expression for Eq. (7) is given by

$$e_n = v_{th}\sigma_n N_c \exp\left(\frac{\Delta S_n}{k}\right)\exp\left(-\frac{\Delta H_n}{kT}\right). \tag{9}$$

In an Arrhenius plot of the emission rate for different temperatures, the slope of the activation curve gives the enthalpy value if the temperature dependencies of the pre-exponential factors in Eq. (9) are known.

For traps in semiconductors, the entropy, ΔS_n, is often temperature dependent [19]. As the thermodynamic relationship between entropy and free energy is $\Delta S_n = -\partial \Delta G_n / \partial T$, this means that the temperature dependence of ΔG_n is nonlinear. In practice it exhibits a shape shown schematically in Fig. 2. In the present treatment, we assume for simplicity, that ΔS_n is temperature independent, meaning that ΔG_n is a linear function of T. For such a case it is readily shown from Eq. (2) that the mean value of the occupied eigenenergy levels, $E_{i,r}$, fulfills a relation to the free energy and the entropy as given by Eq. (8). For a one-electron trap with only one energy level, this means that the energy eigenvalue distance from the conduction band, ΔE_n of this level fulfills the relation:

$$\Delta E_n = \Delta G_n + \Delta S_n T \,. \qquad (10)$$

Hence, the enthalpy and the energy eigenvalue are identical for this specific case.

2.3. Influence of entropies on the energy distribution of interface states

When measuring D_{it} by Fermi-probe technique, the energy resolution is determined by a weight function [4], $w = \partial P(1) / \partial(\Delta G_n)$, which is obtained from Eq. (4) as

$$w(\Delta G_n) = \frac{1}{kT} \frac{\exp\left(\frac{\Delta\mu - \Delta G_n}{kT}\right)}{\left\{1 + \exp\left(\frac{\Delta\mu - \Delta G_n}{kT}\right)\right\}^2}, \qquad (11)$$

where $\Delta\mu = G_c - \mu$.

The interface state density distribution on a free energy scale is then given by

$$D_{it}(\Delta G_n) = \int_0^{\Delta E_g} D_{it}(\Delta E_n)\big[w\big(\Delta G_n(\Delta E_n)\big)\big] d(\Delta E_n), \qquad (12)$$

where the function $\Delta G_n(\Delta E_n)$ is given by Eq. (10) and ΔE_g is the semiconductor energy band gap.

In order to study the influence of a varying entropy, we assume that D_{it} calculated on an energy eigenvalue scale is constant. We will see that changes in the entropy distribution make clear changes in D_{it} when based on a free energy scale. As an example, taking the entropy distribution on an eigenenergy scale as shown in Fig. 3a, the corresponding interface state distribution is demonstrated in Fig. 3b on a free energy scale. Here we notice how $D_{it}(\Delta G_n)$ is merged towards the conduction band by the $\Delta S_n(\Delta E_n)$ function chosen in Fig. 3a. This shape of $\Delta S_n(\Delta E_n)$ gives the standard features often observed for D_{it} as measured by C-V technique: the increased interface state concentration close to the band edge and a peak at about 0.3 eV marking the Pb-center. The position of the 0.3 eV peak depends on the value of the constant part of the entropy curve in Fig. 3a while the step at about $\Delta E_n = 0.3$ eV determines the position of the flank at lower energies.

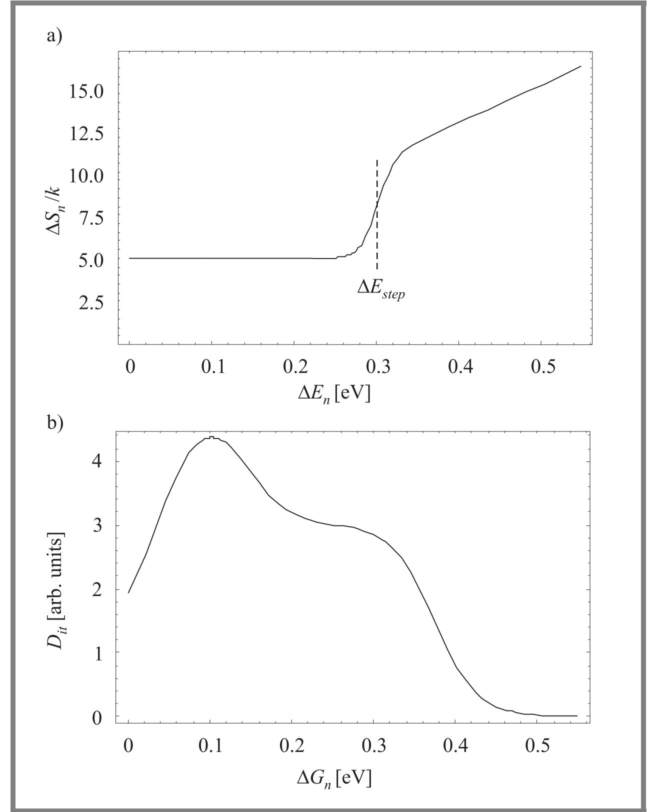


Fig. 3. Entropy in units of k as a function of eigenenergy (a) used for calculating the interface state density, D_{it} as a function of free energy (b) for a D_{it} distribution which is constant on an eigen energy scale.

2.4. Interface traps following the Meyer-Neldel rule

Considering a set $\{j\}$, constituting a "family" of electron traps with properties given by $(e_j, \sigma_j, \Delta S_j, \Delta H_j)$, the thermal emission rate for each member, j, can be expressed as

$$e_j = v_{th}\sigma_j N_c \exp\left(\frac{\Delta S_j}{k}\right)\exp\left(-\frac{\Delta H_j}{kT}\right) = v_{th}\sigma_j N_c \exp\left(-\frac{\Delta G_j}{kT}\right), \quad j = 1, 2, 3, \ldots \qquad (13)$$

Assuming that there exists an "iso-kinetic" temperature T_c such that the emission rate, e_c, is equal for all $j = 1, 2, 3, \ldots$, we have for $e_c \equiv e_j(T_c)$:

$$e_c = v_{th}\sigma_j N_c \exp\left(\frac{\Delta S_j}{k}\right)\exp\left(-\frac{\Delta H_j}{kT_c}\right) = v_{th}\sigma_j N_c \exp\left(-\frac{\Delta G_j}{kT_c}\right). \qquad (14)$$

For an Arrhenius plot, this means that all activation curves would intercept at the iso-kinetic temperature, T_c, with a common emission rate, e_c. Such a feature indicates that the family, $\{j\}$, obeys the Meyer-Neldel rule [16]. From Eq. (14), we find

$$\ln \sigma_j = \ln \frac{e_c}{v_{th} N_c} + \frac{\Delta G_j}{kT_c}. \quad (15)$$

Hence, for a set of interface traps following the Meyer-Neldel rule, one would expect the logarithm of the capture cross sections to be a linear function of the free energy, ΔG_j. Such experimental data are commonly found for SiO_2/Si interfaces [9–15] and will be further discussed below.

From Eq. (15) it is also seen that increasing the iso-kinetic temperature, T_c, to high values makes the capture cross section less dependent on the free energy, ΔG_j, which may explain the constant range for capture cross sections often found for the deeper part of the D_{it} distribution of MOS interfaces, at energies larger than about 0.3 eV [9–15]. A second possibility would be that the enthalpy for releasing an electron from the trap is the product between T_c and ΔS_n, i.e., corresponds to the total heat for releasing the electron at this temperature:

$$\Delta H_j = \Delta S_j T_c. \quad (16)$$

Such a relation has been discussed for deep bulk levels by Van Vechten and Thurmond [18]. As a motivation for MNR to be valid for interface states, it was also suggested by Johnston *et al*. [20] and was further used in [21] to explain the existence of MNR among a large number of bulk traps in GaAs. It means that the entropy is an increasing function of enthalpy or, in the theoretical treatment above, with the eigenenergy difference, ΔE_n. For a case where Eq. (16) is valid, we find from Eq. (13) that $\Delta G_j = 0$ which means that all traps of the family have the same capture cross section $\sigma_j \equiv \sigma_c$ for all $j = 1, 2, 3, \ldots$ However, for the present case we find from the slopes of the curves in Figs. 4 and 5 that $T_c \approx 300$ K, which is the temperature where the measurements were made. As ΔG_j is an independent parameter controlled by the applied gate voltage in this experiment, the former interpretation for the constant capture cross sections at higher energies is the most probable.

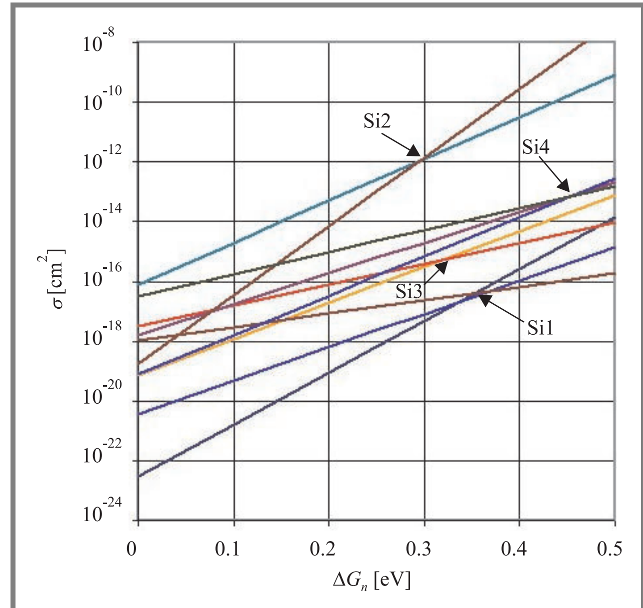


Fig. 5. Capture cross section versus free energy taken from the literature, showing the interception of curves. Each curve represents one interface state "family". Data intercepting at points marked in the diagram represent a "dynasty". The slopes of the curves give the iso-kinetic temperatures and the intercepts with the ordinate axis give the e_c values (Eq. (14)).

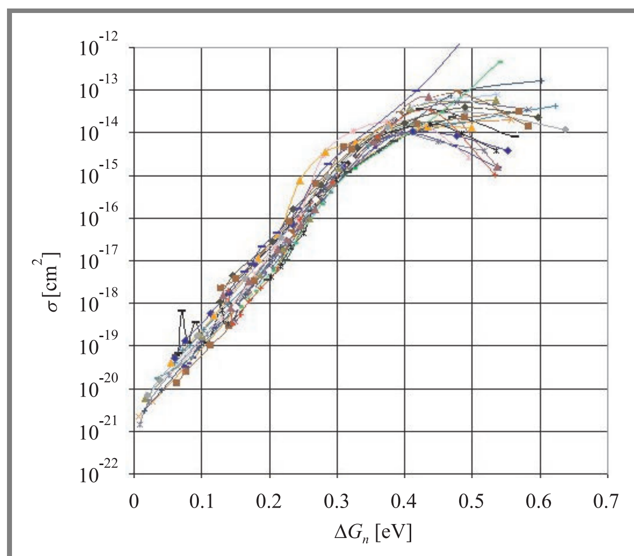


Fig. 4. Electron capture cross sections of Si/SiO_2 interface states as a function of free energy position in relation to the conduction band.

3. Experimental

In order to confirm the results earlier published in the literature [9–15], we repeated measurements on thermally oxidized MOS structures by using the conductance technique [1]. The samples were prepared by oxidizing 4-inch, n-type, phosphorus doped silicon wafers at 1000°C in an N_2/O_2 flow with 31/45 min ramp-up/ramp-down rates to obtain SiO_2 thicknesses of $t_{ox} = 60$ and 160 nm. The wafers were annealed afterwards in nitrogen for $t(N_2)$ = 0, 10, 120 and 1440 min at the temperature $T = 1050$°C with 40/55 min ramp-up/ramp-down rates. Then, 416 nm of aluminium was deposited, followed by the standard post-metallization annealing at 450°C for 20 min in N_2/H_2.

Figure 4 shows the capture cross section for electrons as a function of free energy distance from the conduction band for a large number of samples taken from this series of wafers. For the lower energy part, the capture cross sections increase exponentially over more than six orders of

magnitude. At $\Delta G_n = 0.4$ eV, the increasing feature rolls off and becomes either close to constant or decreases with increasing energy. Often in the literature, a more constant behavior is observed in this energy range [9–15].

Figure 5 depicts corresponding data for the low energy part, redrawn from literature [9–15]. An interesting feature of these results is that the different families, representing different MOS systems, in a couple of cases seem to cluster into common crossing points at about 0.3–0.4 eV. From the slopes of the curves, the iso-kinetic temperature T_c can be calculated while the intercepts at the ordinate axis gives the emission rate e_c, in accordance with Eq. (14). Each family has its own values of these quantities but, interesting enough, in some cases they seem to cluster into "dynasties", i.e., at one common ΔG value some families have one common σ value. Moreover, the different families originate from the results of different authors! The data indicate that for each dynasty there exists at least one common free energy value, ΔG_j with a common capture cross section, governed by the crossing points in Fig. 5. This means that plotting the e_c values versus the inverse T_c values in an Arrhenius plot would give one activation curve for each dynasty. As seen in Fig. 6, this is indeed the case. Because all ΔG_j values at the crossing points in Fig. 5 were taken at about the same room temperature, the slope of the activation plots in Fig. 6 is expected to be equal to the free energies as deduced from Eq. (13).

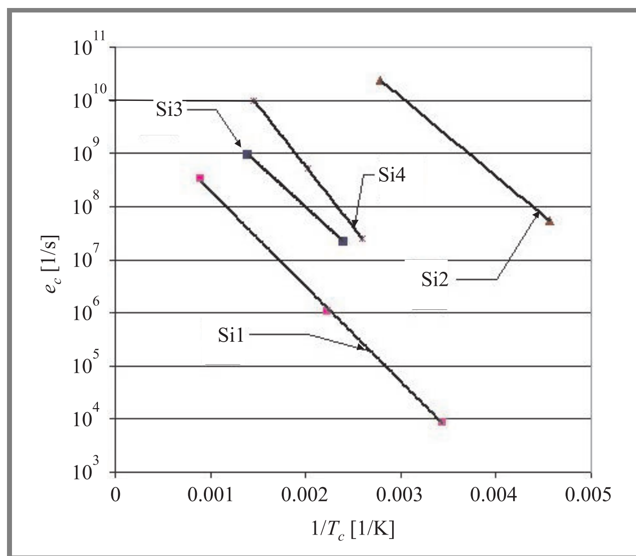


Fig. 6. Dynasties represented in an Arrhenius plot with data taken from Fig. 5.

4. Discussion

The theoretical $D_{it}(\Delta G_n)$ curves in Figs. 3a and 3b were based on an assumption that the value of $D_{it}(\Delta E_n)$ is constant. There is no experimental or theoretical support for such an approach. The idea is to demonstrate the strong influence of entropy variation among the interface states on measured $D_{it}(\Delta G_n)$ distributions. Such consideration is seldom taken into account when comparing theoretically calculated $D_{it}(\Delta E_n)$ data with measured $D_{it}(\Delta G_n)$ results. Theoretical calculations of, for example, the U-shaped interface state distribution by Sakurai and Sugano [22] or of the Pb-center by Edwards [23] were made for eigenenergy scales. Even if these results demonstrate the occurrence of specific distribution characteristics, it is the aim of the present work to point out the influence of electron state degeneracies and ionic vibrational changes [7], together manifesting the total entropy of interface states.

The entropy distribution as shown in Fig. 3a, with a constant value for lower energies and a linear behavior for energies above about 0.3 eV, gives rise to a $D_{it}(\Delta G_n)$ distribution which qualitatively agrees with the features found from Fermi-probe measurements like the C-V technique. Furthermore, it gives a motivation for the shape of the free energy distribution of capture cross sections normally found at SiO_2/Si interfaces by assuming that this kind of states belong to "families" obeying the Meyer-Neldel rule. A fascinating observation is that these families can be portioned into clusters of "dynasties", including centers with equal capture cross sections at equal free energy positions. The condition for making up a dynasty is that one capture cross section is identical for the members of each family. An interesting observation in Figs. 5 and 6 is that this occurs at ΔG_n values of about 0.3–0.4 eV, where the Pb-center is normally found. This would mean that centers giving rise to what is normally labeled as the "Pb-center" may have widely different capture cross sections as given by the interception points between the curves in Fig. 5.

References

[1] E. H. Nicollian and J. R. Brews, *MOS (Metal-Oxide-Semiconductor) Physics and Technology*. New York: Wiley-Interscience, 1982.

[2] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. DeKeersmaecker, "A reliable approach to charge-pumping measurements in MOS transistors", *IEEE Trans. Electron Dev.*, vol. ED-31, pp. 42–53, 1984.

[3] P. Blood and J. W. Onton, *The Electrical Characterization of Semiconductors: Majority Carriers and Electron States*. London: Academic Press, 1992.

[4] O. Engström and A. Alm, "Energy concepts of insulator-semiconductor interface traps", *J. Appl. Phys.*, vol. 54, pp. 5240–5244, 1983.

[5] O. Engström and M. S. Shivaraman, "Statistics for the interpretation of deep level transient spectroscopy on insulator-semiconductor interfaces", *J. Appl. Phys.*, vol. 58, pp. 3929–3930, 1985.

[6] O. Engström, "Influence of entropy properties on measured trap energy distributions at insulator-semiconductor interface", *Appl. Phys. Lett.*, vol. 55, pp. 47–49, 1989.

[7] O. Engström and H. G. Grimmeiss, "Vibronic states of silicon-silicon dioxide interface traps", *Semicond. Sci. Technol.*, no. 4, pp. 1106–1115, 1989.

[8] E. H. Poindexter, "MOS interface states: overview and physicochemical perspective", *Semicond. Sci. Technol.*, no. 4, pp. 961–969, 1989.

[9] W. Fahrner and A. Goetzberger, "Energy dependence of electrical properties of interface states in Si SiO_2 interfaces", *Appl. Phys. Lett.*, vol. 17, pp. 16–18, 1970.

[10] H. Deuling, E. Klausmann, and A. Goetzberger, "Interface states in Si-SiO_2 interfaces", *Solid State Electron.*, vol. 15, pp. 559–571, 1972.

[11] M. Morita, K. Tsubouchi, and N. Mikoshiba, "Measurement of interface-state parameters near the band edge at the Si/SiO_2 interface by the conductance method", *Appl. Phys. Lett.*, vol. 33, pp. 745–747, 1978.

[12] W. D. Eades and R. M. Swanson, "Determination of the capture cross section and degeneracy factor of Si-SiO_2 interface states", *Appl. Phys. Lett.*, vol. 44, pp. 988–990, 1984.

[13] N. Hanei, L. Vishnubhotla, and T. P. Ma, "Possible observation of P_{b0} and P_{b1} centers at irradiated (100)Si/SiO_2 interface from electrical measurements", *Appl. Phys. Lett*., vol. 59, pp. 3416–3418, 1991.

[14] M. Uren, "Interface state capture cross section measurements on vacuum annealed and radiation damaged Si:SiO_2 surfaces", *J. Electrochem. Soc.*, vol. 145, pp. 683–689, 1988.

[15] J. Albohn, W. Füssel, N. D. Sinh, K. Kliefoth, and W. Fuhs, "Capture cross sections of defect states at the Si/SiO_2 interface", *J. Appl. Phys.*, vol. 88, pp. 842–849, 2000.

[16] A. Yelon and B. Movaghar, "Microscopic explanation of the compensation Meyer-Neldel rule", *Phys. Rev. Lett.*, vol. 65, pp. 618–620, 1990.

[17] *Handbook of Semiconductors*, P. T. Landsberg, Ed. New York: Elsevier, 1992, vol. 1, chapter 6.

