A FinFET, a novel double-gate device structure, is capable of scaling well into the nanoelectronics regime. High-performance CMOS FinFETs, fully depleted silicon-on-insulator (FDSOI) devices, have been demonstrated down to 15 nm gate length and are relatively simple to fabricate, which can be scaled to gate length below 10 nm. In this paper, some of the key elements of these technologies are described including sub-lithographic patterning technology, raised source/drain for low series resistance, gate work-function engineering for threshold voltage adjustment as well as metal gate technology, channel roughness on carrier mobility, crystal orientation effect, reliability issues, process variation effects, and device scaling limit.

I. INTRODUCTION

According to Brew’s scaling theory [1], the channel doping concentration in bulk MOSFETs should be continuously increased to suppress short-channel effects - to over $10^{18}$ cm$^{-3}$ for gate lengths below 0.1 μm [2][3]. Unfortunately, such a heavy doping concentration degrades device performance due to decreased mobility, increased junction capacitance, and increased junction leakage. Conventional bulk CMOS devices require aggressive gate oxide scaling, which increases the gate oxide leakage current, and ultra-shallow S/D (source/drain) junctions to suppress short-channel effects. Alternative device structures are considered to circumvent these difficulties and continue to scale MOSFET gate lengths. One of them is a double-gate MOSFET controlling the channel and suppressing off-state leakage current more effectively because the gates control the channel from both sides.

In the past, numerous methods have been proposed and demonstrated [4]-[7] to fabricate double-gate devices. However, many suffer from process complexity. A relatively more manufacturable double-gate structure, the FinFET, was thus proposed [8][9]. Next, simplified FinFET structures reducing gate overlap capacitance were proposed as shown in Fig. 1 [10]-[12]. This process sequence is very similar to the conventional SOI device fabrication. In the FinFET, the fin width ($W_{fin}$ in Fig. 1) is the most important process variable because it governs off-state leakage current and alleviates short-channel effects [8]-[10].

For terabit CMOS device scaling, there are a few key issues for the FinFETs to become one of mainstreams in mass production. In this paper, these key issues are discussed and possible solutions are presented with experimental results.
II. Device Fabrication and Performances

Boron-doped \((1 \times 10^{15} \text{ cm}^{-3})\) (100) SOI wafers served as the starting material. N-type body doping \((1 \times 10^{17} \text{ cm}^{-3} \text{ or } 2 \times 10^{17} \text{ cm}^{-3})\) was achieved with \(P\) implantation. The SOI film with 400 nm buried oxide was reduced from 100 nm to 50 nm by thermal oxidation. Oxide was deposited to a thickness of 50 nm on the silicon film to serve as a hard mask to protect the Si-fin during the subsequent gate etching. A fin pattern to become channel was defined by lithography and plasma etching. Next, a 3 nm sacrificial oxide was grown thermally for alleviation of etch damage of sidewall silicon surface. After removal of the sacrificial oxide by \(HF\), a 2.1 nm gate oxide \(T_{ox}\) was grown and gate material was deposited. Conventionally, the gate electrode was patterned by the lithography and etch. Spacers were formed before S/D implantation. \(N^+\) and \(P^+\) S/D implantation were applied with 0° tilted angle. To reduce S/D series resistance, rotational titled implantation is preferred [11]. Rapid thermal annealing (RTA) was used to activate dopants in the S/D region. Metal interconnection process was used to make probing pads. Finally, forming gas annealing was utilized. Detail process information was described elsewhere [10]. Fig. 2 shows 8 nm gate length FinFET (13 nm gate length in the top view) with 5 nm fin width. Typical \(I-V\) characteristics of 15 nm gate length and 10 nm fin width are shown in Fig. 3.

