Challenges for 10 nm MOSFET process integration

Mikael Östling, Bengt Gunnar Malm, Martin von Haartman, Julius Hällstedt, 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—strained 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 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 $1.3 \times 10^{-7} \Omega \cdot \text{cm}^2$ are assumed. For the year 2008, corresponding to a physical gate length of 23 nm a value of $8.3 \times 10^{-8} \Omega \cdot \text{cm}^2$ 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 purpose 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$_2$O$_3$, 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 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$_2$O$_3$ with $k$ of 9 and HfO$_2$ with $k$ of 25. A sandwiched structure of Al$_2$O$_3$/HfO$_2$/Al$_2$O$_3$ was investigated and compared to single layer Al$_2$O$_3$ 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$_2$O$_3$ 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$_2$O$_3$/HfO$_2$ gate stack and midgap SiGe gate electrode.](image)

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 $D_{it} = 7 \cdot 10^{12} \text{cm}^{-2} \text{eV}^{-1}$ 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 $1.6 \cdot 10^{12} \text{cm}^{-2} \text{eV}^{-1}$ were obtained in the case of SiGe channels and $3.3 \cdot 10^{11} \text{cm}^{-2} \text{eV}^{-1}$ in the case of a Si-channel device with identical gate stack.
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 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$_2$O$_3$ gate. Gate stacks with either HfO$_2$ or HfAlO$_x$ sandwiched between Al$_2$O$_3$ 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-
Challenges for 10 nm MOSFET process integration

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

Challenges for 10 nm MOSFET process integration

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: julius@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