[18] J. A. Van Vechten and C. D. Thurmond, "Entropy of ionization and temperature variation of ionization levels of defects in semiconductors", *Phys. Rev. B*, vol. 14, pp. 3539–3550, 1976.

[19] C. D. Thurmond, "The standard thermodynamic functions for the formation of electrons and holes in Ge, Si, GaAs and GaP", *J. Electrochem. Soc.*, vol. 122, pp. 1133–1141, 1975.

[20] S. W. Johnston, R. S. Crandall, and A. Yelon, "Evidence of the Meyer-Neldel rule in InGaAsN alloys and the problem of determining trap capture cross sections", *Appl. Phys. Lett.*, vol. 83, pp. 908–909, 2003.

[21] O. Engström and M. Kaniewska, "Discovery of classes among deep level centers in gallium arsenide" (submitted to *J. Mat. Sci. B*).

[22] T. Sakurai and T. Sugano, "Theory of continuously distributed trap states at Si-SiO_2 interfaces", *J. Appl. Phys.*, vol. 52, pp. 2889–2896, 1981.

[23] A. H. Edwards, "Theory of the Pb center at the $\langle 111 \rangle$ Si/SiO_2 interface", *Phys. Rev. B*, vol. 36, pp. 9638–9648, 1987.

Olof Engström received the Ph.D. degree in solid state physics from the University of Lund in 1975 and was later employed by ASEA AB for research on high power thyristors, by AB Rifa for development of MOS technology and by the Swedish Defence Research Institute for sensor research. In 1984, he came to Chalmers University of Technology as a Professor in solid state electronics. Between 1996 and 1999, he served as Dean of Chalmers School of Electrical and Computer Engineering and 1999–2002 he was the Director of the Microtechnology Center at Chalmers (MC2). From 2003 he is back in research as a Professor at the Department of Microtechnology and Nanoscience of MC2. His present research interest is in semiconductor quantum structures and interfaces. In 1991, he founded Samba Sensors AB, a company for production of fiberoptical pressure sensors.

e-mail: olof.engstrom@mc2.chalmers.se
Department of Microtechnology and Nanoscience
Chalmers University of Technology
SE-412 96 Göteborg, Sweden

Tomasz Gutt obtained the M.Sc. and Ph.D. degrees in 1976, 1984, respectively, specializing in technological process diagnostics and measurements of MOS structures. In the period of 1980–1992 at the Institute of Electron Technology, Warsaw, Poland, he worked on the development of a computer aided IC manufacturing system. From 1993 until 2003 he was employed by ICL Ltd. and Fujitsu. At present, with the Institute of Electron Technology, he carries out research on semiconductor-insulator interface quality characterization methods.

e-mail: tgutt@ite.waw.pl
Department of MOS System Studies
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland

Henryk M. Przewłocki obtained the M.Sc., Ph.D., and D.Sc. degrees in 1959, 1969, and 2001, respectively, specializing in physics, technology, and measurements of MOS structures. In the period of 1965–1983, he also taught various courses to graduate and undergraduate students of the Technical University of Warsaw, Poland. At present he is head of a research group at the Institute of Electron Technology in Warsaw, specializing in photoelectric properties and measurement methods of MOS structures. He has served as consultant and committeeman to numerous scientific, industrial and educational organizations in Poland and worldwide and has received a number of prestigious awards, including the highest Polish award, the National Award for Outstanding Achievements in Technical Sciences.

e-mail: hmp@ite.waw.pl
Department of MOS System Studies
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland

Invited paper

Modeling of negative bias temperature instability

Tibor Grasser and Siegfried Selberherr

Abstract— Negative bias temperature instability is regarded as one of the most important reliability concerns of highly scaled PMOS transistors. As a consequence of the continuous downscaling of semiconductor devices this issue has become even more important over the last couple of years due to the high electric fields in the oxide and the routine incorporation of nitrogen. During negative bias temperature stress a shift in important parameters of PMOS transistors, such as the threshold voltage, subthreshold slope, and mobility is observed. Modeling efforts date back to the reaction-diffusion model proposed by Jeppson and Svensson thirty years ago which has been continuously refined since then. Although the reaction-diffusion model is able to explain many experimentally observed characteristics, some microscopic details are still not well understood. Recently, various alternative explanations have been put forward, some of them extending, some of them contradicting the standard reaction-diffusion model. We review these explanations with a special focus on modeling issues.

Keywords— reliability, negative bias temperature instability, modeling, simulation, hydrogen, silicon dioxide, defects, interface states, semiconductor device equations.

1. Introduction

After its discovery forty years ago [1, 2] negative bias temperature instability (NBTI) has again moved to the center of scientific attention as a significant reliability concern for highly scaled PMOSFETs [3–7]. This is largely due to the increased electric fields inside the gate-oxide, the presence of nitrogen, and the increased operating temperatures. During bias temperature stress which is normally introduced via a large negative voltage at the gate with drain and source remaining grounded, a shift in device parameters is observed, for instance in the threshold voltage, the subthreshold slope, and the mobility [3, 6]. In particular, the shift of the threshold voltage is often described by a simple power-law

$$\Delta V_{th}(t) = A(T, E_{ox})\, t^n, \quad (1)$$

with A being a coefficient which depends on temperature and the electric field. While in earlier investigations [8, 9] the exponent n in Eq. (1) was given to be in the range $0.2-0.3$, newer investigations [10–12] show that n can be as small as 0.12. In particular it was found that the experimentally determined exponent is very sensitive to the measurement setup. Although this exponent is believed to be the fundamental signature of NBTI [4, 5], the values reported in literature still show a significant scatter. However, the exponent has to be determined as accurately as possible to allow long term extrapolation of device life-times, which are commonly more than ten years, depending on the application, based on relatively short measurements obtained within a couple of days or weeks [13].

Many explanations of NBTI have been given over the years, practically all relying on the depassivation of dangling bonds at the Si/SiO_2 interface during stress. These dangling bonds, which are commonly known as P_b centers [14–16], are present in a considerable number at every Si/SiO_2 interface. During device fabrication they have to be passivated through some sort of hydrogen anneal [3], thereby eliminating the electrically active trap levels. Although the resulting P_bH bonds are very stable, at elevated temperatures and higher electric fields they can be broken, thus reactivating the electrically active trap levels. The charge stored in the P_b centers depends on the position of the Fermi-level and thus on the bias conditions. In addition, fixed positive interface charges might be created, the origin of which is attributed to H^+ or trapped holes.

Of particular importance in that context is the relaxation of the induced damage which is observed as soon as the stress is removed. This recovery can be quite large but the microscopic origin is not completely understood. It was found in 2003 that this effect is extremely important in the understanding of NBTI, because during measurements unintentional recovery had distorted practically all previously available measurement data [10, 12, 17].

Although a lot of progress has been made in the understanding of NBTI, a universally accepted theory is still missing. Many publications focus on refining the classic reaction-diffusion model originally proposed by Jeppson and Svensson [4, 5, 8, 9, 18, 19]. Extended versions of the reaction-diffusion model have been successfully calibrated to a wide range of measurement data reproducing a considerable number of phenomena like temperature dependent slopes via measurement artifacts, AC/DC differences, and saturation effects. However, especially the behavior in the relaxation phase is only qualitatively reproduced. This is demonstrated in Fig. 1 where the NBTI degradation during subsequent stress/relaxation cycles is shown for two different reaction-diffusion models, where good accuracy is only obtained during the first stress phase. During relaxation some apparent saturation is often observed which is not well reproduced. Also, if stress is applied again, the accuracy of the results predicted by the reaction-diffusion model decreases (also see for instance fits to measurements in [4, 20]).

Recently a variety of other explanations for bias temperature instability have been put forward, using for instance

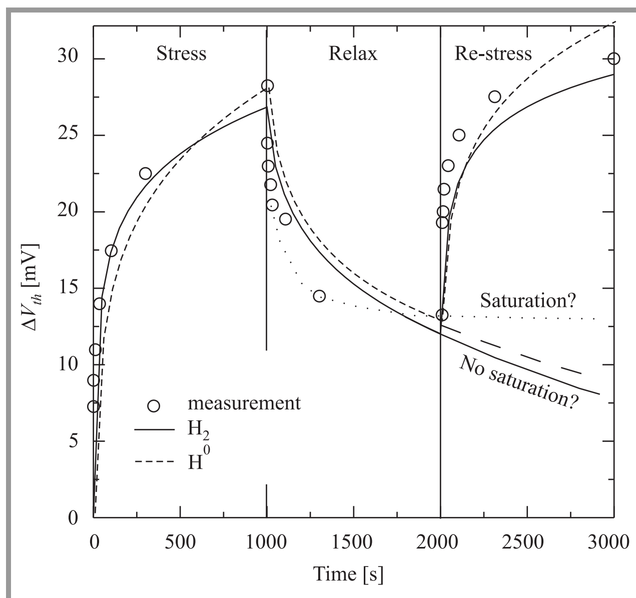


Fig. 1. NBTI degradation during subsequent stress/relaxation cycles. Although good accuracy can be obtained during the first stress phase, the fit is only qualitative and considerably poorer during the second stress phase. Shown are the results of the standard reaction-diffusion model with H^0 and H_2 kinetics together with measurement data from [4].

dispersive transport of the hydrogen species released from the dangling bond [21–23], creation of hole traps [6], a broad distribution of dissociation rates at the interface [6, 24, 25] and interactions with hydrogen from the inversion layer [26]. Some of these extended/alternative models rely on a different microscopic picture, sometimes augmenting – but not always compatible with – the standard reaction-diffusion model.

For modeling NBTI it is often tried to identify a *single dominant* mechanism which determines the asymptotic behavior of NBTI. With various simplifying assumptions a closed form expression for the V_{th} shift over time is sought. As of yet, however, no universally accepted dominant mechanism could be isolated. One can conclude that several mechanisms are at work, each dominant over the other ones in certain devices (nitrided oxides [27], high-k [28], ultrathin oxides [5] compared to power devices with thicker oxides [29]) under certain processing or stressing conditions.

2. Experimental issues

Experimental determination of NBTI induced V_{th} shifts suffers from some fundamental difficulties. To determine the V_{th} shift the stressing voltage on the gate has to be removed, and in addition to $I_D - V_G$ sweeps, capacitance-voltage and charge pumping measurements are often conducted to separate the potential contribution of interface and oxide charges. Unfortunately, with the stressing voltage removed from the gate, the inverse reaction of the depassivation process is favored, resulting in an extremely fast (< 1 ms or even 1 μs) relaxation [6, 10–12]. This relaxation seems to depend on the processing, stressing, and relaxation conditions. While most groups report only partial recovery [6], 100% recovery has also been reported [11] in addition to a contradicting dependence of the recovery rate on the applied gate bias [6, 30–32].

To avoid any interference of this recovery process with the measured data, it has been suggested to measure the drift *without* interruption of the stress condition [6, 17]. One variant just monitors the change in the drain current in the linear regime which is then traced back to a threshold voltage shift via the initial $I_D - V_G$ characteristic [19]. Other variants have been proposed where small variations to the bias conditions can be added allowing the determination of $\Delta I_{d,lin}$ and Δg_m. A drawback of these on-the-fly measurements is that they measure the change in the drain current in the linear regime, where the occupancy of the traps might be different from the one observed during real operating conditions.

Although unintentional measurement delay is detrimental for life-time extrapolation, valuable information about the relaxation physics can be obtained by studying the influence of the measurement delay [10, 19, 30, 33] on the result. Since most of the relaxation occurs within the first milliseconds, extremely fast measurement techniques have been developed [33] which allow to study delays as short as 1 μs.

3. Physical mechanisms

Although the reaction-diffusion model [5, 8, 18] is often successful in describing measurements, knowledge of the underlying microscopic physics is still vague [21–23]. In the following, the most important processes likely to occur during negative bias temperature stress are summarized. They comprise the depassivation and annealing of interface states and the behavior of the released hydrogen inside the surrounding materials. In addition to, or even instead of the chemical reactions underlying the reaction-diffusion model, various other reactions may occur. These reactions are shown schematically in Fig. 2 and explained in more detail in the following.

3.1. Hydrogen in semiconductor devices

Most degradation mechanisms reported in the context of NBTI are closely linked to the existence of hydrogen in SiO_2, Si, and p-Si. Hydrogen in these materials is amphoteric and occurs for instance as H^0, H^+, H^-, and H_2. Due to its negative-U character H^0 is unstable at room temperature [34] and depending on the position of the Fermi-level turns into H^+ or H^-, or dimerizes within a fraction of a second [15]. However, atomic hydrogen occurs as a transient quantity during various reactions. In particular, release of atomic hydrogen is assumed in the reaction-diffusion model which then quickly dimerizes into H_2. However, both H^0 and H_2 are extremely fast diffusers and atomic hydrogen is released from spatially

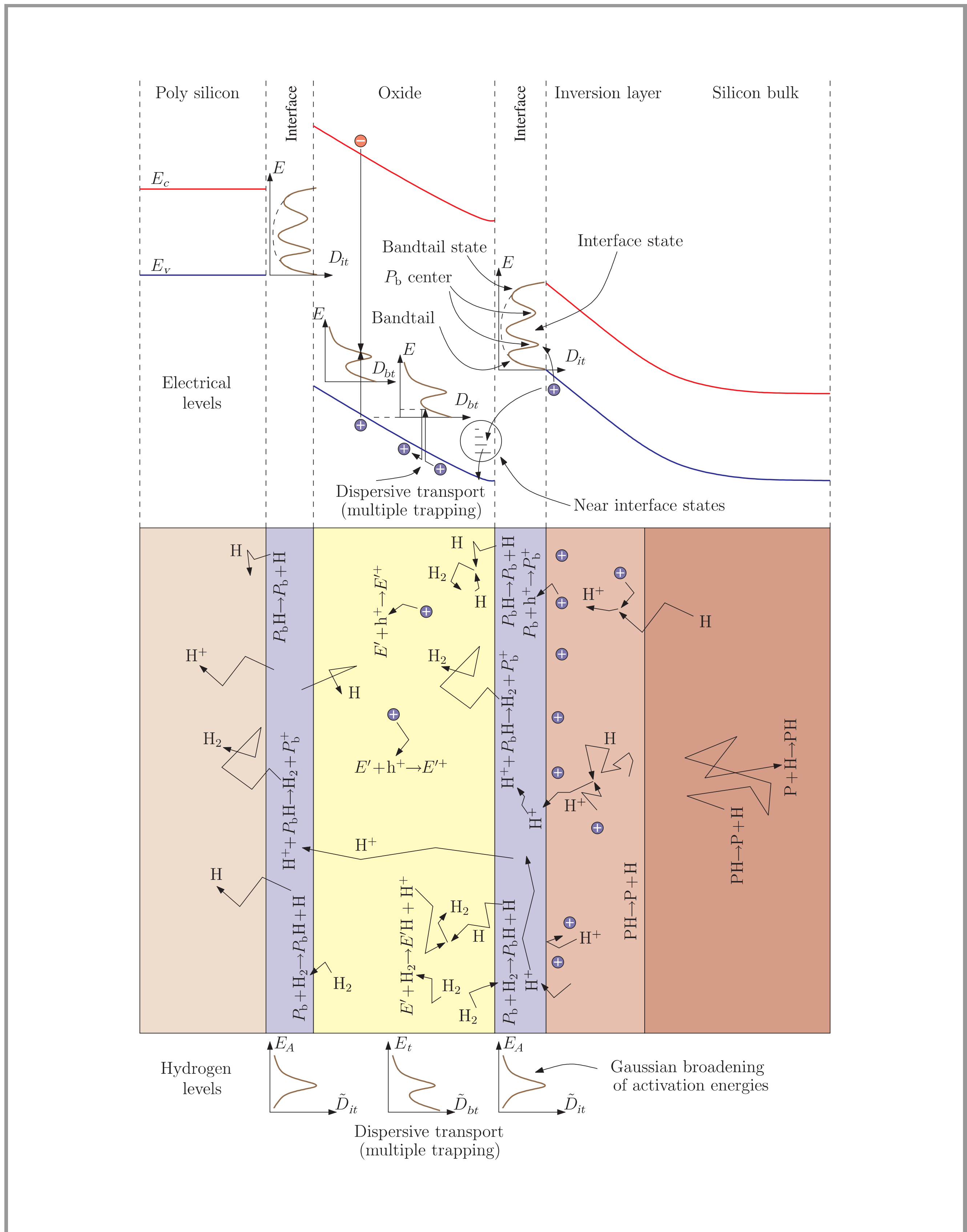


Fig. 2. Various processes reported in the context of NBTI.

separated dangling bonds which might reduce the dimerization rate. Therefore, it is not straightforward to decide whether dimerization occurs first, whether atomic hydrogen leaves the oxide before being able to dimerize, or whether atomic hydrogen turns into H^+, its energetically most favorable state [34].

Another interesting issue stems from the fact that in semiconductor devices hydrogen is present in large amounts. Hydrogen is either unintentionally introduced during the various processing steps or intentionally during the required hydrogen anneals to passivate defects at the Si/SiO_2 interface and dangling bonds in the p-Si-gate. For instance, [35] reported a background H concentration of $10^{19}\,cm^{-3}$ in p-Si. With a "free" hydrogen concentration as large as that, the reverse reaction in the reaction-diffusion model would always be very large, and consequently a very small time exponent would be expected, in contradiction to measurement results. Thus, one needs to assume that most hydrogen is trapped, possibly on shallow bond-center cites or in deep traps formed by dangling bonds, for instance at grain boundaries in p-Si. Trapped hydrogen, however, is not part of the standard reaction-diffusion picture.

Based on first-principle studies [26] it was proposed that the release of the proton is energetically preferable. However, dimerization of H^+ is unlikely due to electrostatic repulsion. In addition, depending on the process conditions, H^+ can be either extremely stable or highly reactive [36]. A proper understanding of the various hydrogen species in SiO_2, Si, p-Si, is thus essential [34] to justify the microscopic picture underlying the reaction-diffusion or alternative models. Matters become further complicated due to the various interactions of hydrogen species with dopands [37] and some additional effects occurring for instance in nitrided oxides [38, 39].

3.2. Dispersive transport

Although the reaction-diffusion model relies on conventional diffusive transport, dispersive transport equations are often used to describe the motion of the hydrogen species in dielectrics and amorphous materials [40–44]. In particular, the type of transport seems to depend on the hydrogen concentration, being diffusive for hydrogen concentrations larger than the trap-density and dispersive otherwise. In the dispersive case the traveling particle packet slows down [45] due to the trapping in states with a broad distribution of release times. As a consequence, the shape of the particle packet becomes non-Gaussian. As a measure of dispersivity one may look at the ratio of the mean and the standard deviation of the particle packet [46]. While for the Gaussian packet this ratio increases with time, it stays roughly constant in the dispersive case, indicating an anomalous spreading of the particle packet. A typical impulse response of the average flux of a dispersive system in comparison to a diffusive system is shown in Fig. 3. Characteristic for dispersive transport is the rapid decline in the beginning (where particles start to fill the traps), followed by a broad shoulder which eventually develops a long tail (note the logarithmic time scale). The response of a diffusive system, on the other hand, vanishes after a transit time of approximately L^2/D, where L is the sample thickness and D the diffusion coefficient.