III. Key Issues of Double-Gate FinFets

1. Sub-Lithographic Pattering

Off-state leakage current increases dramatically as the body thickness increases because gate control of the
channel is worsened [13]. In the FinFET, the fin width is corresponding to the body thickness of SOI devices. It is thus important to achieve small and controllable dimensions for the fin width. In addition, because adequate suppression of short-channel effects requires that the fin width ($W_{\text{fin}}$ in Fig. 1) be approximately half of the gate length [9], a sub-lithographic patterning technology is needed. This is clearly impossible to accomplish with standard lithography technologies when $L_g$ is at the limit of lithography. A spacer lithography technology is thus proposed to overcome this limitation [14]-[16]. The essence of the technique is based on the conformal deposition of material during low pressure chemical vapor deposition (LPCVD) process. By depositing a material that has a different etching property than the sacrificial layer and directionally etching the material on the top of the step, the sacrificial layer can then be removed selectively, leaving only the material deposited on the sidewall (spacer). These spacers will serve as a mask for pattern transferring. Thus, minimum-sized features are no longer tied to the Rayleigh’s criteria $R = \frac{k\lambda}{\NA}$, and sub-lithographic patterning is feasible beyond limit of optical or electron beam lithography. A 6.5 nm line width was achieved by this spacer lithography [14]. CD uniformity of this technique is better than that of electron beam lithography because the fin is defined by the thickness of a film deposited by low pressure chemical vapor deposition (LPCVD). LPCVD film thickness uniformity across a wafer is generally better than lithographic CD uniformity in standard silicon processing. To obtain higher drive current, a large channel width is required. This can be achieved by placing many fins in parallel - all straddled by a single gate line [17]. A benefit of spacer lithography technology is that it doubles the pattern density achievable by lithography because two spacers are always formed from single sacrificial structure. As a result, it doubles the drive current for a given layout area. If this spacer lithography is used $n$ times in succession, $2^n$ lines can generated from a single lithographically defined line [16]. In terms of lithographic pattern fidelity, e.g. notching phenomena, spacer lithography is more attractive than electron beam or advanced optical lithography because the sidewall spacer was deposited conformally, thus the line width along the sacrificial layer is the same as shown in Fig. 4 (a). Line edge roughness effects are substantially reduced.

2. Series Resistance

As the fin width is reduced, the parasitic series resistance in the source and drain increases. If the gate is misaligned to the source and drain pads, the output characteristics can be changed significantly [10]. A selective Ge deposition process was proposed [10][18] to alleviate the effect of misalignment and to reduce the series resistance. The use of Ge is attractive because it is a low temperature process, which is important for future integration with high-K gate dielectrics and metal gates. Furthermore, Ge provides an in-situ cleaning effect because GeH$_4$ can remove native oxide [19]. On-current was enhanced to 28% with selective Ge raised S/D process [10]. For further reduction of series resistance, metal germanide such as nickel-germanide can be attractive.

3. Adjustment of Threshold Voltage and Metal Gate

A major challenge in the realization of a CMOS FinFET technology is the achievement of appropriate threshold voltages for both NMOS and PMOS devices. For an ideal double-gate structure without channel
doping, this requires that a technique be available to adjust the work function of the gate electrode [20]. In this work, molybdenum (Mo) with nitrogen implantation is used to study gate work function engineering for threshold voltage ($V_T$) control. A benefit of metal gate is to eliminate the gate depletion effect and dopant penetration through the dielectric. Ordinarily, two metals with appropriate work functions need to be used on a single Si substrate in order to obtain low and symmetric $V_T$ devices in CMOS. For minimal process complexity, a method for tuning the gate work function over the required range would be highly desirable.

![Graph showing variation of Mo work function with thermal annealing.](image)

**Fig. 5.** Variation of Mo work function with thermal annealing. All anneals were 15 min long except for the 900 °C anneal (15 sec) [38].

There have been several reports on the use of ion implantation to change the work function of thin metal films [21]-[26]. Nitrogen implantation has been shown to be an effective way to nitridize the Mo film and controllably lower its work function [22]. As shown in Fig. 5, the Mo work function is dependent on nitrogen implant dose, energy, and annealing temperature. The range of work functions obtained makes this approach attractive for FDSOI-CMOS applications.

Fig. 6 shows cross-sectional TEM photographs of Mo-gate FinFET. Nitrogen implantation was applied after Mo sputtering. Next, poly-silicon was deposited as a cap layer. More process description was reported elsewhere [26]. Threshold voltage of an unimplanted Mo-gated PMOS FinFET shows –0.2V (defined at 100nA/μm $I_{ds}$). Nitrogen-implanted Mo-gated devices exhibit a $V_T$ which is shifted in the negative direction, with $|V_T|$ increasing with implant dosage as shown in Fig. 7 (a). It should be noted that no special post-nitrogen-implant annealing step was performed in this study, and that larger $V_T$ shifts can be attained by adding such a step, as well as by using higher implant doses [22][26]. These would be needed in order to achieve the low gate work function values required for NMOS devices. Subthreshold swing is degraded by nitrogen implantation as nitrogen dose increases as shown in Fig 7 (a). This side effect and hysteresis are alleviated by deuterium annealing as shown in Fig. 7 (b).