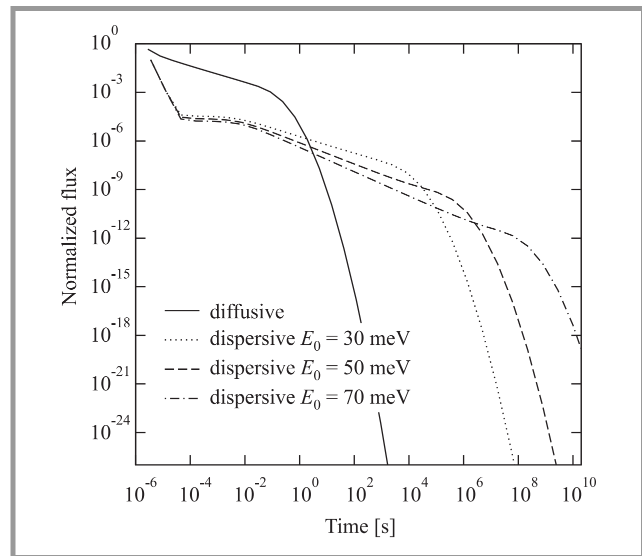


Fig. 3. Impulse response of a diffusion system in comparison to a dispersive system for various characteristic trap energies. Typical for dispersive transport is the rapid decline in the beginning, the shoulder, followed by a long tail.

Dispersive transport models were first applied to describe the movement of holes in amorphous materials [41] and H^+ after irradiation damage [44]. While the first studies were based on the continuous time random walk (CTRW) theory developed by Scher and Montroll [40, 41, 44], multiple trapping (MT) models were proposed soon afterwards [42, 43, 45]. Both models exhibit similar features [47–49] and simplified versions were used to describe NBTI [21–23]. The basic quantity in the CTRW approach is the hopping time distribution which gives the hopping probability from one state to the next. As such, this approach is well suited for Monte-Carlo techniques but analytic solutions for the CTRW equations cannot be given for the general case. Approximate solutions have been commonly sought using the inverse Laplace transformation of the system's Green's function through suitable trial functions, either analytically or numerically [44, 46, 50, 51].

The MT model, on the other hand, uses partial differential equations compatible to the equations conventionally used in process and device simulation and are therefore – in our opinion – more suitable for the inclusion into a numerical simulator. In the MT model the species $X(\mathbf{x},t)$ consists of free (conducting) particles $X_c(\mathbf{x},t)$ and particles residing on various trap levels E_t. The energy density of those trapped particles is given by $\rho(\mathbf{x},E_t,t)$ and the total concentration is calculated as

$$X(\mathbf{x},t) = X_c(\mathbf{x},t) + \int \rho(\mathbf{x},E_t,t)\,\mathrm{d}E_t. \qquad (2)$$

The continuity equation for the total concentration of the species X reads

$$\frac{\partial X(\mathbf{x},t)}{\partial t} = -\nabla \cdot \mathbf{F}_{X_c}(\mathbf{x},t)\,, \tag{3}$$

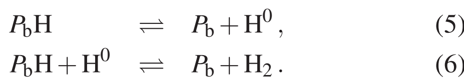
where the flux is only determined by particles in the conduction states. At each trap level a balance equation accounts for the newly trapped particles versus the released ones. The release rate is proportional to the trapped charge on that level, assuming appropriate space in the extended states:

$$\frac{\partial \rho(\mathbf{x},E_t,t)}{\partial t} = c(E_t)X_c(\mathbf{x},t)\big(g(E_t)-\rho(\mathbf{x},E_t,t)\big) - r(E_t)\rho(\mathbf{x},E_t,t)\,. \tag{4}$$

Here, $c(E_t)$ and $r(E_t)$ are the energy-dependent capture and release rates, respectively, and $g(E_t)$ is the trap density of states, where commonly an exponential distribution is assumed. In that context some caution is in order: although hydrogen motion can be phenomenologically described by equations similar to electron transport in semiconductors there are some fundamental differences [52]. First, the configuration of the host material can permanently change as a consequence of hydrogen motion. Second, transport does probably not occur near a mobility edge but rather through hopping from individual shallow states. In addition, hydrogen clusters, known as platelets [37], may form making the application of the trap occupancy concept more involved. Nevertheless, relatively simple models based on two trap levels (shallow and deep) have been successfully used to describe hydrogen motion in a wide range of materials.

3.3. Interface states

The central mechanism in NBT degradation is the dissociation of P_bH bonds located at the Si/SiO$_2$ interface, most possibly through reactions like



$$P_b\mathrm{H} \rightleftharpoons P_b + \mathrm{H}^0\,, \tag{5}$$
$$P_b\mathrm{H} + \mathrm{H}^0 \rightleftharpoons P_b + \mathrm{H}_2\,. \tag{6}$$

The forward reaction in Eq. (5) is commonly assumed to be the dominant dissociation mechanism. Although the reverse reaction, which is responsible for the passivation through atomic hydrogen, is highly effective, its total impact is normally insignificant [53] due to the low concentration of H^0. However, for the special case of radiation environments, where large quantities of atomic hydrogen are generated in the oxide, and for hot carrier effects it becomes important [54, 55]. In particular, combined NBTI/hot carrier/irradiation stress would be a good probe for the validity of a general model.
Equations (5) and (6) also explain the dual behavior of hydrogen as being able to passivate and to depassivate dangling bonds. The reverse reaction through H_2 *without* preliminary cracking [53] is sometimes assumed to be the dominant reaction in the case of NBT stress [56]. Activation energies for the first-order reaction given through Eq. (6) were traditionally estimated to be around 1.6 eV. Recent work has shown that although the first-order kinetics can be confirmed, a Gaussian distribution of activation energies around 1.5 eV with a standard deviation of 0.15 eV has to be considered [24, 57].

3.4. Oxide defects

Another interesting issue is the creation or modification of defects by diffusing hydrogen. Although for ultra-thin oxides the influence of oxide traps on NBTI is controversial [5, 6], for thick oxides these defects seem to be important [29]. Some investigations report that roughly the same number of positive fixed charges as depassivated P_b centers are created [58], while others attribute NBTI induced V_{th} shifts totally to depassivated P_b centers [5], provided proper stressing conditions are chosen ($E_{ox} < E_{crit}$). Most positive charges are located close to the interface and have been identified as E' centers (thermal oxide hole traps) [15]. E' centers have been reported to dominate oxide hole trapping with their density being strongly process dependent [15]. It has been shown that E' centers react rapidly with H_2, even at room temperature, turning them into hydrogen complexed E' centers (E'H) according to [53]

$$\mathrm{H}_2 + E' \rightleftharpoons E'\mathrm{H} + \mathrm{H}^0. \tag{7}$$

In addition, trapping of H^0 has been reported [59]

$$\mathrm{H}^0 + E' \rightleftharpoons E'\mathrm{H}\,. \tag{8}$$

Of particular interest in the case of NBTI is the annealing of E' centers through H_2, which was reported to bring up roughly the same amount of P_b centers [15], possibly through the following reaction, with H_2 formally being a catalyst

$$P_b\mathrm{H} + \mathrm{H}_2 + E' \rightleftharpoons P_b + \mathrm{H}_2 + E'\mathrm{H}. \tag{9}$$

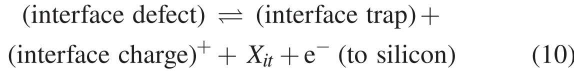
The atomic hydrogen released in the various reactions is commonly assumed to either quickly dimerize into H_2 and diffuse towards the poly gate [4, 5], assuming classical diffusion, or to move dispersively as H^+ [21, 22].
Hydrogen motion in the silicon bulk is normally neglected. This might be justified in the case of H_2 based models by the large diffusion barrier found in theoretical studies [60], or in the case of H^+ by the negative bias driving the protons towards the gate. Provided that the breaking of PH bonds in the Si bulk is an important source of H^0 and H^+ [26], transport in Si must be included in a rigorous model.

4. Models

Using a combination of some selected mechanisms summarized above a considerable number of different NBTI models has been proposed. Of particular interest in that case is the fact that although these models rely on different microscopic contexts, they all seem to reproduce the measurement data published alongside them.

4.1. The reaction-diffusion model

The reaction-diffusion model dates back to the work of Jeppson and Svensson [8]. They assumed an electrochemical reaction of the form



$$\begin{aligned}&(\text{interface defect}) \rightleftharpoons (\text{interface trap}) + \\ &(\text{interface charge})^{+} + X_{it} + \mathrm{e}^{-}\ (\text{to silicon})\end{aligned} \tag{10}$$

at the interface where the hydrogen species created at the interface ($X_{it}(t) = X(0,t)$) is assumed to diffuse away into the oxide. The exact chemical composition of the diffusing species is still under debate, although strong arguments for H_2 have been presented [4, 5, 12, 19]. In the reaction-diffusion model H_2 results in a characteristic time exponent of $1/6$, closest to experimentally observed values. Atomic hydrogen, on the other hand, with an exponent of $1/4$ was favored in older publications, consistent with measurement data available at that time. H^+, previously neglected, because it results in an exponent of $1/2$, recently entered the scene [21–23] because dispersive transport models seem to allow to adjust the slope to smaller values. As of yet, however, the scatter in the experimentally observed time exponents is still too large to settle for a single diffusing species and one explanation could be the simultaneous creation, diffusion, and interaction of several hydrogen species [61].

The kinetic equation describing the interface reaction is [9, 56, 62]

$$\frac{\partial N_{it}}{\partial t} = k_f \left(N_0 - N_{it}\right) - k_r N_{it} X_{it}^{1/a}, \tag{11}$$

where N_{it} is the surface state concentration, N_0 the initial concentration of passivated interface defects, k_f and k_r the field and temperature dependent rate coefficients, while a is the kinetic exponent (1 for H^0 and H^+, and 2 for H_2). Note that the same equation is obtained assuming instantaneous dimerization of H^0 into H_2 [56].

Transport of the species X away from the interface is assumed to be controlled by conventional drift-diffusion

$$\frac{\partial X}{\partial t} = -\nabla \cdot \mathbf{F}_X, \tag{12}$$

$$\mathbf{F}_X = -D_X \nabla X + Z_X X \mu_X \mathbf{E}. \tag{13}$$

Here, D_X, μ_X, and Z_X are the diffusion coefficient, the mobility and the charge state of species X in the medium. Diffusivity and mobility are assumed to be independent of the electric field and to be related via the Einstein relation [61]:

$$\mu_X = \frac{q D_X}{k_B T_L} = \frac{D_X}{V_T}. \tag{14}$$

Note that this is not the case for a dispersive medium where strongly different temperature dependencies can be observed in the equilibrium regime. At the boundary we have to consider the influx of the newly created species

$$a\frac{\partial N_{it}}{\partial t} = \mathbf{F}_X \cdot \mathbf{n}. \tag{15}$$

For the general case, Eqs. (11)–(15) can be solved numerically. However, for some special cases analytical approximations can be given [18, 56, 63] which are helpful for the understanding of the basic kinetics. One finds different phases, starting from the reaction dominated regime with slope $n = 1$, where the reverse rate is negligible due to the lack of available X, a transition regime with slope $n = 0$, the quasi-equilibrium regime with $\partial N_{it}/\partial t \approx 0$, which is the dominant regime and displays the characteristic time exponent depending on the created species, and a saturation

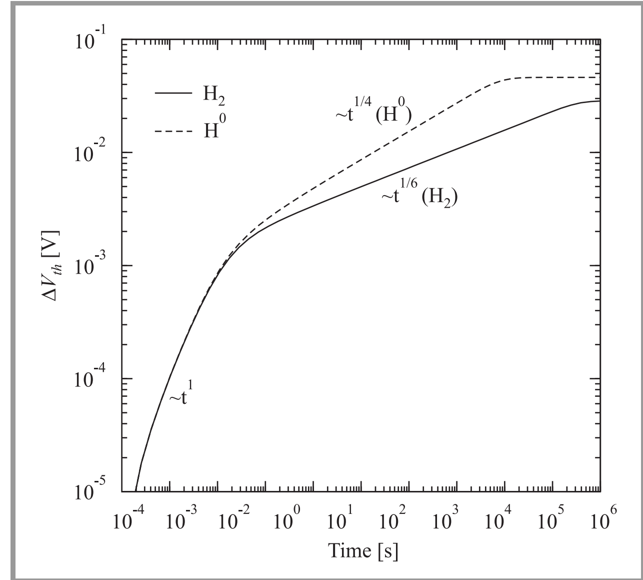


Fig. 4. The three most important phases in the reaction-diffusion model. Shown are the results for H^0 and H_2 kinetics. The time exponent $n = 1$ is the signature of the reaction limited phase while $n = 1/4$ and $n = 1/6$ result from the diffusion limited phase.

regime. These phases are shown in Fig. 4 for H^0 and H_2 kinetics. Of particular interest is the result for $\partial N_{it}/\partial t \approx 0$, which is the one normally measured during NBTI stress. For atomic hydrogen one obtains

$$N_{it}(t) = \sqrt{\frac{k_f N_0}{2k_r}} \left(D_X t\right)^{1/4}, \tag{16}$$

while molecular hydrogen results in

$$N_{it}(t) = \left(\frac{k_f N_0}{2k_r}\right)^{2/3} \left(D_X t\right)^{1/6} \tag{17}$$

and the hydrogen proton gives

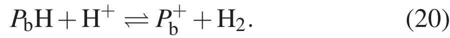
$$N_{it}(t) = \sqrt{\frac{k_f N_0}{k_r}} \left(\mu_X E_{ox} t\right)^{1/2}. \tag{18}$$

Of course, the characteristic exponents for each species given above rely on the validity of the reaction-diffusion model and different exponents can be envisaged using an alternative model [21–23].

From the calculated interface state density N_{it} the threshold voltage shift has to be determined. This is performed by assuming that all traps are positively charged, an assumption fulfilled during strong negative bias where the Fermi-level is close to the valence band edge. In addition, the generated oxide charges are often neglected and we obtain

$$\Delta V_{th} = -\frac{\Delta Q_{it}(E_F) + \Delta Q_{ot}}{C_{ox}} \approx -\frac{q\Delta N_{it}}{C_{ox}} \,. \qquad (19)$$

Due to the increased Coulomb scattering at the interfacial layer caused by charged interface states the carrier mobility decreases [64, 65]. However, for the analysis of NBTI, any potential mobility degradation has so far been neglected [20].
Based on first-principles calculations, the dissociation of P_bH through H^+ has been suggested as the dominant reaction [26], thereby replacing reaction Eq. (5):



$$P_b\mathrm{H} + \mathrm{H}^+ \rightleftharpoons P_b^+ + \mathrm{H}_2. \qquad (20)$$

The required H^+ is provided through broken PH bonds in the silicon bulk inversion layer. After overcoming the migration barrier at the interface, some H^+ diffuses along the interface before depassivating P_b centers. Alternatively, some H^+ can surmount the energy barrier towards the SiO_2 where they quickly drift to the gate due to the strong electric field.

4.2. Models assuming dispersive transport

Based on simplified solutions for the MT problem NBTI models have been developed [22, 30]. For the extremely non-equilibrium case Arkhipov and Rudenko derived [43]

$$\frac{X(\mathbf{x},t) - X_0(\mathbf{x})}{\tau(t)} = -\nabla \cdot \mathbf{F}_X(\mathbf{x},t) \qquad (21)$$

which describes the broadening of the initial distribution $X_0(\mathbf{x})$. Here the flux is given through an "effective" flux of the total concentration of the species X rather than the concentration in the conduction states. Note, that there is no time derivative in Eq. (21) and the dynamics of the system have been incorporated into $\tau(t)$ which directly depends on the density of states. Starting from Eq. (21), assuming an exponential density of states with a characteristic energy E_0 and density N_t, and $\partial N_{it}/\partial t \approx 0$, Kaczer *et al.* derived [22]

$$N_{it}(t) = \sqrt{\frac{k_f N_0}{k_r}} \left(\frac{D_X}{\nu_0} \frac{N_c}{N_t} \right)^{1/4} (\nu_0 t)^{\alpha/4}, \qquad (22)$$

with $\alpha = k_B T_L/E_0$. Interestingly, the time exponent is given here through $n = \alpha/4$ and thus explicitly depends on temperature, a phenomenon sometimes observed experimentally [4, 6, 22], but also ascribed to a measurement artifact [66]. Note that α is in the range $0 \le \alpha \le 1$ with $\alpha = 1$ being the diffusive limit.
Using simple arguments from statistical mechanics in addition to the assumption of dispersive transport of H^+ inside the oxide, Zafar [23] derived a stretched-exponential relation for the threshold voltage shift as

$$\frac{\Delta V_{th}(t)}{\Delta V_{th,\max}} = 1 - \exp\left(\left(\frac{t}{\tau} \right)^{-\alpha} \right) . \qquad (23)$$

Here, $\Delta V_{th,\max}$ gives the maximum threshold voltage shift, τ is the characteristic time constant, and α the dispersion parameter. At early times, the above equation assumes a similar form as the expression derived by Kaczer *et al.* [22].

4.3. Reaction-limited models

Houssa *et al.* [67] base their NBTI model for nitrided oxides on the assumption that the dissociation rate is determined by electron and hole tunneling currents. Employing a Gaussian distribution of activation energies, interface traps are generated by releasing a proton which is later trapped inside the oxide forming oxide charges. With these assumptions the contributions due to electrons and holes can be separated, indicating a dominance of hole induced damage at operating voltages. Consequently, if for accelerated tests higher voltage levels are used, the electron contribution begins to dominate which makes long-term extrapolation difficult.
This Gaussian distribution of activation energies was also used as the main ingredient in the reaction-limited model of Huard *et al.* [68] who assume an energy-dependent distribution of the dissociation activation energy

$$g(E_d, \sigma) = \frac{1}{\sigma} \frac{\exp\left(\frac{E_{dm} - E_d}{\sigma}\right)}{\left(1 + \exp\left(\frac{E_{dm} - E_d}{\sigma}\right)\right)^2} \,. \qquad (24)$$

In the above equation the median dissociation energy E_{dm} was assumed to depend on the oxide electric field in order to accommodate the reported field dependence. The threshold voltage shift was then derived as

$$\frac{\Delta V_{th}}{\Delta V_{th,\max}} = \frac{1}{1 + \left(\frac{t}{\tau}\right)^{-\alpha}} \qquad (25)$$

with $\tau = \tau_0 \exp(E_d(E_{ox})/k_B T_L)$ and $\alpha = k_B T_L/\sigma$. Again, as with the dispersive transport model, a temperature dependent slope is obtained.

5. Other modeling issues

In addition to the issues raised above, some further considerations are required in order to develop a comprehensive model. They are summarized in the following.

5.1. Boundary conditions

An important issue is the behavior of the hydrogen species when they encounter the SiO_2/p-Si interface. Commonly, simplified boundary conditions to the diffusion equation

are assumed, either perfect reflection [9, 69], perfect absorber [9], or perfect transmitter [4] (no trapping). However, for a rigorous treatment one might have to consider the energy barriers [60], the creation and passivation of P_b centers [28], and re-emission of hydrogen on the p-Si side, analogous to the Si/SiO_2 interface and models used in process-simulation [70].

5.2. Geometry dependence

Negative bias temperature instability is commonly assumed to be a one-dimensional process [71], which is in agreement with many reported results, while only the closely related damage caused by hot-carrier injection is acknowledged to require a two-dimensional treatment of the diffusion equation. Even if all processes leading to NBTI were one-dimensional, inhomogeneous doping profiles [72], variable oxide thicknesses such as found in high-voltage devices, inhomogeneities observed around shallow trench isolations, or inhomogeneous stress conditions ($V_{DS} \neq 0$) [72] require a two- or even three-dimensional description of the problem. Even for homogeneous stress ($V_{DS} = 0$) a gate length dependence is occasionally reported [3]. For highly scaled MOSFETs or MOSFETs with a narrow channel the geometry influences the diffusion of the released hydrogen species, an effect contained in the classic reaction-diffusion model [73]. Other explanations are based on diffusion of H^+ along the interface as observed experimentally [74] and confirmed theoretically [36].

5.3. Coupling to semiconductor equations

A commonly neglected issue in NBTI modeling is the coupling of the "hydrogen equations" to the semiconductor device equations for current transport. In particular, the dynamic creation and annihilation of P_b and E' centers influences the electric field distribution and thus the reaction rates and the transport properties. This issue is of particular importance when annealing during measurements [17] is to be understood. Some issues need to be resolved when such a coupling is attempted. First, the charge trapped in the amphoteric P_b centers depends on the position of the Fermi-level and thus on the bias conditions. To model this effect, the density of P_b centers created needs to be coupled to the electrically active interface trap density-of-states $D_{it}(\mathcal{E})$ in a surface recombination process [75]. A lot of information on D_{it} is available and it is known that in addition to the band-tail states P_b centers introduce two distinct peaks in the Si band gap [14, 55]. The shape of these peaks has been described using Fermi functions [13] where the two peak values evolve differently in time with each width staying roughly constant [15, 55]. Regarding the contribution of trapped holes in the oxide, precise statements on where exactly these charges are located are important to properly model the shape of the band-edges in SiO_2, which directly influence the oxide field and thus charge carrier transport and tunneling rates.

A specific coupling issue concerns the influence of holes which are commonly assumed to be "available". The dissociation rate in Eq. (5) is often assumed to depend on the concentration of the inversion layer holes, a quantity not directly available in NBTI models. Here, a rigorous coupled solution should provide better estimates. Although the importance of holes in this process is widely acknowledged, the mechanisms have not yet been evaluated rigorously and it is not clear in which way they influence the forward rate. Furthermore, electrons and holes might be required to properly account for charging and discharging of oxide and near-interface defects.

6. Conclusions

Although significant progress regarding the understanding of NBTI has been made in the last decade, and the reaction-diffusion model gives good qualitative agreement with many measurements, various microscopic details are still unclear. Among those are the oxide charges created during stress, the nature and transport mechanism of the created species, and the relaxation behavior. Many more consistent sets of experiments are required to aid the development and evaluation of more detailed models. Thereby the relative importance and the complex interplay between the various processes reported could be clarified.

References

[1] A. Goetzberger and H. E. Nigh, "Surface charge after annealing of Al-SiO_2-Si structures under bias", *Proc. IEEE*, vol. 54, no. 10, pp. 1454–1454, 1966.

[2] E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, "Characteristics of the surface-state charge (Qss) of thermally oxidized silicon", *J. Electrochem. Soc.*, vol. 114, no. 3, p. 266, 1967.

[3] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: road to cross in deep submicron silicon semiconductor manufacturing", *J. Appl. Phys.*, vol. 94, no. 1, pp. 1–18, 2003.

[4] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation", *Microelectron. Reliab.*, vol. 45, no. 1, pp. 71–81, 2005.

[5] S. Mahapatra, M. A. Alam, P. B. Kumar, T. R. Dalei, D. Varghese, and D. Saha, "Negative bias temperature instability in CMOS devices", *Microelectron. Eng.*, vol. 80, no. suppl., pp. 114–121, 2005.

[6] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: from physical mechanisms to modelling", *Microelectron. Reliab.*, vol. 46, no. 1, pp. 1–23, 2006.

[7] J. H. Stathis and S. Zafar, "The negative bias temperature instability in MOS devices: a review", *Microelectron. Reliab.*, vol. 46, no. 2–4, pp. 270–286, 2006.

[8] K. O. Jeppson and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices", *J. Appl. Phys.*, vol. 48, no. 5, pp. 2004–2014, 1977.

[9] S. Ogawa and N. Shiono, "Generalized diffusion-reaction model for the low-field charge build up instability at the Si/SiO_2 interface", *Phys. Rev. B*, vol. 51, no. 7, pp. 4218–4230, 1995.

[10] M. Ershov, R. Lindley, S. Saxena, A. Shibkov, S. Minehane, J. Babcock, S. Winters, H. Karbasi, T. Yamashita, P. Clifton, and M. Redford, "Transient effects and characterization methodology of negative bias temperature instability in PMOS transistors", in *Proc. Int. Rel. Phys. Symp.*, Dallas, USA, 2003, pp. 606–607.

[11] S. Rangan, N. Mielke, and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability", in *Proc. Int. Electron Dev. Meet.*, Washington, USA, 2003, pp. 341–344.

[12] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. Alam, "On the dispersive versus arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: measurements, theory, and implications", in *Proc. Int. Electron Dev. Meet.*, Washington, USA, 2005, pp. 1–4.

[13] A. Haggag, W. McMahon, K. Hess, K. Cheng, J. Lee, and J. Lyding, "High-performance chip reliability from short-time-tests", in *Proc. Int. Rel. Phys. Symp.*, Orlando, USA, 2001, pp. 271–279.

[14] E. H. Poindexter, G. J. Gerardi, M.-E. Rueckel, P. J. Caplan, N. M. Johnson, and D. K. Biegelsen, "Electronic traps and P_b centers at the Si/SiO_2 interface: band-gap energy distribution", *J. Appl. Phys.*, vol. 56, no. 10, pp. 2844–2849, 1984.

[15] P. M. Lenahan and J. F. Conley Jr., "What can electron paramagnetic resonance tell us about the Si/SiO_2 system?", *J. Vac. Sci. Technol. B*, vol. 16, no. 4, pp. 2134–2153, 1998.

[16] J. P. Campbell, P. M. Lenahan, A. T. Krishnan, and S. Krishnan, "Direct observation of the structure of defect centers involved in the negative bias temperature instability", *Appl. Phys. Lett.*, vol. 87, no. 20, pp. 1–3, 2005.

[17] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. Rey-Tauriac, and N. Revil, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's", in *Proc. Int. Electron Dev. Meet.*, San Francisco, USA, 2004, pp. 109–112.

[18] M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs", in *Proc. Int. Electron Dev. Meet.*, Washington, USA, 2003, pp. 345–348.

[19] A. T. Krishnan, C. Chancellor, S. Chakravarthi, P. E. Nicollian, V. Reddy, A. Varghese, R. B. Khamankar, and S. Krishnan, "Material dependence of hydrogen diffusion: implications for NBTI degradation", in *Proc. Int. Electron Dev. Meet.*, Washington, USA, 2005, pp. 688–691.

[20] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, and S. Krishnan, "NBTI impact on transistor and circuit: models, mechanisms and scaling effects", in *Proc. Int. Electron Dev. Meet.*, Washington, USA, 2003, pp. 14.5.1–14.5.4.

[21] M. Houssa, M. Aoulaiche, S. De Gendt, G. Groeseneken, M. M. Heyns, and A. Stesmans, "Reaction-dispersive proton transport model for negative bias temperature instabilities", *Appl. Phys. Lett.*, vol. 86, no. 9, pp. 1–3, 2005.

[22] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Temperature dependence of the negative bias temperature instability in the framework of dispersive transport", *Appl. Phys. Lett.*, vol. 86, no. 14, pp. 1–3, 2005.

[23] S. Zafar, "Statistical mechanics based model for negative bias temperature instability induced degradation", *J. Appl. Phys.*, vol. 97, no. 10, pp. 1–9, 2005.

[24] A. Stesmans, "Passivation of P_{b0} and P_{b1} interface defects in thermal (100) Si/SiO_2 with molecular hydrogen", *Appl. Phys. Lett.*, vol. 68, no. 15, pp. 2076–2078, 1996.

[25] M. Houssa, J. L. Autran, A. Stesmans, and M. M. Heyns, "Model for interface defect and positive charge generation in ultrathin SiO_2/ZrO_2 gate dielectric stacks", *Appl. Phys. Lett.*, vol. 81, no. 4, pp. 709–711, 2002.

[26] L. Tsetseris, X. J. Zhou, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Physical mechanisms of negative-bias temperature instability", *Appl. Phys. Lett.*, vol. 86, no. 14, pp. 1–3, 2005.

[27] K. Kushida-Abdelghafar, K. Watanabe, J. Ushio, and E. Murakami, "Effect of nitrogen at SiO_2/Si interface on reliability issues", *Appl. Phys. Lett.*, vol. 81, no. 23, pp. 4362–4364, 2002.

[28] M. Houssa, "Modelling negative bias temperature instabilities in advanced p-MOSFETs", *Microelectron. Reliab.*, vol. 45, no. 1, pp. 3–12, 2005.

[29] N. Stojadinović, D. Danković, S. Djorić-Veljković, V. Davidović, I. Manić, and S. Golubović, "Negative bias temperature instability mechanisms in p-channel power VDMOSFETs", *Microelectron. Reliab.*, vol. 45, no. 9-11, pp. 1343–1348, 2005.

[30] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification", in *Proc. Int. Rel. Phys. Symp.*, San Jose, USA, 2005, pp. 381–387.

[31] T. Yang, C. Shen, M. F. Li, C. H. Ang, C. X. Zhu, Y.-C. Yeo, G. Samudra, and D.-L. Kwong, "Interface trap passivation effect in NBTI measurement for p-MOSFET with SiON gate dielectric", *IEEE Electron Dev. Lett.*, vol. 26, no. 10, pp. 758–760, 2005.

[32] D. S. Ang, "Observation of suppressed interface state relaxation under positive gate biasing of the ultrathin oxynitride gate p-MOSFET subjected to negative-bias temperature stressing", *IEEE Electron Dev. Lett.*, vol. 27, no. 5, pp. 412–415, 2006.

[33] H. Reisinger, O. Blank, W. Heinrigs, A. Mühlhoff, W. Gustin, and C. Schlünder, "Analysis of NBTI degradation- and recovery-behavior based on ultra fast V_t-measurements", in *Proc. Int. Rel. Phys. Symp.*, San Jose, USA, 2006, pp. 448–453.

[34] C. G. Van de Walle and B. R. Tuttle, "Microscopic theory of hydrogen in silicon devices", *IEEE Trans. Electron Dev.*, vol. 47, no. 10, pp. 1779–1786, 2000.

[35] N. H. Nickel, A. Yin, and S. J. Fonash, "Influence of hydrogen and oxygen plasma treatments on grain-boundary defects in polycrystalline silicon", *Appl. Phys. Lett.*, vol. 65, no. 24, pp. 3099–3101, 1994.

[36] S. N. Rashkeev, D. M. Fleetwood, D. Schrimpf, and S. T. Pantelides, "Dual behavior of H^+ at Si-SiO_2 interfaces: mobility versus trapping", *Appl. Phys. Lett.*, vol. 81, no. 10, pp. 1839–1841, 2002.

[37] C. G. Van de Walle, P. J. H. Denteneer, Y. Bar-Yam, and S. T. Pantelides, "Theory of hydrogen diffusion and reactions in crystalline silicon", *Phys. Rev. B*, vol. 39, no. 15, pp. 10791–10808, 1989.

[38] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, M. Bidaud, G. Ribes, D. Barge, L. Vishnubhotla, B. Tavel, Y. Rey-Tauriac, F. Perrier, N. Revil, F. Arnaud, and P. Stolk, "New hole trapping characterization during NBTI in 65 nm node technology with distinct nitridation processing", in *Proc. Int. Integr. Reliab. Worksh.*, Lake Tahoe, USA, 2004, pp. 121–124.

[39] D. S. Ang, S. Wang, and C. H. Ling, "Evidence of two distinct degradation mechanisms from temperature dependence of negative bias stressing of the ultrathin gate p-MOSFET", *IEEE Electron Dev. Lett.*, vol. 26, no. 12, pp. 906–908, 2005.

[40] E. W. Montroll and H. Scher, "Random walks on lattices. IV. Continuous-time walks and influence of absorbing boundaries", *J. Stat. Phys.*, vol. 9, no. 2, pp. 101–135, 1973.

[41] H. Scher and E. W. Montroll, "Anomalous transit-time dispersion in amorphous solids", *Phys. Rev. B*, vol. 12, no. 6, pp. 2455–2477, 1975.

[42] J. Noolandi, "Multiple-trapping model of anomalous transit-time dispersion in *a*-Se", *Phys. Rev. B*, vol. 16, no. 10, pp. 4466–4473, 1977.

[43] V. I. Arkhipov and A. I. Rudenko, "Drift and diffusion in materials with traps", *Phil. Mag. B*, vol. 45, no. 2, pp. 189–207, 1982.

[44] D. B. Brown and N. S. Saks, "Time dependence of radiation-induced trap formation in metal-oxide-semiconductor devices as a function of oxide thickness and applied field", *J. Appl. Phys.*, vol. 70, no. 7, pp. 3734–3747, 1991.

[45] J. Orenstein, M. A. Kastner, and V. Vaninov, "Transient photoconductivity and photo-induced optical absorption in amorphous semiconductors", *Phil. Mag. B*, vol. 46, no. 1, pp. 23–62, 1982.

[46] F. B. McLean and G. A. Ausman, "Simple approximate solution to continuous-time random-walk transport", *Phys. Rev. B*, vol. 15, no. 2, pp. 1052–1061, 1977.

[47] V. I. Arkhipov, "Trap-controlled and hopping modes of transport and recombination: similarities and differences", in *Proc. Int. Symp. Elect. Insul. Mater.*, Tokyo, Japan, 1995, pp. 271–274.

[48] J. Noolandi, “Equivalence of multiple-trapping model and time-dependent random walk”, *Phys. Rev. B*, vol. 16, no. 10, pp. 4474–4479, 1977.

[49] F. W. Schmidlin, “Theory of trap-controlled photoconduction”, *Phys. Rev. B*, vol. 16, no. 6, pp. 2362–2385, 1977.

[50] S. Lathi and A. Das, “Seminumerical simulation of dispersive transport in the oxide of metal-oxide semiconductor devices”, *J. Appl. Phys.*, vol. 77, no. 8, pp. 3864–3867, 1995.

[51] N. Talwalkar, A. Das, and J. Vasi, “Dispersive transport of carriers under nonuniform electric field”,, *J. Appl. Phys.*, vol. 78, no. 7, pp. 4487–4489, 1995.

[52] W. B. Jackson and C. C. Tsai, “Hydrogen transport in amorphous silicon”, *Phys. Rev. B*, vol. 45, no. 12, pp. 6564–6580, 1992.

[53] K. L. Brower, “Passivation of paramagnetic Si/SiO_2 interface states with molecular hydrogen”, *Appl. Phys. Lett.*, vol. 53, no. 6, pp. 508–510, 1988.

[54] E. Cartier, J. H. Stathis, and D. A. Buchanan, “Passivation and depassivation of silicon dangling bonds at the Si(111)/SiO_2 interface by atomic hydrogen”, *Appl. Phys. Lett.*, vol. 63, no. 11, pp. 1510–1512, 1993.

[55] E. Cartier and J. H. Stathis, “Hot-electron induced passivation of silicon dangling bonds at the Si(111)/SiO_2 interface”, *Appl. Phys. Lett.*, vol. 69, no. 1, pp. 103–105, 1996.

[56] A. T. Krishnan, S. Chakravarthi, P. Nicollian, V. Reddy, and S. Krishnan, “Negative bias temperature instability mechanism: the role of molecular hydrogen”, *Appl. Phys. Lett.*, vol. 88, no. 15, pp. 1–3, 2006.

[57] A. Stesmans, “Dissociation kinetics of hydrogen-passivated P_b defects at the (111)Si/SiO_2 interface”, *Phys. Rev. B*, vol. 61, no. 12, pp. 8393–8403, 2000.

[58] M. Denais, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, D. Roy, and A. Bravaix, “Perspectives on NBTI in advanced technologies: modelling & characterization”, in *Proc. ESSDERC Conf.*, Grenoble, France, 2005, pp. 399–402.

[59] D. L. Griscom, “Diffusion of radiolytic molecular hydrogen as a mechanism for the post-irradiation buildup of interface states in SiO_2-on-Si structures”, *J. Appl. Phys.*, vol. 58, no. 7, pp. 2524–2533, 1985.

[60] L. Tsetseris and S. T. Pantelides, “Migration, incorporation, and passivation reactions of molecular hydrogen at the Si-SiO_2 interface”, *Phys. Rev. B*, vol. 70, no. 24, pp. 1–6, 2004.

[61] S. Chakravarthi, A. T. Krishnan, V. Reddy, C. F. Machala, and S. Krishnan, “A comprehensive framework for predictive modeling of negative bias temperature instability”, in *Proc. Int. Rel. Phys. Symp.*, Phoenix, USA, 2004, pp. 273–282.

[62] S. Ogawa, M. Shimaya, and N. Shiono, “Interface-trap generation at ultrathin SiO_2 (4 nm–6 nm)-Si interfaces during negative-bias temperature aging”, *J. Appl. Phys.*, vol. 77, no. 3, pp. 1137–1148, 1995.

[63] J. B. Yang, T. P. Chen, S. S. Tan, and L. Chan, “Analytical reaction-diffusion model and the modeling of nitrogen-enhanced negative bias temperature instability”, *Appl. Phys. Lett.*, vol. 88, no. 17, pp. 1–3, 2006.

[64] C. T. Sah, T. H. Ning, and L. L. Tschopp, “The scattering of electrons by surface oxide charges and by lattice vibrations at the silicon-silicon dioxide interface”, *Surf. Sci.*, vol. 32, no. 3, pp. 561–575, 1972.

[65] M. Kondo and H. Tanimoto, “Accurate Coulomb mobility model for MOS inversion layer and its application to NO-oxynitride devices”, *IEEE Trans. Electron Dev.*, vol. 48, no. 2, pp. 265–270, 2001.

[66] M. A. Alam, “NBTI: a simple view of a complex phenomena”, in *Proc. Int. Rel. Phys. Symp.*, San Jose, USA, 2006.

[67] M. Houssa, M. Aoulaiche, J. L. Autran, C. Parthasarathy, N. Revil, and E. Vincent, “Modeling negative bias temperature instabilities in hole channel metal-oxide-semiconductor field effect transistors with ultrathin gate oxide layers”, *J. Appl. Phys.*, vol. 95, no. 5, pp. 2786–2791, 2004.

[68] V. Huard, M. Denais, F. Perrier, N. Revil, C. Parthasarathy, A. Bravaix, and E. Vincent, “A thorough investigation of MOSFETs NBTI degradation”, *Microelectron. Reliab.*, vol. 45, no. 1, pp. 83–98, 2005.

[69] H. Aono, E. Murakami, K. Okuyama, A. Nishida, M. Minami, Y. Ooji, and K. Kubota, “Modeling of NBTI saturation effect and its impact on electric field dependence of the lifetime”, *Microelectron. Reliab.*, vol. 45, no. 7-8, pp. 1109–1114, 2005.

[70] F. Lau, L. Mader, C. Mazure, Ch. Werner, and M. Orlowski, “Model for phosphorus segregation at the silicon-silicon dioxide interface”, *Appl. Phys. A*, vol. 49, no. 6, pp. 671–675, 1989.

[71] H. Kufluoglu and M. A. Alam, “A geometrical unification of the theories of NBTI and HCI time-exponents and its implications for ultra-scaled planar and surround-gate MOSFETs”, in *Proc. Int. Electron Dev. Meet.*, San Francisco, USA, 2004, pp. 113–116.

[72] C. Schlünder, R. Brederlow, B. Ankele, W. Gustin, K. Goser, and R. Thewes, “Effects of inhomogeneous negative bias temperature stress on p-channel MOSFETs of analog and RF circuits”, *Microelectron. Reliab.*, vol. 45, no. 1, pp. 39–46, 2005.

[73] H. Kufluoglu and M. A. Alam, “Theory of interface-trap-induced NBTI degradation for reduced cross section MOSFETs”, *IEEE Trans. Electron Dev.*, vol. 53, no. 5, pp. 1120–1130, 2006.

[74] R. E. Stahlbush and E. Cartier, “Interface defect formation in MOSFETs by atomic hydrogen exposure”, *IEEE Trans. Nucl. Sci.*, vol. 41, no. 6, pp. 1844–1853, 1994.