![Cross-sectional TEM photographs of Mo-gate FinFET.](image)

**Fig. 6.** Cross-sectional TEM photographs of Mo-gate FinFET. Poly-Si caps over Mo-gate for subsequent processes.
Fig. 7. $I_d-V_g$ characteristics of Mo-gate FinFET (a) and subthreshold swing as well as hysteresis after deuterium anneal (b). Device dimensions are $L_g$=80nm, $W_{fin}$=10nm, and $T_{ox}$=2.1nm. Deuterium annealing (30 min at 400 °C) is effective to restore the subthreshold swing and reduce the hysteresis after nitrogen implantation.

4. Surface Roughness

Another challenge in fabrication of the FinFET structure is control of surface roughness along the fin sidewall, which may be caused by line edge roughness in the resist (if conventional lithography is used to pattern the fins) and plasma etch damage. The sidewall roughness can cause mobility degradation. Thus, pure hydrogen annealing after the silicon fin formation was proposed to smoothen the sidewall surface. Reduction of surface roughness by pure hydrogen annealing was verified by SEM analysis, low frequency noise measurement, mobility measurement, and gate oxide reliability test. With SEM analysis as shown in Fig. 8, surface roughness was significantly improved due to the high surface Si migration rate in the pure hydrogen annealing process. To observe the clear improvement of sidewall roughness, polycrystalline-silicon pattern was used instead of crystalline silicon structure. Physical sidewall roughness was improved after 5 min annealing at 900 °C in 100% hydrogen as shown in Fig. 8.

Previous experimental results [27] showed that electron mobility was degraded due to surface roughness while hole mobility still followed the universal curve. This is likely because the inversion charge centroid for electrons is closer to the silicon-oxide interface than that for holes [28] since the heavy electron effective mass results in a strong quantum confinement effect; this is confirmed by a coupled Schrödinger-Poisson solver [29]. If the interface in a small region of the channel is away from the averaged position by a quantity $\Delta$, which was the root-mean-square (rms) value of rough surface, the average potential felt by the centroid of the carriers will be changed by $E_{eff}\Delta$. It is well known that a potential change causes additional scattering, thus it can be treated perturbatively during carrier transport. Because of mobility degradation by surface roughness scattering is represented by $\mu_{SR} \propto (E_{eff}\Delta)^2$, electron mobility is more significantly degraded at high effective field in NMOS than in PMOS with the same $\Delta$. This behavior is consistent with the previous report in [27] and the measured mobility versus effective field in the FinFETs (Fig. 9). Therefore, relatively low NMOS current and high PMOS current as shown in Fig. 3 can be partly explained.

To investigate the quality of the Si-SiO$_2$ interface formed by the etching, low-frequency noise characteristics were investigated [26][30]. It was found that the noise characteristic is flat at low current levels (weak inversion regime), and is attributed to carrier-number fluctuations due to traps at the Si-SiO$_2$ interface. At high current levels (strong inversion regime) the noise increases quadratically, and is attributed to mobility fluctuations. The noise level and dependence on the drain current are comparable and similar for the poly-Si-gated hydrogen-annealed FinFETs as for bulk-Si MOSFETs [31].
hydrogen [32]. Therefore, the hydrogen annealing process is effective for achieving high-surface-quality fins.

5. Crystal Orientation

Relatively high PMOS and low NMOS drive currents for FinFETs as compared to bulk-Si MOSFETs have been observed as shown in Fig. 3. Except for surface roughness effects to degrade electron mobility, there is another factor to cause low electron mobility and high hole mobility, which comes from an anisotropy of carrier mobilities in silicon [33]. It is known that hole mobility is enhanced, while electron mobility is degraded, for a (110) surface as compared to a (100) surface in Si. The field-effect mobilities for holes and electrons were extracted from FinFET $C_g$-$V_g$ and $I_{ds}$-$V_{gs}$ measurements. The hole mobility surpasses the universal mobility curve for (100) bulk-Si PMOSFETs while the electron mobility is lower than the universal curve for (100) bulk-Si NMOSFETs. These results are expected, and are consistent with the previous report of FinFET $g_m$ dependence on crystal orientation [11]. These anisotropy effects become even more important in trigate [34] or gate-all-around devices [35]. To simultaneously achieve high NMOS and PMOS drive currents, a (100) sidewall surface for NMOS and (110) sidewall surface for PMOS is desirable. One way to implement these two different crystal orientations is to align silicon fins to be perpendicular or parallel to the flat zone of a (100) wafer for PMOS and at a 45° rotation for NMOS as shown in Fig. 11. Such a scheme, which may incur a small area penalty, will depend on lithographic capabilities.
6. Reliability