[75] W. D. Eades and R. M. Swanson, “Calculation of surface generation and recombination velocities at the Si-SiO_2 interface”, *J. Appl. Phys.*, vol. 58, no. 11, pp. 4267–4276, 1985.

Tibor Grasser was born in Vienna, Austria, in 1970. He received the diplomingenieur degree in communications engineering, the Ph.D. degree in technical sciences, and the venia docendi in microelectronics from the Technische Universität Wien in 1995, 1999, and 2002, respectively. He is currently employed as an Associate Professor at the Institute for Microelectronics. Since 1997 he has headed the Minimos-NT development group, working on the successor of the highly successful MiniMOS program. He was a visiting research engineer for Hitachi Ltd., Tokyo, Japan, and for the Alpha Development Group, Compaq Computer Corporation, Shrewsbury, USA. In 2003 he was appointed head of the Christian Doppler Laboratory for TCAD in Microelectronics, an industry-funded research group embedded in the Institute for Microelectronics. Doctor Grasser is the co-author or author of nearly 200 articles in scientific books, journals, and conferences proceedings, is the editor of a book on advanced device simulation, a Senior Member of IEEE, and co-chair of SISPAD 2007. His current scientific interests include circuit and device simulation, device modeling and reliability issues.

e-mail: Grasser@iue.tuwien.ac.at
Christian Doppler Laboratory for TCAD
in Microelectronics
Institute for Microelectronics, TU Wien
Gußhaustraße 27-29
A-1040 Wien, Austria

Siegfried Selberherr was born in Klosterneuburg, Austria, in 1955. He received the degree of diplomingenieur in electrical engineering and the doctoral degree in technical sciences from the Technische Universität Wien in 1978 and 1981, respectively. He has been holding the venia docendi on computer-aided design since 1984. Since 1988 he has been the chair Professor of the Institute for Microelectronics. From 1998 to 2005 he served as dean of the Fakultät für Elektrotechnik und Informationstechnik. Professor Selberherr is a Fellow of the IEEE since 1993. His current research interests are modeling and simulation of problems for microelectronics engineering.

e-mail: Selberherr@iue.tuwien.ac.at
Institute for Microelectronics, TU Wien
Gußhausstraße 27–29
A-1040 Wien, Austria

Regular paper

Modeling of voice data integrated traffic in 3G mobile cellular network

Imdadul Islam, Jugal Krishna Das, and Siddique Hossain

Abstract— The most important feature of 3G mobile cellular network is introduction of voice data integrated service under multilayered cell environment to support overflow traffic of lower layered cells by upper ones. This paper deals with traffic model of three layered cells, i.e., micro cell, macro cell and satellite cell. Here a new call admission control is introduced for three layered cell of 3G mobile cellular network. State transition chain is designed for theoretical analysis of above mentioned traffic. Blocking probability of data call, new voice call and handover failure of voice call, probability of utilization of micro cell channel, macro cell channel and satellite cell channel are analyzed against different traffic parameters and yield logical results.

Keywords—3G mobile, voice data integrated service, micro cell, macro cell, satellite cell, Markovian chain and thinning scheme.

1. Introduction

Mobile cellular communication system is the most rapidly changing technology in the field of telecommunications. The coverage area of mobile cellular network has been expanded from urban micro cell to indoor pico cell. As the mobile cellular network moves towards 3G technology, one of the objectives is the use of single wireless system to support a variety of services including voice, data and voice in various forms mentioned in [1, 2]. The 3G mobile technology had combined both satellite and terrestrial networks together, and three- or four-layer cell structure (several micro cells are overlaid by a macro cell and again few macro cells are overlaid by a big satellite cell) is provided to combat network congestion summarized in [3–5]. Channel assignment and power control is really a cumbersome job for this type of cell configuration, but power control part is beyond the scope of this paper; the authors analyze here only channel allocation. Usually in multilayered cell any new arrival will first search for a channel from micro cell and will route to macro cell in case of unavailability of free channel in a given micro cell. A call will only route to satellite cell if both micro and macro cells are found occupied. If any call continuing in macro or satellite cell finds free channel in micro cell, it will take-back [6] to that micro cell; take-back might take place even between satellite cell and macro cell. Multilayer cell architecture of 3G system is shown in Fig. 1 for three layer cell structure, where any mobile station (MS) can make intra system handover from one layer to another depending on its mobility.

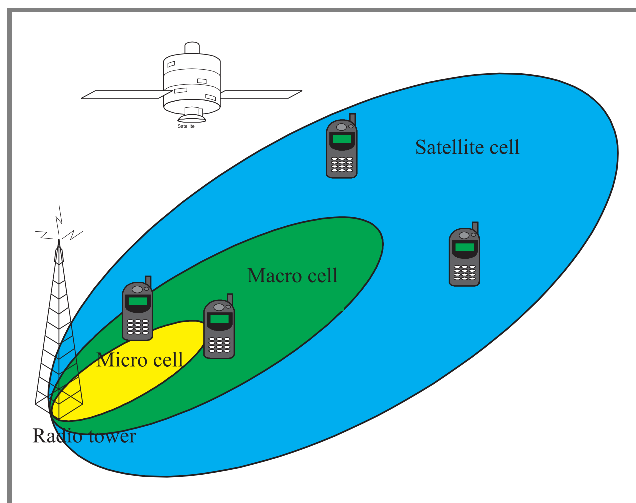


Fig. 1. Multilayer cell structure of 3G system (IMT-2000).

Section 2 of the paper deals with overflow traffic of three layered cell in generalized form and we call it composite traffic model. Section 3 presents the proposed call admission control scheme, while Subsection 3.1 implements the traffic model of Section 3 based on Markovian chain and mathematical expressions of data, new voice call and handover voice call blocking probability are derived. Section 4 depicts the results graphically for different traffic parameters. In this paper we ignored the take-back phenomenon in our new call admission scheme for simplicity of analysis; of course, the take-back is included in composite traffic mode of Section 2.

2. Composite traffic model

Here overlaid cell carries overflow traffic from underlain cell. At the same time take-back alleviates overloading of any upper layer cell in order to reduce blocking probability. State transition diagram for three layer cell structure of Fig. 1 is shown in Fig. 2, where number of channel of micro cell is n, that of macro cell is m and that of satellite cell is k. Few states are unreachable here, like [7, 8] due to the take-back phenomenon. These states are marked by crosses in Fig. 2.

Here:

μ_1 – termination rate of micro cell,

μ_2 – termination rate of macro cell,

μ_3 – termination rate of satellite cell,

λ – average arrival rate of new or handoff call on any cell.

Solution of the Markovian chain yields following equations using cut equations of [9, 10] and [18, 19]. Probability of

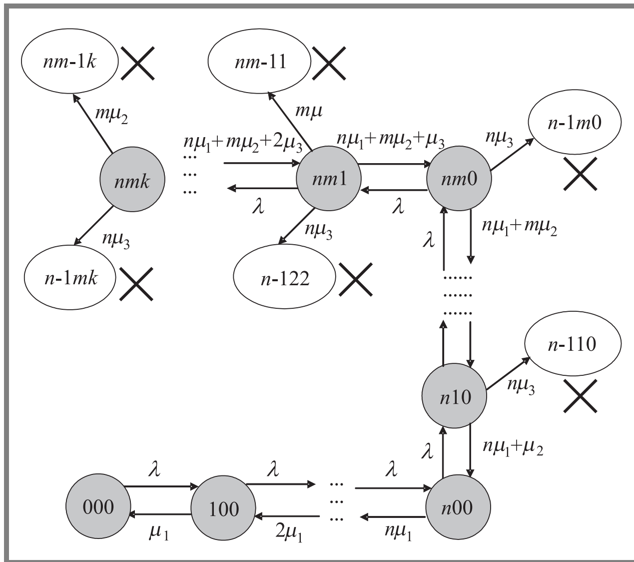


Fig. 2. Markov chain of three layer cell structure.

utilization of micro cell channel taking equal termination rate, $\mu_1 = \mu_2 = \mu_3 = \mu$ and $A = \lambda/\mu$ is derived like:

$$Mi(m) = \frac{\sum_{i=0}^{n} \frac{A^i}{i!}}{\sum_{i=0}^{n} \frac{A^i}{i!} + \frac{A^n}{n!} \sum_{j=1}^{m} \frac{A^j}{\prod_{s=1}^{j}(n+s)} + \frac{A^{n+m}}{n! \prod_{i=1}^{m}(n+i)} \sum_{t=1}^{k} \frac{A^t}{\prod_{z=1}^{t}(r+z)}}, \tag{1}$$

where $r = m+n$.

Probability of utilization of macro cell channel is defined by the following formula:

$$Ma(m) = \frac{\frac{A^n}{n!} \sum_{j=1}^{m} \frac{A^j}{\prod_{s=1}^{j}(n+s)}}{\sum_{i=0}^{n} \frac{A^i}{i!} + \frac{A^n}{n!} \sum_{j=1}^{m} \frac{A^j}{\prod_{s=1}^{j}(n+s)} + \frac{A^{n+m}}{n! \prod_{i=1}^{m}(n+i)} \sum_{t=1}^{k} \frac{A^t}{\prod_{z=1}^{t}(r+z)}}. \tag{2}$$

And the probability of utilization of satellite cell channel is

$$Sa(m) = \frac{\frac{A^{n+m}}{n! \prod_{i=1}^{m}(n+i)} \sum_{t=1}^{k} \frac{A^t}{\prod_{z=1}^{t}(r+z)}}{\sum_{i=0}^{n} \frac{A^i}{i!} + \frac{A^n}{n!} \sum_{j=1}^{m} \frac{A^j}{\prod_{s=1}^{j}(n+s)} + \frac{A^{n+m}}{n! \prod_{i=1}^{m}(n+i)} \sum_{t=1}^{k} \frac{A^t}{\prod_{z=1}^{t}(r+z)}}. \tag{3}$$

Call blocking probability equal is

$$B(m) = \frac{\frac{A^{m+n+k}}{n! \prod_{i=1}^{m}(n+i)} \sum_{t=1}^{k}(r+t)}{\sum_{i=0}^{n} \frac{A^i}{i!} + \frac{A^n}{n!} \sum_{j=1}^{m} \frac{A^j}{\prod_{s=1}^{j}(n+s)} + \frac{A^{n+m}}{n! \prod_{i=1}^{m}(n+i)} \sum_{t=1}^{k} \frac{A^t}{\prod_{z=1}^{t}(r+z)}}. \tag{4}$$

3. New call admission control scheme

In the previous section we did not distinguish between data and voice traffic where call arrival rate λ is simply ruled by Poisson's distribution. Arrivals include data call, new voice call and voice handover call. In this call admission scheme we put emphasis on voice traffic rather than data traffic. Here both data and voice traffic share channels of a micro cell. Any data arrival beyond the capacity of micro cell will be blocked, but overflow voice traffic will be supported by the overlaid macro cell.

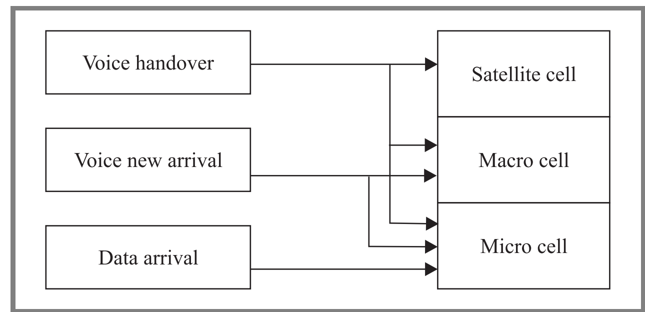


Fig. 3. Proposed call admission control.

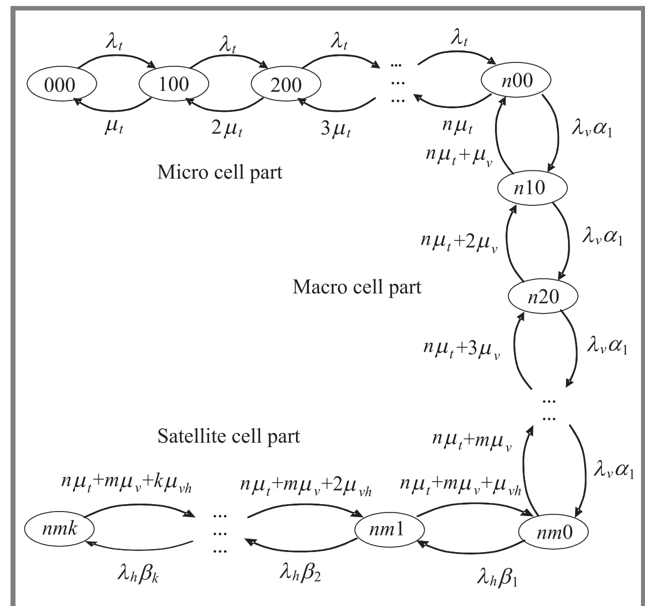


Fig. 4. State transition chain of the proposed scheme.

$$Bd(n,m,k) = \frac{\sum_{i=n}^{n} \frac{A_t^i}{i!}}{\sum_{i=0}^{n} \frac{A_t^i}{i!} + \frac{A_t^n}{n!} \sum_{s=1}^{m} \frac{\lambda_v^s \prod_{i=1}^{s} \alpha_i}{\prod_{j=1}^{s} (n\mu_t + j\mu_v)} + \frac{\lambda_v^m \prod_{i=1}^{m} \alpha_i}{\prod_{j=1}^{m} (n\mu_t + j\mu_v)} \frac{A_t^n}{n!} \sum_{r=1}^{k} \frac{\lambda_h^r \prod_{j=1}^{r} \beta_j}{\prod_{i=1}^{r} (n\mu_t + m\mu_v + i\mu_{vh})}} \tag{5}$$

$$V(n,m,k) = \frac{\frac{A_t^n}{n!} \sum_{s=m}^{m} \frac{\lambda_v^s \prod_{i=1}^{s} \alpha_i}{\prod_{j=1}^{s} (n\mu_t + j\mu_v)}}{\sum_{i=0}^{n} \frac{A_t^i}{i!} + \frac{A_t^n}{n!} \sum_{s=1}^{m} \frac{\lambda_v^s \prod_{i=1}^{s} \alpha_i}{\prod_{j=1}^{s} (n\mu_t + j\mu_v)} + \frac{\lambda_v^m \prod_{i=1}^{m} \alpha_i}{\prod_{j=1}^{m} (n\mu_t + j\mu_v)} \frac{A_t^n}{n!} \sum_{r=1}^{k} \frac{\lambda_h^r \prod_{j=1}^{r} \beta_j}{\prod_{i=1}^{r} (n\mu_t + m\mu_v + i\mu_{vh})}} \tag{6}$$

$$HF(n,m,k) = \frac{\frac{\lambda_v^m \prod_{i=1}^{m} \alpha_i}{\prod_{j=1}^{m} (n\mu_t + j\mu_v)} \frac{A_t^n}{n!} \sum_{r=k}^{k} \frac{\lambda_h^r \prod_{j=1}^{r} \beta_j}{\prod_{i=1}^{r} (n\mu_t + m\mu_v + i\mu_{vh})}}{\sum_{i=0}^{n} \frac{A_t^i}{i!} + \frac{A_t^n}{n!} \sum_{s=1}^{m} \frac{\lambda_v^s \prod_{i=1}^{s} \alpha_i}{\prod_{j=1}^{s} (n\mu_t + j\mu_v)} + \frac{\lambda_v^m \prod_{i=1}^{m} \alpha_i}{\prod_{j=1}^{m} (n\mu_t + j\mu_v)} \frac{A_t^n}{n!} \sum_{r=1}^{k} \frac{\lambda_h^r \prod_{j=1}^{r} \beta_j}{\prod_{i=1}^{r} (n\mu_t + m\mu_v + i\mu_{vh})}} \tag{7}$$

Only the handover part of voice arrival will further be supported by satellite cell. The phenomenon mentioned above is shown in Fig. 3 based on concept presented in papers [11–13].

Channels of macro and satellite cells resemble the guard channel of handover call, hence probability of forced termination/handover failure will be reduced to minimum. The whole phenomenon is represented by state transition diagram of Fig. 4 like [14, 15] and equations of call blocking probability, probability of utilization of micro cell channel, macro cell channel and satellite cell channel are derived from cut equations of Markovian chain.

3.1. Traffic modeling

State transition diagram for call admission scheme of previous section is depicted in Fig. 4. As channels of macro and satellite cells are shared among several micro cells, the probability of getting free channel of macro and satellite cells will be gradually decreased, hence thinning scheme of [16, 17] is applicable to this traffic model.

Here:

λ_t – total call arrival rate of voice data integrated call,

μ_t – total termination rate of voice and data call,

λ_v – voice call arrival rate,

μ_v – termination rate of voice call,

λ_h – voice handover call arrival rate,

μ_{vh} – termination rate of voice handover call,

n – number of channels of micro cell,

m – number of channels of macro cell,

k – number of channels of satellite cell,

$\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4 > \ldots$ – probabilities of thinning scheme for micro cell,

$\beta_1 > \beta_2 > \beta_3 > \beta_4 > \ldots$ – probabilities of thinning scheme for macro cell.

Solution of the Markovian chain yields the following equations using cut equations presented in [18,19].

Equations (5)–(7) shown at the top of this page define the following parameters:

- blocking probability of data call $Bd(n,m,k)$ – Eq. (5);
- blocking probability of new voice call $V(n,m,k)$ – Eq. (6);
- probability of handover failure of voice call $HF(n,m,k)$ – Eq. (7).

4. Results

Figure 5 shows graphs of Eqs. (1)–(4) plotted against number of channels. The graph reveals that probability of utilization of micro cell channel increases but that of macro and satellite cell channel decrease with rising number of micro cell channels. Again, probability of utilization of micro cell channel remains fixed, that of macro cell increases, but that of satellite cell channel decreases with increased number of macro cell channels. Probability of utilization of both micro and macro cell channel remains approximately constant, but that of satellite cell channel increases with increase in number of satellite cell channels k. Call blocking probability decreases for all three cases with increase in number of channels.

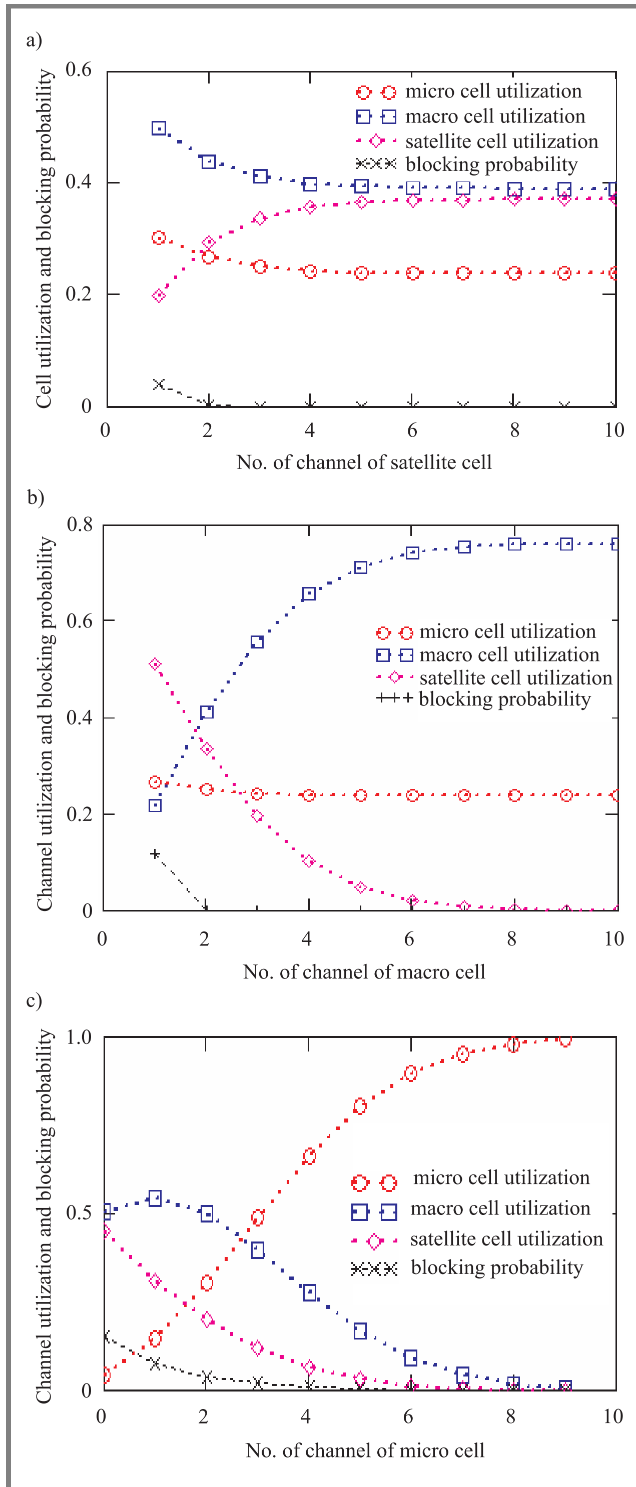


Fig. 5. Channel utilization and blocking probability for three different cases: (a) $n = 2$, $m = 2$; (b) $n = 2$, $k = 3$; (c) $m = 2$, $k = 1$.

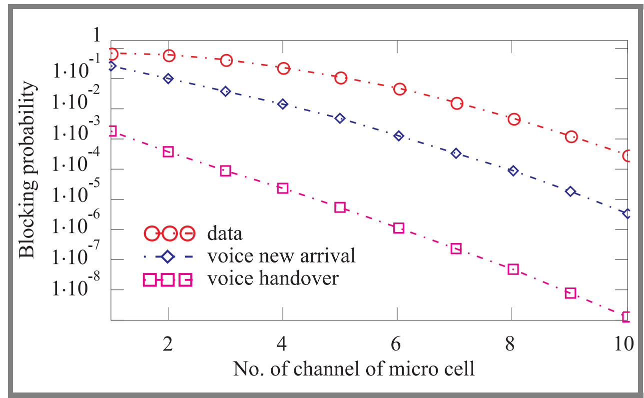


Fig. 6. Call blocking probability against channel of micro cell.

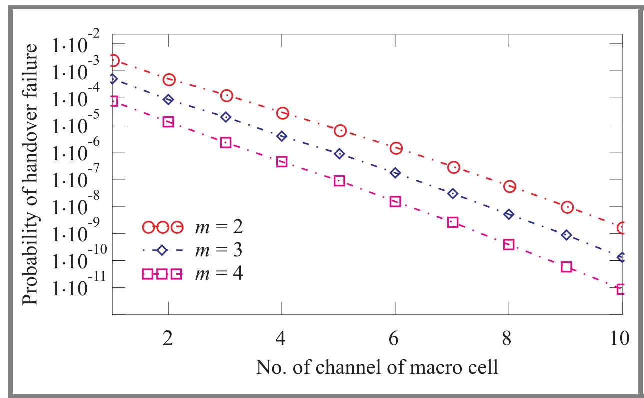


Fig. 7. Impact of number of macro cell channels on handover failure.

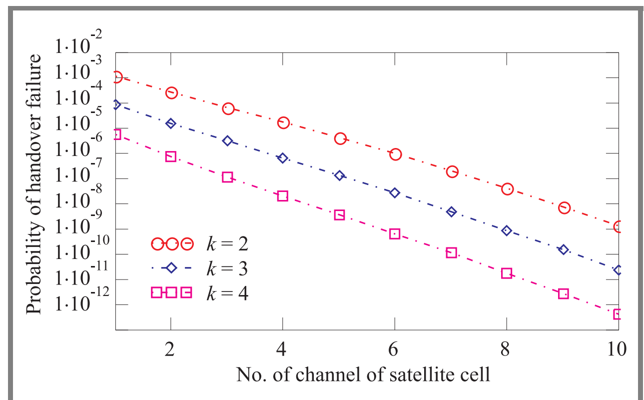


Fig. 8. Impact of number of satellite cell channels on handover failure.

In our proposed call admission scheme the blocking probability of micro cell includes only blocking of data call, blocking of macro cell only covers new arrival part of voice call and finally blocking of satellite cell results in handover failure.

Effects of those phenomena are shown in Fig. 6, assuming: $\lambda_t = 10$ calls/min, $\mu_t = 4$ calls/min, $\lambda_v = 6$ calls/min, $\mu_v = 3$ calls/min, $\lambda_h = 1.2$ calls/min, $\mu_{vh} = 0.8$ calls/min, $\alpha_1 = 0.9$, $\alpha_2 = 0.8$, $\alpha_3 = 0.75$, $\alpha_4 = 0.7$, $\beta_1 = 0.9$, $\beta_2 = 0.8$, $\beta_3 = 0.75$, and $\beta_4 = 0.7$. The probability of handover failure against number of channel of micro cell is plotted in Figs. 7 and 8, taking channel count of macro and satellite cell as parameter.

5. Conclusions

It is demonstrated in previous sections that, when increasing the channel capacity of any layer, the cell or channel utilization of that layer increases with decreased or constant utilization of other layers, which validates the composite traffic analysis. Falling slope and decreased height of curves with increase of m and k validates the analysis of the proposed call admission scheme. In this traffic model, the handover voice call gets maximum opportunity of using trunk of overlaid cells to minimize probability of forced termination. Handover part of data call were included in total arrival part of traffic. It would not be a difficult job to implement a different call admission control scheme providing guard channel of macro cell to carry data handover traffic. Here all traffic parameters were evaluated based on one dimensional Markovian chain of Figs. 2 and 4. Still there is a room for improvement of results by considering three dimensional chain of three types of arrival traffic (data, new voice call and handover voice arrival). Inclusion of impatient users and take-back could further improve modeling accuracy.

References

[1] K. Zhang and K. Pahlavan, “An integrated voice/data system for mobile indoor radio networks”, *IEEE Trans. Veh. Technol.*, vol. 39, no. 1, pp. 75–82, 1990.

[2] K. Pahlavan and P. Krisnamurthy, *Principles of Wireless Networks*. New Delhi: Prentice-Hall, 2003.

[3] B. Jabbari and W. F. Fuhrmann, “Traffic modeling and analysis of flexible hierarchical cellular networks with speed-sensitive handoff strategy”, *IEEE J. Selec. Areas Commun.*, vol. 15, no. 8, pp. 1539–1548, 1997.

[4] ———, “An analytical model of traffic performance of mobile cellular network in underlay overlay cell system”, in *Proc. 6th Int. Conf. Comput. Inform. Technol. ICCIT*, Dhaka, Bangladesh, 2003, vol. 1, pp. 230–234.

[5] K. L. Yeung and S. Nanda, “Channel management in micro/macro cellular radio systems”, *IEEE Trans. Veh. Technol.*, vol. 45, no. 4, pp. 14–23, 1996.

[6] P. Fitzpatrick, C. S. Lee, and B. Warfield, “Teletraffic performance of mobile radio networks with hierarchical cells and overflow”, *IEEE J. Selec. Areas Commun.*, vol. 15, no. 8, pp. 1549–1557, 1997.

[7] G. D. Morley and W. D. Grover, “Strategies to maximize carried traffic in dual-mode cellular systems”, *IEEE Trans. Veh. Technol.*, vol. 49, no. 2, pp. 357–366, 2000.

[8] J. Mar and J. P. Huang, “Traffic performance analysis of integrated dual-band cellular radio network”, *IEE Proc. Commun.*, vol. 147, no. 3, pp. 180–187, 2000.

[9] ———, “An analytical model of performance evaluation of SDMA based mobile cellular network for limited user traffic case”, *Jahangirnagar Univ. J. Sci.*, vol. 27, pp. 134–143, 2004.

[10] ———, “An analytical model of low dense network traffic for LEO mobile satellite systems”, *J. Electron. Comput. Sci.*, vol. 5, pp. 31–44, 2004.

[11] B. Li, L. Li, B. Li, K. M. Sivalingam, and X.-R. Cao, “Call admission control for voice/data integrated cellular networks: performance analysis and comparative study”, *IEEE J. Selec. Areas Commun.*, vol. 22, no. 4, pp. 706–718, 2004.

[12] M. Mahdavi, R. M. Edwards, and S. R. Cvetkovic, “Policy for enhancement of traffic in TDMA hybrid switched integrated voice/data cellular mobile communications systems”, *IEEE Commun. Lett.*, vol. 5, no. 6, pp. 242–244, 2001.

[13] Q. Hung, S. Chan, K.-T. Ko, and M. Zukerman, “An enhanced handoff control scheme for multimedia traffic in cellular networks”, *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 195–197, 2004.

[14] T. Veerarajan, *Probability, Statistics and Random Process*. New Delhi: Tata McGraw-Hill, 2003.

[15] S. M. Ross, *Introduction to Probability Models*. San Diego: Academic Press, 2001.

[16] R. Ramjee, R. Nagarajan, and D. Towsley, “On optimal call admission control in cellular networks”, *Wirel. Netw. J.*, vol. 3, no. 1, pp. 29–41, 1997.

[17] Y. Fang, “Thinning scheme for call admission control in wireless networks”, *IEEE Trans. Comput.*, vol. 52, no. 5, pp. 685–687, 2003.

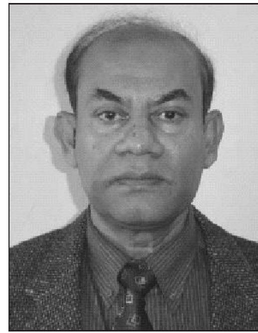
[18] ———, “Analysis of two dimensional limited source and mixed traffic model for a BTS of small mobile cellular network”, in *Proc. 4th Int. Conf. Comput. Inform. Technol. ICCIT*, Dhaka, Bangladesh, 2001, pp. 190–196.

[19] G. M. Galvan-Tejada and J. G. Gardiner, “Theoretical model to determine the blocking probability for SDMA systems”, *IEEE Trans. Veh. Technol.*, vol. 50, no. 5, pp. 1279–1288, 2001.

Imdadul Islam has completed his B.Sc. engineering in electrical and electronic engineering from Bangladesh University of Engineering and Technology, Dhaka, in 1993 and M.Sc. engineering from the same institute in 1998. Now he is perusing Ph.D. at the department of CSE, J.U., Dhaka, in the field of teletraffic engineering. He worked as an Assistant Engineer in Sheba Telecom (Pvt.) Ltd. (a joint venture company between Bangladesh and Malaysia, for mobile cellular and WLL), from Sept. 1994 to July 1996. He has very good field experiences in installation of radio base station and switching center for WLL. He is now working as an Associate Professor, at the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. His research field is network traffic, OFDMA, WCDMA and array antenna

system. He has more than fifity papers in national and international journal and conference proceedings.
e-mail: imdad22000@yahoo.com
Department of Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka, Bangladesh

Jugal Krishna Das was born in Shariatpur in 1964. He received his M.Sc. and Ph.D. in computer engineering from Donetsk Technical University of Ukraine in 1989 and 1993, respectively. He joined as a lecturer at the Department of Electronics and Computer Science, Jahangirnagar University in 1994 and now serving as a Professor at the Department of Computer Science and Engineering of the same university. His research interests are in mobile ad hoc network and distributed system.
e-mail: drdas64@yahoo.com
Department of Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka, Bangladesh



Siddique Hossain is the senior most Professor of Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka. He has more than 34 years of teaching, research and administrative experience both at home and abroad. Worked as the Head, EEE Department, Head, CSE Department, Dean, EEE Faculty and Director, BUET Computer Centre. He worked as a visiting faculty member of more than ten universities. He is a Senior Member of IEEE and was engaged as Chief of IEEE in Bangladesh in 1997–1998. He is interested in the field of computers and communication engineering, 3G mobile communication, wideband code division multiple access (WCDMA) for universal mobile telecommunication system (UMTS), etc. He has more than 40 publications in national, international journals and conference proceedings.
e-mail: sdq@eee.buet.ac.bd
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Regular paper

Simple method for characterization of photonic crystal fibers

Mieczysław Szustakowski, Norbert Pałka, and Waldemar Grabiec

Abstract— We report on our experimental characterization of index-guiding photonic crystal fibers (PCF) from their far field intensity distribution. The algorithm presented below makes it possible to determine the geometrical parameters of the PCF (core diameter, air hole spacing and air hole diameter) from its far field pattern. We obtained good agreement with the manufacturer's data for all fibers tested.

Keywords—photonic crystal fibers, far field distribution, characterization.

1. Introduction

Over the past few years, a substantial progress has been made in fabricating photonic crystal fibers (PCF) – new dielectric structures with a refractive index that varies periodically in the transverse plane, with a period of the order of an optical wavelength [1]. Regular morphological microstructure incorporated into the material radically alters its optical properties. Two guiding mechanisms are possible in PCFs: index guiding (IG) or photonic band gap (PBG) [2]. As far as the IG case is concerned (Fig. 1a), a waveguide consists of a solid core and a cladding with an array of air holes. The guided modes may be trapped in a core with a higher refractive index than an averaged index of the cladding. On the other hand, the PBG fibers have a *hollow* core and also an array of air holes in a cladding. Now, the modes are trapped in a core of *lowered* index by a photonic band gap effect.

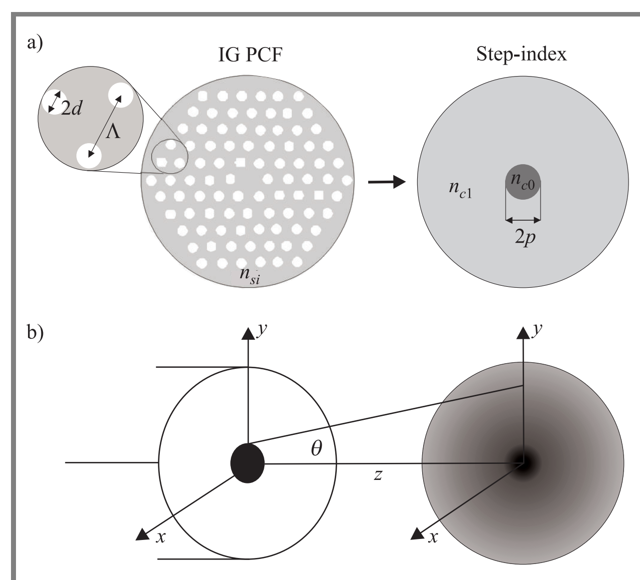


Fig. 1. The idea of effective index method (a) and far field intensity distribution scheme with angle θ and distance z (b).

This mechanism is based on Bragg reflection and prevents the light from propagating in a cladding material – so it has to propagate inside the hollow core. These two guiding mechanisms as well as a possibility of the tailoring fiber geometry determine the diverse nature and properties of PCFs [2, 3].

Because of their wide applications, PCFs require a simple method for characterization of their basic geometrical parameters: core diameter (ρ), air hole spacing (Λ) and air hole diameter (d) – see Fig. 1a. We report here on our experimental evaluation of characterization of index-guiding photonic crystal fibers from far field intensity distribution proposed by Varshney and Sinha [4].

2. Theory

An effective index method (EIM) is a simple tool that can provide a good description of the index guiding photonic crystal fibers [5]. The fundamental idea behind this method is to replace the periodic array of holes in silica structure by a properly chosen effective index (Fig. 1a) that can be described in terms of the propagation constant of the lowest-order mode that could propagate in the infinite cladding material [6]. As a result, a step-index fiber that consists of a cladding region with refractive index n_{c1} and a silica core with refractive index n_{c0} (equal to the refractive index of a pure silica n_{si}) is obtained.

The core radius (ρ) of the step-index fiber can be estimated from the formula $\rho = 0.64\Lambda$. Solving scalar wave equation one can obtain an effective refractive index (n_{eff}) and a modal field of the fundamental mode in the step-index fiber.

Using the effective index method and the Kirchoff-Huygens theory of diffraction, the normalized far field intensity distribution ($I(\alpha)$) of a step-index fiber (substitution of PCF)

as a function of the normalized angle (α) can be written as [4]:

$$I(\alpha)=\left\{\frac{U^2W^2}{\left(U^2-\alpha^2\right)\left(W^2+\alpha^2\right)}\left[J_0(\alpha)-\alpha J_1(\alpha)\frac{J_0(U)}{UJ_1(U)}\right]\right\}^2, \tag{1}$$

where: $U=k_0\rho\sqrt{n_{c0}^2-n_{eff}^2}$, $W=k_0\rho\sqrt{n_{eff}^2-n_{c1}^2}$, J_0 and J_1 are the Bessel function of first kind of zero and first order, respectively, α is the normalized angle given by [4]:

$$\alpha=k_0\rho\sin\theta\,, \tag{2}$$

where θ is the angle (Fig. 1b), k_0 is the free space propagation constant and V_{eff} is the effective normalized frequency [7]:

$$V_{eff}=k_0\rho\sqrt{n_{c0}^2-n_{c1}^2}\,. \tag{3}$$

3. Algorithm of characterization of PCFs

The algorithm for characterization of PCFs makes use of the set of curves which can be analytically calculated for the required wavelength according to principles presented in Section 2. Far field intensity pattern of the step-index fiber that is a substitution of LMA-10 PCF is presented in Fig. 2a. Figure 2b shows curves depicting the variation of α_x/α_h ratio and α_h with effective normalized frequency (V_{eff}) for $\lambda=670$ nm. The α_x is the normalized angle of first minimum of the far field intensity distribution and α_h is the normalized angle where the intensity reaches half of its maximum. Figure 2c presents variation of effective normalized frequency (V_{eff}) with the air hole spacing (Λ) for the normalized air hole size (d/Λ) in the 0.2–0.6 range.

The algorithm of characterization is a follows. From the measured far field pattern one can determine the angle of first minimum (θ_x) and the angle of half intensity (θ_h). Next, taking into account that $\sin\theta_x/\sin\theta_h=\alpha_x/\alpha_h$ (Eq. (2)), two values: V_{eff} and α_h can be determined from the curves shown in Fig. 2b. Further, with knowledge of the normalized half intensity angle (α_h), the core radius (ρ) can be calculating from the Eq. (2). The air hole spacing (Λ) can be determined from the formula: $\rho=0.64\Lambda$. From Fig. 2c, the normalized air hole size (d/Λ) can be estimated using Λ and V_{eff}. Finally, knowing the ratio d/Λ and Λ, we can determine the air hole diameter (d). The procedure is presented schematically in Fig. 2b and 2c.

It is shown, that the algorithm presented above makes it possible to determine the geometrical parameters of the PCF: core diameter (ρ), air hole spacing (Λ) and air hole diameter (d) from analysis of the far field intensity distribution.