It is timely to investigate reliability of FD-SOI CMOS FinFETs. For suppression of short-channel effects, the fin width is as narrow as possible. However, it can be a big concern that the extremely narrow fin width would be desirable in the device reliability point of view. Thus, DC hot-carrier tests were performed to evaluate device reliability for various device dimensions. Gate oxide charge-to-breakdown ($Q_{BD}$) distributions are also studied. The gate dielectric in all devices was 2.1 nm thermally grown SiO$_2$. There is significant boron penetration through the gate oxide into the channel/fin resulted in a high threshold voltage ($V_T$) in $P^+$ Si$_{0.5}$Ge$_{0.4}$ gate and gate-underlapped structure for the n-channel FinFETs. The $p$-channel FinFETs, however have a conventional gate-overlapped structure. First, the worst-case bias condition for hot-carrier stress was determined for $V_d=1.8V$. The $V_T$ shift was found to be larger when the gate voltage was equal to the drain voltage ($V_d=V_g$), compared to $V_g=V_d/2$. Degradation in device performance parameters under $V_g=V_d=1.8V$ stress was then tracked over time, for FinFETs of various fin widths ($W_{fin}$). $Q_{BD}$ distributions were also obtained for FinFETs of various sizes.

The magnitude of the $V_T$ shift increases with increasing fin width as shown in Fig. 12 (a). To explain this dependence on $W_{fin}$, the electron-hole pair generation rate ($G$) due to impact ionization – and electron temperature ($T_e$) profile in an $n$-channel double-gate MOSFET were studied using the device simulator MEDICI [36]. The gate is assumed to be perfectly aligned (zero overlap) to the abrupt source/drain (S/D) junctions, and its doping concentration is 2x10$^{20}$cm$^{-3}$. It was found that both $T_e$ and $G$ increase with increasing $W_{fin}$, which is consistent with the reliability measurement data [32]. Fig. 12 (b) shows that the potential profile becomes increasingly "one-dimensional" as the body thickness decreases. In the thicker body case, hot electrons are more strongly driven towards the by the 2D curvature of the potential. The paths of generated hot carriers are indicated schematically by the bold arrows in Fig. 12 (b). Thus, hot-carrier immunity is improved by thinning the body of a double-gate MOSFET, which also improves short-channel effects.

![Image](67x247 to 153x286)

**Fig. 12.** (a) relative shifts in $V_T$ and $I_{dss}$ for NMOS FinFETs (gate-underlapped structure) vs. stress time, for different fin widths. Better hot-carrier immunity is achieved with a narrower fin. (b) simulated 2-D potential profile ($V_g=V_d=1.2V$) in a double-gate MOSFET with $T_S=5$ nm and $T_G=42$ nm ($L_g = 79$). The gate is perfectly aligned (zero overlap) to the abrupt S/D junctions ($N_{SOI}=2x10^{20}$cm$^{-3}$). The potential lines are more straight than for the thick-body device at the drain side. The highest potential gradient is located well under the gate, in contrast with the case of the thinner-body structure.

![Image](169x244 to 268x310)

**Fig. 13.** Hot-carrier DC lifetime for CMOS FinFETs ($P^+$ poly-$Si_{0.5}Ge_{0.4}$ gate, no hydrogen anneal after fin etch).
Hot-carrier DC lifetime is plotted in Fig. 13. The criteria for failure used for these plots are 10% change of $V_T$, 10% change of $g_m$ (transconductance), and 10% change of $I_{dsat}$. The lifetime is improved slightly for devices that received a post-fin-etch anneal in $H_2$ [32]. These results indicate that FinFETs should be able to meet the 10-year lifetime requirement for normal operating conditions, if $W_{fin}$ is sufficiently small.

7. The Effects of Process Variations

In such thin-body devices (FinFETs), process variations of dimensions and thicknesses can be a big concern for mass production. A statistical calculation of the $\sigma$ values of device electric parameter variations are studied by simulation. It was assumed that the contributions to device electrical parameter variations by small change of physical parameters are independent. Nominal device parameters for the device simulation are $L_{gate}=20$ nm, $W_{fin}=5$ nm, $T_{ox}=1$ nm, $T_{body}=4$ nm, $T_{body}=5$ nm, and $T_{body}=10$ nm, and $N_{body}=10^{15}$ cm$^{-3}$, $I_{off}=1$ nA/µm, and $N_{peak of S/D}=2x10^{20}$ cm$^{-3}$ with a 2 nm/dec Gaussian lateral gradient. Assuming that the process parameter variations are $3\sigma_{T_{Si}}=1$ nm, $3\sigma_{L_{gate}}=2$ nm, and $3\sigma_{I_{off}}=1$ A, threshold voltage variation is in the range of $40$mV while drive current variation is within 8%~10%. With these same process parameter variations, the subthreshold swing is $68\pm6$mV for NMOS and $70\pm6$mV for PMOS [38]. Therefore, it is concluded that the device parameter fluctuations can be kept within reason.

8. Scaling Limits

It is clear that an ultimate limit of MOSFET gate length scaling will eventually be reached. As the MOSFET channel length is reduced to 50 nm and below, the suppression of off-state leakage current becomes an increasingly difficult technological challenge—one that will ultimately limit the scalability of the conventional MOSFET structure. By using the off-state leakage current as criteria for the double-gate MOSFET scaling, its scaling limit can be evaluated. Scaling of the body thickness in the double-gate eliminates leakage paths that are not well controlled by the gate—those that are physically far from the gate electrode. As a result, short-channel effects are minimized at small gate lengths, thus reduced leakage. The scaling limit of double-gate MOSFETs is a strong function of the body thickness. Previous work has suggested that the minimum acceptable body thickness for a thin-body MOSFET will be ~ 5 nm [20]. This is because the series resistance in the source and drain regions introduced by such a thin film may become unacceptable even with the incorporation of a raised S/D technology. In addition, quantum confinement in the thin-body may cause an intolerable threshold voltage when $T_{Si}$ is thinner than 5 nm [37]. Thus, a scaling limit for double-gate MOSFETs can be obtained under the assumption that the minimum body thickness is 5 nm. The equivalent oxide thickness will likely be limited to ~1nm due to gate leakage current, even with the advent of alternative gate dielectric materials. At these limits, Figure 14 shows that off-state leakage current targets [13] can be met at gate lengths down to ~10 nm conservatively. This scaling limit is also dependent upon the choice of threshold voltage and the lateral S/D doping gradient [13]. In short, the value of the scaling limit based upon off-state leakage criteria becomes smaller as the body thickness and equivalent oxide thickness are decreased, but is raised by an abrupt source/drain doping profile and a low threshold voltage [38].
IV. CONCLUSIONS

A FinFET, a novel double-gate device structure is demonstrated for sub-10nm device scaling. Key issues including sub-lithographic patterning technology, raised S/D for low series resistance, gate work-function engineering for threshold voltage adjustment as well as metal gate technology, channel roughness on carrier mobility, crystal orientation effect, reliability issues, process variation effects, and device scaling limit are discussed. While a number of challenges remain to be overcome, it appears to us that the continued evolution of CMOS integrated technology will not be impeded by basic limitations of the underlying technology. This implies that device scaling will continue for yet another 15-20 years before the ultimate device limits for CMOS are reached [39]. The double-gate FinFETs will lead the further device scaling.

ACKNOWLEDGEMENTS

Author would like to thank Professor Chenming Hu, Jeffrey Bokor, and Tsu-Jae King at the University of California-Berkeley for their support to perform this research. Author also thanks to colleagues, Leland Chang, Daewon Ha, Pushkar Ranade, and Shiyiing Xiong for their collaboration. The FinFET devices were fabricated at the Microlab of the University of California-Berkeley.

REFERENCES


Yang-Kyu Choi received B.S., M.S. degree from the Seoul National University, Korea, in 1989 and 1991, M.S. and Ph. D. degrees from the University of California-Berkeley in 1999, 2001, respectively. He was a post-doctoral researcher working on nanobiotechnology interdisciplinarily at the University of California-Berkeley from 2001 to 2003. He is currently an assistant professor at the department of electrical engineering in the Korea Advanced Institute of Science and Technology (KAIST). From January 1991 through July 1997, he had worked for Hynix Co., Ltd., Kyungki-Do, Korea, where he developed 4M, 16M, 64M, 256M DRAM as a process integration engineer. His research interests are novel MOSFET structure such as UTBFET and FinFET, nanofabrication technology for bio-electronics as well as NEMS, and an investigation of quantum phenomena for nanoscale CMOS. He received the Sakrison Award for the best dissertation in the department of Electrical Engineering and Computer Sciences at Berkeley in 2002. He also received Samsung Humantech Thesis Prize in 2000 and one of the best student paper at the International Semiconductor Device Research Symposium in 2001 and the International SOI conference. His biographic profile was published in 57th Marquis Who’s Who in America. He has authored or co-authored over 50 papers, and holds 6 U.S. patents as well as 99 Korea patents.