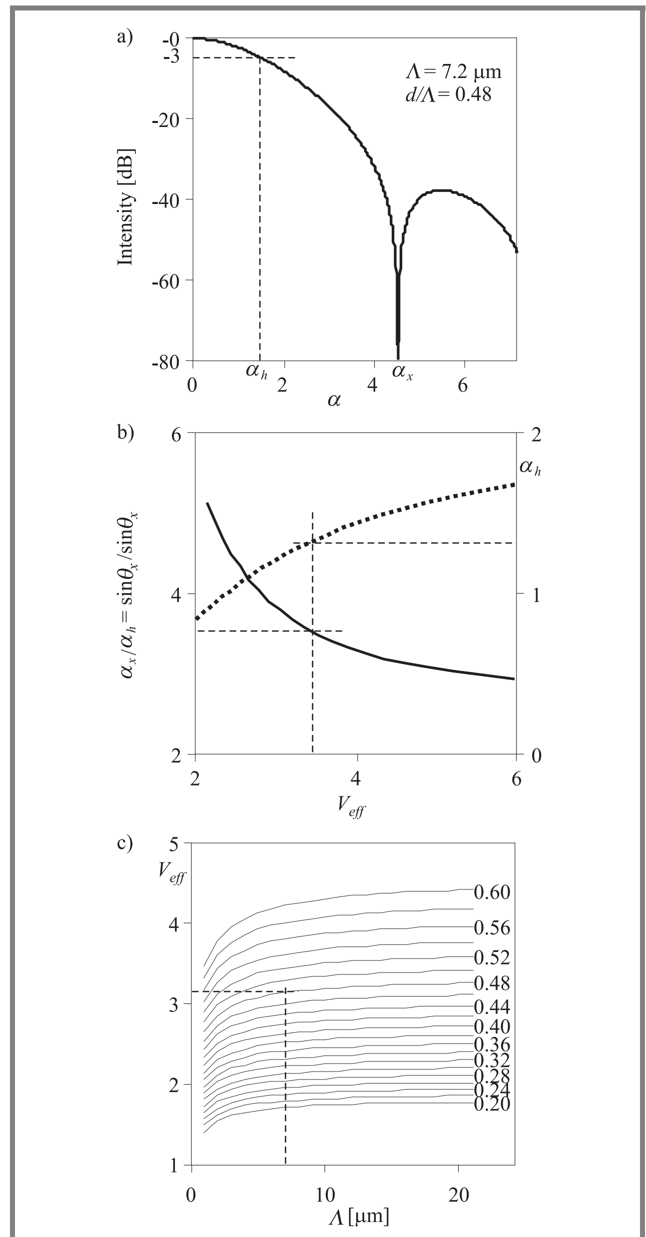


Fig. 2. Far field intensity pattern for the step-index fiber (a); variation of the ratio α_x/α_h (solid) and α_h (dotted) with the effective normalized frequency (V_{eff}) for $\lambda=670$ nm (b); variation of the effective normalized frequency (V_{eff}) with air hole spacing (Λ) for the normalized air hole size (d/Λ) in the 0.2–0.6 range (c). The dashed lines show results calculated for LMA-10 fiber.

4. Evaluation experiment

In order to prove correctness of the presented characterization method, we build a setup shown schematically in Fig. 3a. Light from the laser ($\lambda=670$ nm) is coupled into the 10 meters long PCF: LMA-8 or LMA-10 made

by photonic fibre A/S. Fiber data are shown in Table 1. The far field intensity distribution was observed using BCi4 CMOS C-Cam Technologies Camera (1280 × 1024 pixels, 7 μm pixel size, 12 bit resolution) and 3D beam laser software from MS Macrosystem. All necessary calculations were conducted by the Matlab-based software.

For every fiber we recorded two images for two fiber-camera distances (z). The first image (Fig. 3b) was obtained with normal exposure and used to find the angle of half intensity (θ_h), while the second image (Fig. 4a) was overexposed in order to find the angle of first minimum (θ_x). We show only a quarter of the intensity patterns because of symmetry properties of PCFs.

Figure 3c presents cross sections of intensity distribution for two angles: $\delta = \pi/6$ and $\delta = \pi/3$. The distance to the half intensity position (L_h) can be determined without ambiguity because the far field pattern is cylindrically symmetric for the half intensity values. Thus, knowing the fiber-camera distance (z) and pixel size, one can calculate the half intensity angle (θ_h) from simple geometrical considerations.

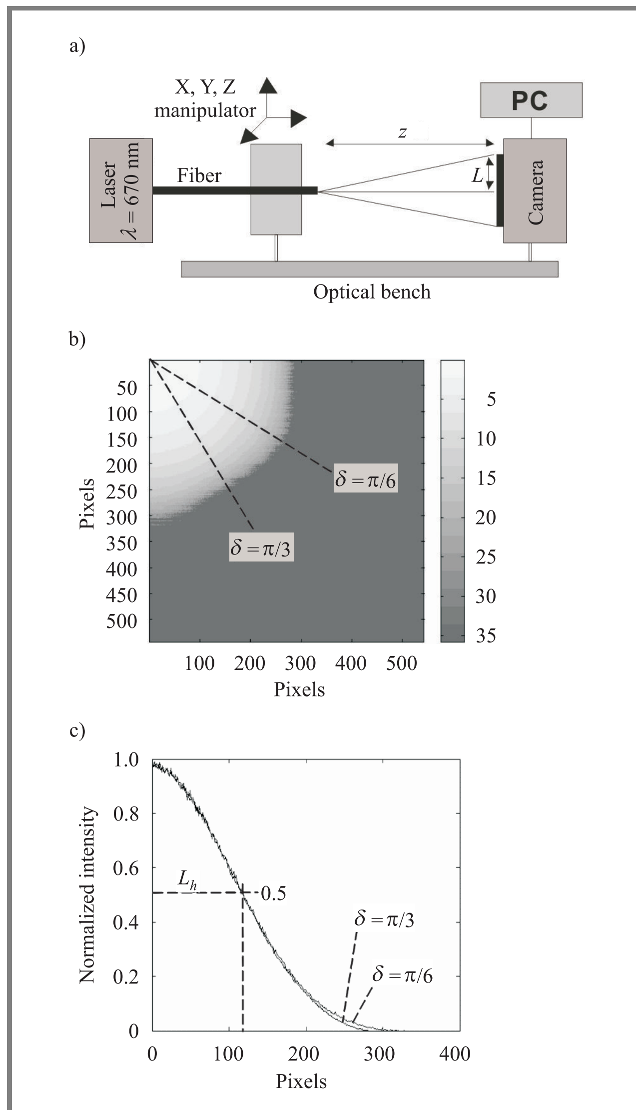


Fig. 3. Far field arrangement (a); normal far field intensity distribution – logarithmic scale (b); cross sections for two angles $\delta = \pi/6$ and $\delta = \pi/3$ (c). LMA-10 fiber, $z = 22$ mm, $\lambda = 670$ nm.

The situation is more complicated for the first minimum angle (θ_x), where the position of the minimum intensity depends on the angle δ. Figure 4b presents cross sections of intensity distribution for two angles: $\delta = \pi/6$ and $\delta = \pi/3$.

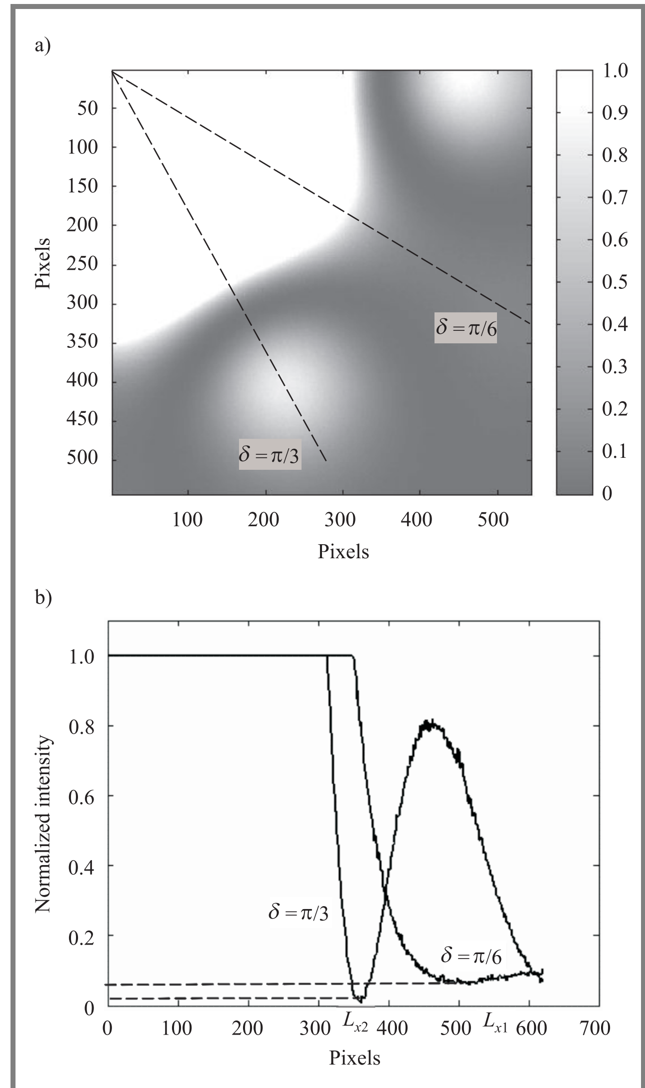


Fig. 4. Overexposed far field intensity distribution (a); cross sections for two angles $\delta = \pi/6$ and $\delta = \pi/3$ (b). LMA-10 fiber, $z = 22$ mm, $\lambda = 670$ nm.

To solve this problem, we assumed that the distance to the first minimum (L_x) equals the mean value of L_{x1} and L_{x2}. Value L_{x1} corresponds to the biggest L_x ($\delta = \pi/6$), while L_{x2} to the smallest one ($\delta = \pi/3$). Now, the first intensity angle (θ_x) can be easily calculated.

Taking advantage of the Fig. 2b and 2c as well as procedure presented in Section 3 one can estimate the parameters

Table 1
Experimental results

Fiber	Manufacturer's data			Measurement data							
	Λ [μm]	d [μm]	V_{eff}	z [μm]	α_x/α_h	α_h	V_{eff}	Λ [μm]	d [μm]	δ_Λ [%]	δ_d [%]
LMA-8	5.6	2.58	2.922	15	3.878	1.141	2.967	5.43	2.55	3.0	1.2
				26	3.866	1.146	2.980	5.56	2.67	0.7	3.5
LMA-10	7.2	3.46	3.112	22	3.762	1.184	3.104	7.13	3.42	1.0	1.2
				28	3.752	1.188	3.118	6.95	3.34	3.5	3.5

of the PCF. The dashed lines in Fig. 2b, 2c, 3b, 3c, 4a, and 4b show results of measurements and calculations for LMA-10 PCF, and $z = 22$ mm, $\lambda = 670$ nm. Table 1 presents the measured parameters for two PCFs: LMA-8 and LMA-10.

It is seen from Table 1 that there is a good agreement between parameters of photonic crystal fibers measured using our method and data provider by fiber manufacturer. Generally, the measured parameters (Λ and d) are lower than the real values. Relative errors of the air hole spacing (δ_Λ) and air hole diameter (δ_d) estimation are below 4%.

5. Conclusions

We present an experimental evaluation of the simple method of characterization of index guiding photonic crystal fibers by means of their far field pattern. In this method, the PCF in question is approximated by a step-index-fiber. The set of calibration curves can be calculated analytically for required wavelength using the far field intensity distribution of the step-index fiber. These curves are used to determine the geometrical parameters of PCF.

We have proposed a measurement procedure that makes use of normally exposed and overexposed images to find the characteristic points of the far field pattern. We also overcame the problem of determination of the first minimum of intensity distribution for photonic crystal fibers with six fold symmetry.

We have obtained a good agreement with manufacturer's data for all fibers tested – the relative errors of measurement for geometrical parameters of PCFs are less than 4%.

The method presented in this paper can be used for easy and quick determination of geometrical parameters of the index-guided PCFs.

Acknowledgements

The authors would like to acknowledge the Polish State Committee for Scientific Research for founding this work.

References

[1] T. A. Birks *et al.*, "Full 2-D photonic bandgaps in silica/air structures", *Electron. Lett.*, vol. 31, pp. 1941–1945, 1995.

[2] J. Laegsgaard *et al.*, "Photonic crystal fibers", in *Proc. Conf. SBMO/IEEE MTT-S 2003*, Foz do Iguaçu, Brazil, 2003, pp. 259–264.

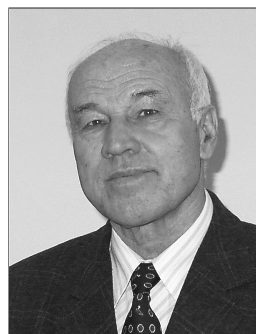
[3] J. C. Knight, "Photonics crystal fibers", *Nature*, vol. 424, pp. 847–850, 2003.

[4] S. K. Varshney and R. K. Sinha, "Characterization of photonic crystal fibers from far field measurements", *J. Microw. Optoelectron.*, vol. 2, pp. 32–42, 2002.

[5] T. A. Birks *et al.*, "Endlessly single mode photonic crystal fibers", *Opt. Lett.*, vol. 22, p. 961, 1997.

[6] J. Broeng, D. Mogilevstev, S. E. Barkou, and A. Bjarklev, "Photonic crystal fibers: a new class of optical waveguides", *Opt. Fib. Technol.*, vol. 5, pp. 305–330, 1999.

[7] A. Bjarklev, J. Broeng, and A. S. Bjarklev, *Photonic Crystal Fibers*. Boston [etc.]: Kluwer, 2003.



Mieczysław Szustakowski was born in Perespa, Poland, in 1933. He received the M.Sc. degree in quantum and solid body electronics, the Ph.D. degree in technical physics and Assistant Professor degree in optoelectronics, all from the Military University of Technology (MUT), Warsaw, Poland, in 1968, 1973 and 1976, respectively. Since 1983 he has been Professor of optoelectronics. In 1968 he joined the Military University of Technology and worked there as a chief of Coupled Field Laboratory, dean of Chemical and Technical Physics Faculty. From 1987 to 1991 he had been working as a Director of Institute of Plasma Physics and Laser Microfusion. Since 1991 he has returned to the Institute of Applied Physics, Military University of Technology. Now he works as Professor. Since 1995 his scientific interests have been in acoustooptics, fiber optics in local area network, fibre optics sensor and integrated optics. At present his main field of interests is acoustooptics signal processing, photonic crystal fibres and fibre optic security systems.

Professor M. Szustakowski is a Member of Polish Chapter of the International Society for Optical Engineering, Polish Committee of Optoelectronics, International Fundation for Research of Optoelectronics and Member of Committee of Electronics and Telecommunication of Polish Science Academy. He is a co-author of more than 60 publication of fiber optic and acoustooptic physics and application, 2 books and 5 patents.
e-mail: mszustakowski@wat.edu.pl
Institute of Optoelectronics
Military University of Technology
Kaliski st 2
00-908 Warsaw, Poland

Norbert Pałka was born in Gniezno, Poland, in 1975. He received the M.Sc. degree in physics – fiber optics technology, the Ph.D. degree in optoelectronics, all from the Military University of Technology (MUT), Warsaw, Poland, in 1999 and 2004, respectively. In 1999 he joined the Military University of Technology as an research worker in Institute of Optoelectronics. Since 1999 his scientific interests have been in fiber optics sensors, especially interferometric and fiber Bragg gratings. At present his main field of interests are photonic crystal fibres and fibre optic security systems.
e-mail: npalka@wat.edu.pl
Institute of Optoelectronics
Military University of Technology
Kaliski st 2
00-908 Warsaw, Poland

Waldemar Grabiec was born in Dębica, Poland, in 1962. He received the M.Sc. degree in telecommunications from the Military University of Technology (MUT), Warsaw, Poland. Since 1998 he has been working as an Assistant and Lecturer at Military University of Technology. His current research areas: optical telecommunications, all optical networks (AON), photonic crystal fibers.
e-mail: waldemar.grabiec@wel.wat.edu.pl
Institute of Telecommunications
Military University of Technology
Kaliski st 2
00-908 Warsaw, Poland

Regular paper

Availability analysis and comparison of different WDM systems

Ivan Rados

Abstract— We begin reasons why high system availability is important. Furthermore, the basic terms are introduced pertaining to the availability with particular review of the parallel structure. Then the availability of different wavelength division multiplexing (WDM) systems is analysed: point to point, chain and ring with 1+1 protection of the wavelength channel, plus the influence of availability of nodes and links on the total system availability. The data on failure intensity and mean time to repair of certain components were taken from various literature sources. We assumed in the analysis a WDM system with 16 wavelengths and 2.5 Gbit/s capacity per wavelength channel. Finally, results of calculation and comparisons of availability of different WDM systems and proposals for improvement of their availability are presented.

Keywords— WDM, wavelength channel, availability, protection.

1. Introduction

Modern industrialised societies are dependent on telecommunication services. Growth of data transmission and growth of internet traffic require transmission systems of huge capacities. Owing to the development of photonic technologies, wavelength division multiplexing (WDM) systems based have been introduced, offering high transmission speeds through a single fiber. In such networks, interruption of service for any reason, either defective equipment or human error, can cause isolation from telecommunication services and huge losses to both customers and network operators. Therefore, protection in such systems is extremely important for network operators [1]. The contracts between operators and their customers are always based on the service level agreements (SLAs), which are very strict regarding network availability and service quality. The possibility to guarantee high availability is the key for the operators in order to keep the existing and attract new customers, and is mainly reflected in the price of their services. Operators can invest a lot of money to equip their network with high quality hardware [2].

However, increase of quality of these elements above the level of current commercial systems is difficult and beyond their control. Far more acceptable way of guaranteeing high availability is the use of various protection strategies. Although there are several protection strategies, in this paper we will focus on 1+1 protection of wavelength channel which is very similar to path protection in synchronous digital hierarchy (SDH) networks [3]. In this paper we will first analyse the availability of link between two terminals of WDM system without protection (point to point, chain) and later the availability of WDM ring which uses 1+1 protection of wavelength channel.

2. On availability in general

The availability of some system A is the probability that the system will be functional in the given time, i.e., the ratio of time in which the system is operational compared to the total time. Unavailability U is a probability complementary to availability. When reporting on system performance, unavailability is frequently expressed as MDT (mean down time) in minutes per year:

$$\mathrm{MDT} = \frac{\mathrm{MTTR}}{\mathrm{MTTR}+\mathrm{MTTF}}, \tag{1}$$

where the MTTF (mean time to failure) is a mean time till the failure occurs and MTTR (mean time to repair) the mean time of repair [4]. In this paper we assume that the MTTF of the wavelength channel is constant regardless of a component age. The MTTF = $1/\lambda$, where λ is the failure rate, usually expressed in FIT (failure in time), where 1 FIT = 1 failure in 10^9 hours.

The main transport entity in the WDM systems is the wavelength channel. This is a one way connection type of a standard bit rate, for example 2.5 Gbit/s, between two nodes. Wavelength channel in the unprotected system (serial structure) in general passes through nodes which can be optical add/drop multiplexers and optical cross connects and optical links. Optical links consist of an optical cable and optical amplifiers which in terms of availability can be considered as one entity because they constitute a serial structure in which failure at any amplifier leads to interruption of communication between two nodes.

Availability of the serial structure is equal to the product of availabilities of individual elements, i.e.,

$$A_s = \prod_{i=1}^{n} A_i\,. \tag{2}$$

Such a structure does not have a possibility to restore communication between two nodes in case of failure of any element, i.e., the connection is functional if and only if all elements of the structure are working.

A significant improvement of availability can be achieved by applying a parallel structure, which means that besides the working path there is a protective one. Of course it is important that these two paths are completely independent, i.e., they do not have common elements [5]. If we mark the availabilities of the working and the protective paths as two

completely independent events (w and p), then the availability of such a structure is a union of two nondisjunctive events, i.e., events which do not exclude each other mutually because complementary variables do not exist

$$A_p = P(w \cup p)\,. \tag{3}$$

For these events to be able to exclude each other mutually it is necessary to turn this union of nondisjunctive events into a union of disjunctive events in which complementary variables appear.

By conducting a simple analysis using De Morgan laws, one finds that the availability is [6] ruled by formula:

$$A = P\left\lfloor w \cup (\overline{w} \cap p)\right\rfloor. \tag{4}$$

By using the law of distribution and bearing in mind the union and intersection

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{5}$$

we get that

$$A_p = P(w \cup \overline{w}) \cap (w \cup p)\,, \tag{6}$$

since $w \cup \overline{w} = 1$ (complete set), we can write:

$$A_p = P\Big[1 \cap (w \cup p)\Big]\,, \tag{7}$$

and according to the identity $1 \cap p = p$, we obtain:

$$A_p = P(w \cup p) \tag{8}$$

this being the availability of parallel structure.

The availability of connection for a given wavelength channel k is [7]

$$A_p = A_k = A_w + A_p - A_w A_p\,. \tag{9}$$

Connection between two nodes in case of 1+1 protection will fail if and only if there is a failure both of the working and protection paths. Of course, the connection fails if a failure occurs on terminal nodes because they are common elements for working and protection paths.

3. Availability analysis of different WDM systems

In this part we will first analyse the availability of WDM point to point system, chain (systems without redundancy) and WDM ring with 1+1 protection of wavelength channel.

We assume that network consists of N nodes and N links which connect those nodes [8]. In order to define the availability of wavelength channel between two nodes we introduce terms which refer to availability of the optical link connecting two nodes:

- a_{OL_i}, availability of ith link of the working path;
- a_{n_j}, availability of $j$the node, which belongs to the working path.

An optical link consists of an optical fibre and an optical amplifier (a_{OA}). In order for the connection to be functional, the optical fibre and amplifier must work properly. The most frequent cause of failure of the optical link is optical fibre cut. Usually all the fibres in the cable are cut, therefore instead of availability of the fibre the availability of the cable (a_{CA}) is used, since failures of fibre and cable are completely mutually dependent

$$a_{OL_i} = a_{CA}\, a_{OA}\,. \tag{10}$$

Nodes in which the wavelength channel is added/dropped are called "termination" nodes. The nodes through which wavelength channels pass from "west" to the "east" side, located between end nodes are called "transit" nodes. Since the wavelength channel passes through different components within these two types of nodes, their availability is different:

- a_{n_jt}, when termination node is working;
- a_{n_jp}, when transit node is working.

If the all nodes of the same type are the same, we have

- $a_{n_jt} = a_{nt}, \forall j$;
- $a_{n_jp} = a_{np}, \forall j$.

3.1. Availability of WDM point to point systems

Installation of a multiple SDH line systems between two nodes is expensive because large number of fibres is occupied, so a need arose for high capacity systems which require only two fibres. Such are WDM systems based on the multiplexing of wavelengths. Introduction of WDM technology begins with a point to point system. With such systems only end nodes are WDM nodes; between them the optical amplifiers are placed at certain distances. At the end nodes the wavelength channel passes through transmitter (TX), multiplexer (MUX), demultiplexer (DMUX), boost optical amplifier (BOA), optical preamplifier (POA) and receiver (RX), so the availability of the node is equal to the product of availability of several elements [9], i.e.,

$$a_{nt} = a_{TX}\, a_{MUX}\, a_{DMUX}\, a_{BOA}\, a_{POA}\, a_{RX}\,. \tag{11}$$

Optical link consists of optical cable and m optical amplifiers, so its availability is naturally equal to the product of the availability of cable and optical amplifiers

$$a_{OL_i} = a_{CA} \prod_{i=1}^{m} a_{OA_i}\,. \tag{12}$$

Total availability of the point to point system is equal to the product of availability of nodes and optical link [10]

$$a_{st}(P) = \prod_{i,j \in P} a_{OL_i}\, a_{n_j} = \prod_{i \in P} a_{OL_i} \prod_{j=1}^{2} a_{n_j}$$

$$= (a_{nt})^2 \left[\prod_{i \in P}^{m} a_{OL_i}\right]. \tag{13}$$

3.2. Availability of WDM chain

When instead of optical amplifiers we install optical add/drop multiplexers which have the possibility to separate wavelength channels, we will get a WDM chain which consists of termination nodes and transit nodes. At the termination nodes the wavelength channel passes through transmitter, multiplexer, demultiplexer, booster optical amplifier, optical preamplifier and receiver so that the availability of termination node is the same as for the point-point system.

At the transit nodes the wavelength channel passes through optical preamplifier, demultiplexer, multiplexer and booster optical amplifier, so the availability of the transit nodes is

$$a_{np} = a_{MUX}\, a_{DMUX}\, a_{BOA}\, a_{POA}\,. \tag{14}$$

A failure of any component within the nodes or the optical link leads to interruption of communication between termination nodes.

If the wavelength channel in the unredundant structure passes m optical links and nodes, availability of the communication between the termination nodes in such a structure is equal to the product of the availability of the optical links and nodes [10], i.e.,

$$\begin{aligned} a_{st}(P) &= \prod_{i,j\in P} a_{OL_i}\, a_{n_j} = \prod_{i\in P} a_{OL_i} \prod_{j\in P} a_{n_j} \\ &= (a_{nt})^2\, (a_{np})^{m-1} \left[\prod_{i\in P}^{m} a_{OL_i}\right]. \end{aligned} \tag{15}$$

It is important to stress here that the optical link consist of optical cable only.

3.3. Availability of WDM ring with 1+1 protection of wavelength channel

In the ring network which uses 1+1 protection of the wavelength channel, the wavelength channel in the source node is duplicated and transmitted simultaneously in both directions of the ring – the working and protection directions [11]. Under normal circumstances, the receiver at the termination node receives two copies of the signal (with different delay) and chooses the better one. In case of failure in the working path, the receiver selects the signal coming from the protection path. This is the so called **single ended** protection because switching is carried out at only one (reception) side. It is important here that working and protection paths do not have common elements, so a single failure does not cause total interruption of communication, and that means failures of the elements in the working and protection paths are completely independent. For each wavelength channel of the working path the appropriate wavelength channel of the protection path is reserved, and therefore we speak about 1+1 protection of the wavelength channel [8]. Since the wavelength channels on the protection path are reserved in advance, we speak about the so called **dedicated** (intended) protection where active cross connects are not necessary. Wavelength channels on working and protection paths can have the same wavelength but do not have to. If different wavelengths are used, wavelength converters which do not have significant influence on availability of WDM ring must be used.

Wavelength channel in the termination nodes passes through transmitter, splitter, multiplexer, booster optical amplifier, optical preamplifier, demultiplexer, switch and receiver, so the availability of the termination nodes is

$$a_{nt} = a_{TX}\, a_{SPL}\, a_{MUX}\, a_{DMUX}\, a_{BOA}\, a_{POA}\, a_{SW}\, a_{RX}\,. \tag{16}$$

Availability of transit nodes is the same as for WDM chain because wavelength channel passes throught the same components. If the wavelength channel of the working path P_0 passes m optical links between end nodes, availability of the working path is equal to the product of the availability of optical links and nodes through which the wavelength channel passes [10]

$$\begin{aligned} a_{st}(P_0) &= \prod_{i,j\in P_0} a_{OL_i}\, a_{n_j} = \prod_{i\in P_0} a_{OL_i} \prod_{j\in P_0} a_{n_j} \\ &= (a_{nt})^2\, (a_{np})^{m-1} \left[\prod_{i\in P_0}^{m} a_{OL_i}\right]. \end{aligned} \tag{17}$$

In case of failure in the working path, wavelength channel passes $N-m$ optical links and $N-m-1$ nodes at the protective path P_1 (Fig. 1).

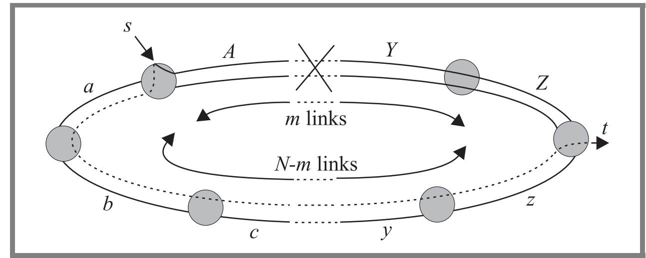


Fig. 1. 1+1 protection of the wavelength channel in case of working path failure.

Availability of the protection path is [10]

$$\begin{aligned} a_{st}(P_1) &= \prod_{i,j\in P_1} a_{OL_i}\, a_{n_j} = \prod_{i\in P_1} a_{OL_i} \prod_{j\in P_1} a_{n_j} \\ &= (a_{nt})^2\, (a_{np})^{N-m-1} \left[\prod_{i\in P_1}^{N-m} a_{OL_i}\right]. \end{aligned} \tag{18}$$

Availability of the wavelength channel between s and t nodes is completely determined by these two paths, so

the availability of wavelength channel in case of 1+1 protection is calculated as availability of parallel structure A_{st} composed of two branches, whose failures are completely independent:

$$A_{st}(a) = a_{st}(P_0) + a_{st}(P_1) - \left[a_{st}(P_0)a_{st}(P_1)\right], \qquad (19)$$

$$\begin{aligned} A_{st}(a) =& (a_{nt})^2(a_{np})^{m-1}\left[\prod_{i\in P_0}^{m} a_{OL_i}\right] \\ &+ (a_{nt})^2(a_{np})^{N-m-1}\left[\prod_{i\in P_1}^{N-m} a_{OL_i}\right] \\ &- \left\{\begin{matrix} (a_{nt})^2(a_{np})^{m-1}\left[\prod_{i\in P_0}^{m} a_{OL_i}\right](a_{nt})^2 \\ \times(a_{np})^{N-m-1}\left[\prod_{i\in P_1}^{N-m} a_{OL_i}\right] \end{matrix}\right\}, \end{aligned} \qquad (20)$$

where a marks availability of optical links and nodes.

Although in the above equation's large brackets we have a product of two equal members $(a_{nt})^2(a_{nt})^2$, only one is present in the formula for availability $(a_{nt})^2$ because the cause of node failure is the same, so we get:

$$A_{st}(a) = (a_{nt})^2 \left\{\begin{matrix} (a_{np})^{m-1}\left[\prod_{i\in P_0}^{m} a_{OL_i}\right] \\ +(a_{np})^{N-m-1}\left[\prod_{i\in P_1}^{N-m} a_{OL_i}\right] \\ -(a_{np})^{N-2}\left[\prod_{i\in P_0,P_1}^{N} a_{OL_i}\right] \end{matrix}\right\}. \qquad (21)$$

If we assume that optical links have equal lengths, their availability is equal, i.e.,

$$a_{OL_i} = a_{OL}, \forall i.$$

In this case the availability between nodes s and t is dictated by the following formula:

$$A_{st}(a) = (a_{nt})^2 \left[\begin{matrix} (a_{np})^{m-1}(a_{OL})^m \\ +(a_{np})^{N-m-1}(a_{OL})^{N-m} \\ -(a_{np})^{N-2}(a_{OL})^N \end{matrix}\right]. \qquad (22)$$

4. Calculation and analysis of results

In order to calculate availability of WDM system it is necessary to know availability of nodes and optical links. For calculation of availability in general, it is necessary to know the failure rates and time to repair of certain elements. These data are taken from various papers previously published [9, 12].

As it can be seen in Tables 1, 2 and 3, decrease of time to repair of both the cable and equipment significantly decreases mean down time and influences availability of the entire system.

We will carry out an analysis of availability of systems, assuming they comprise 8 and 12 nodes, carry 16 wavelengths, 560 and 960 km long optical links and have 80 km spacing between optical amplifiers, i.e., nodes. We will also analyse the influence of reduction of cable and equipment repair time on system availability.

Table 1
Failure rate, unavailability and MDT for optical link (MTTR = 21 and 15 h)

λ [FIT/km]	$U \times 10^{-6}$	MDT [min/year]
114	2.394	1.26
114	1.710	0.90

Table 2
Unavailability and MDT nodes for the system point to point and chain

MTTR = 6 h	$U \times 10^{-5}$	MDT [min/year]
Termination	5.759	30.27
Transit	4.319	22.71
MTTR = 4 h		
Termination	3.840	20.18
Transit	2.880	15.14

Table 3
Unavailability and MDT nodes of the WDM ring with 1+1 protection of wavelength channel

MTTR = 6 h	$U \times 10^{-5}$	MDT [min/year]
Termination	4.049	21.29
Transit	4.319	22.71
MTTR = 4 h		
Termination	2.700	14.19
Transit	2.880	15.14

As can be seen in Tables 4, 5, 6 and 7, the point to point system has the highest unavailability. Although the chain system has WDM nodes in places of optical amplifiers, its unavailability is somewhat smaller in comparison with to point to point system because of very high failure rates of optical amplifiers [9]. Naturally, the WDM ring with protection of wavelength channel has the lowest unavailability (35 times smaller compared to point to point and 31 times compared to chain system) which is understandable because we analyze a parallel structure where failures of certain elements are completely independent.

Table 4
Unavailability and MDT for different systems
(N = 8 nodes)

MTTR = 21 and 6 h	$U \times 10^{-5}$	MDT [min/year]
Point to point	191.82	1008.51
Chain	171.31	900.48
Ring (1+1)	8.16	42.91
MTTR = 15 and 4 h		
Point to point	136.54	717.83
Chain	120.64	634.02
Ring (1+1)	5.43	28.55

Table 5
Unavailability and MDT for different systems
(N = 8 nodes)

MTTR = 21 and 4 h	$U \times 10^{-5}$	MDT [min/year]
Point to point	188.02	988.33
Chain	158.86	834.89
Ring (1+1)	5.45	28.68
MTTR = 15 and 6 h		
Point to point	140.45	738.01
Chain	133.12	699.61
Ring (1+1)	8.13	42.77

Table 6
Unavailability and MDT for different systems
(N = 12 nodes)

MTTR = 21 and 4 h	$U \times 10^{-5}$	MDT [min/year]
Point to point	314.27	1651.76
Chain	265.78	1396.92
Ring (1+1)	556.64	29.26
MTTR = 15 and 6 h		
Point to point	230.75	1212.55
Chain	218.65	1149.01
Ring (1+1)	8.26	43.14

Table 7
Unavailability and MDT for different systems
(N = 12 nodes)

MTTR = 21 and 6 h	$U \times 10^{-5}$	MDT [min/year]
Point to point	318.13	1671.94
Chain	284.05	1492.79
Ring (1+1)	8.28	43.55
MTTR = 15 and 4 h		
Point to point	226.86	1192.37
Chain	200.32	1053.15
Ring (1+1)	5.49	28.87

Decrease of cable repair time from 21 to 15 hours significantly decreases unavailability of point to point system (around 27%) and chain system (around 22%), but difference for WDM ring is insignificant (less than 1%). The reason is that contribution of links to total unavailability is around 95% for a point to point system, around 80% for chain system and only between 0.5 and 1% for a ring with 1+1 protection of wavelength channel.

Decrease of equipment repair time from 6 to 4 hours causes largest decrease of unavailability of the ring structure, around 23%; for chain system the reduction is around 7% and for point to point around 2%. Such situation is due to significantly higher contribution of nodes to total unavailability of ring structure (between 98 and 99%) than of the chain (around 20%) and point to point (around 2%); therefore it is understandable that decrease of equipment repair time has exactly such effect. Of course, in the WDM ring the end nodes have the highest influence on the total unavailability because their failure leads to interruption of communication between two nodes. Termination nodes influence the unavailability of both point to point system and chain system, but the influence is smaller due to significantly higher influence of cable on the total unavailability.

The lowest system unavailability is obtained by decreasing the repair time of both cable (from 21 to 15 hours) and equipment (from 6 to 4 hours) and, expressed as percentage, it ranges from around 30% for point to point, to 29% for chain and 33% for ring. As it can be seen, the difference in the total decrease of unavailability is small because in certain cases the influence of cable is dominant, so almost the same decrease of unavailability is obtained by decrease of repair time of cables and nodes.

5. Conclusion

Results of analysis of presented the in this paper show that the point to point system has the lowest availability. Significant improvement of availability is achieved in WDM ring with 1+1 protection of wavelength channels, because it is actually a parallel structure, where failures of both paths are completely independent. Only after failure on both sides of the ring and at the termination nodes the system becomes unavailable. Point to point and chain systems have several times higher unavailability because they in fact have a serial structure in which failure of any component leads to interruption of communication between two nodes. It is also noticeable that, in general, the availability of different WDM systems decreases with increase of the optical link length and number of nodes, in particular for non-redundant systems. Significant availability improvement of all WDM systems can be achieved by decreasing the repair time both of optical links and equipment, e.g., by better organization of maintenance teams. We need to build ring systems because only this way we can offer the users services of high availability.

References

[1] C. Coltro, "Evolution of transport network architectures", *Alcatel Telecommun. Rev.*, pp. 10–18, 1st Quarter 1997.

[2] G. Willems, P. Arijs, V. W. Parys, and P. Demeester, "Capacity vs. availability trade-offs in mesh-restorable WDM networks", in *Proc. Des. Reliab. Commun. Netw. DRCN*, Budapest, Hungary, 2001.

[3] A. D. Schupke, "Reliability models of WDM self-healing rings", in *Proc. Int. Worksh. Des. Reliab. Commun. Netw. DRCN*, Munich, Germany, 2000.

[4] I. Jurdana and B. Mikac, "An availability analysis of optical cables", in *Proc. WAON'98*, Zagreb, Croatia, 1998, pp. 153–160.

[5] R. Inkret, "Shaping of entirely optical transmission network with wave multiplex". Masters thesis, Zagreb, Faculty of Electrical Engineering and Computing/IT, University in Zagreb, 1998.

[6] R. Inkret, "Availability modelling of multi-service all-optical transmission network". Ph.D. thesis, Zagreb, Faculty of Electrical Engineering and Computing, University of Zagreb, 2004.

[7] D. Arci, G. Maier, A. Pattavina, D. Petecchi, and M. Tornatore, "Availability models for protection techniques in WDM networks", in *Proc. Des. Reliab. Commun. Netw. DRCN*, Banff, Canada, 2003, pp. 158–166.

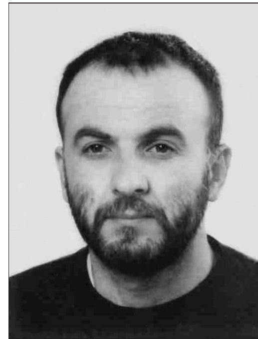
[8] M. R. Wilson, "The quantitative impact of survivable network architecture on service availability", *IEEE Commun. Mag.*, pp. 122–127, May 1998.

[9] R. Inkret and B. Mikac, "Availability analysis of different WDM network protection scenarios", in *Proc. WAON'98*, Zagreb, Croatia, 1998, pp. 121–128.

[10] I. Rados, "Influence of common path on availability of the ring network", *J. Telecommun. Inform. Technol.*, no. 3, pp. 71–75, 2004.

[11] P. Tomsu and C. Schmutzer, *Next Generation Optical Networks*. Upper Saddle River: Prentice Hall, 2002.

[12] W. D. Grover, *Mesh-Based Survivable Networks, Options and Strategies for Optical, MPLS, SONET and ATM Networks*. Upper Saddle River: Prentice Hall, 2004.



Ivan Rados received the B.Sc. degree in electrical engineering from the University of Split, Croatia, in 1983, and the M.Sc. degree from the University of Zagreb, in 2000. In 1985 he joined the PTT (Post and Telecommunication) at Tomislavgrad. Since 1992 he works at the Department of Transmission Systems of the HT Mostar (Croatian Telecommunication). His research interests include: digital transmission systems, optical systems and networks, availability and reliability of telecommunication systems. He has published 10 papers in international conference proceedings.

e-mail: Ivan.Rados@htmobilne.ba
HT d.o.o. Mostar
Kneza Branimira bb
88000 Mostar, Bosnia and Herzegovina