Selected Topics on HfO₂ Gate Dielectrics for future ULSI CMOS Devices

M.F.Li¹ ² , H.Y.Yu¹ , Y.T.Hou¹ ² , J.F.Kang¹ ³ , X.P.Wang¹ , C.Shen¹ , C.Ren¹ , Y.C.Yeo¹ , C.X.Zhu¹ , D.S.H.Chan¹
Albert Chin⁴ , D.L.Kwong⁵
¹ Silicon Nano Device Lab, ECE Dept, National University of Singapore, Singapoer 119260
Email : elelimf@nus.edu.sg
² Institute of Microelectronics, Singapore 117685
³ Institute of Microelectronics, Peking University, Beijing 100871, P.R.China
⁴ Dept Electronics Eng, National Chiao Tung University, Hsinchu, Taiwan, R.O.C
⁵ Microelectronics Research center, Dept ECE , The University of Texas, Austin, TX 78712, USA

Abstract: Based on our recent investigation on HfO₂ high-k gate dielectrics, we review the Hf-based gate dielectric in the future ULSI CMOS devices in the following aspects: How long HfO₂ can be used satisfactorily, assessed from the gate tunneling and scalability; how thin EOT can be grown and scalability; how thin EOT can be grown; how high operating voltage can be employed for 10 years reliable operation, assessed by charge trapping and threshold voltage shift.

I. Introduction

HfO₂ has emerged as one of the most promising high-k gate dielectrics and will be used in next generation CMOS transistors in the industry mass production in 2007. HfO₂ has the following advantages as gate dielectric: high dielectric constant (~25), large electron (1.9 eV) and hole (2.3 eV) band offsets to Si, and therefore effectively reduce the gate leakage current; excellent high frequency response; thermodynamic stability in contact with Si. However, HfO₂ also has some serious problems such as low crystallization temperature, low mobility degradation, threshold voltage instability due to charge trapping and de-trapping.

II. How long HfO₂ can be used satisfactorily? assessed from the gate tunneling current and scalability.

Although superior electrical characteristics have been demonstrated for HfO₂ and HfO₂ gate stacks, the simulations are fitted tunneling mass \( m_{HfO_2} \) is 0.18 \( m_0 \). When the uncertainty of HRTEM is 1 Å, the result of \( m_{HfO_2} \) is 0.18 \( m_0 \), and a value of 0.18 \( m_0 \) fits the experiments well in both the inversion and accumulation polarities. Good agreements between simulations and measurements are obtained.

In our model, electron and hole voltages in both the substrate and gate electrode are accurately treated, including the band mixing effect in hole quantization. The tunneling current from the discrete 2-D subbands in inversion or accumulation layers in MOS devices is obtained by a modified WKB approximation, which takes into account the wave function reflections at the dielectric-Si barrier interface. In our model, the interfacial layer (IL) between high-k and Si substrate is readily included. In addition, both tunneling through channel and source/drain extension region (SDE) overlapped with the gate, as shown in Fig.1, can be calculated.

In our simulation, the valence band offsets \( \Delta E_v \) between (HfO₂) \((Al₂O₃)_{1-x} Si \) and Si are determined from the valence band XPS and the energy gap \( E_g \) is determined by the O 1s energy loss spectra. For high-k stacks, the simulations are compared with electrical data from ultra-thin CVD and PVD HfO₂ devices with and without a N₂H₃-based interface layer, and with either poly-Si or TaN as gate electrodes. The flat-band voltages for all samples were obtained from C-V measurements.

![Fig.1: The gate current of a MOSFET is composed of tunneling currents from the channel (CH) and source/drain extension (SDE) region overlapped with the gate.](image)

**Tunneling in HfO₂ and Al₂O₃ gate stacks**

Fig.2 shows the comparison between simulation and experimental data of PVD HfO₂ on p-Si. The physical thickness of HfO₂ (38 Å) and IL (6 Å) used in the simulation were determined by XTEM. The only fitting parameter in our simulation is the electron tunneling effective mass in HfO₂, \( m_{HfO_2} \), and a value of 0.18 \( m_0 \) fits the experiments well in both the inversion and accumulation polarities. Good agreements between simulations and measurements are obtained.

![Fig.2: Simulated gate current of a n+ poly/HfO₂/SiO₂/p-Si device. The measured data are from [3], the physical thickness are HfO₂(38 Å)/IL(6 Å) from HRTEM and IL is likely SiO₂ from XPS. The fitted tunneling mass \( m_{HfO_2} \) is 0.18 \( m_0 \). When the uncertainty of HRTEM is 1 Å, the result of \( m_{HfO_2} \) is ±0.02 \( m_0 \). The dashed lines are simulations with \( m_{HfO_2} = 0.20 \) and 0.16 \( m_0 \).](image)

Fig.3 shows the comparison between simulation and experimental data of ALD Al₂O₃/SiO₂ stacks. A single value of \( m_{AlO_2} = 0.28 \( m_0 \) is able to give the best fitting to the four experimental curves with different equivalent oxide thickness (EOT) and bias polarities.

**Tunneling in (HfO₂)ₓ(Al₂O₃)_{1-x}**

The electron effective mass and the dielectric constant values of (HfO₂)ₓ(Al₂O₃)_{1-x} are linearly interpolated from those
It is shown that SiO2 or optimized SiON can meet the ITRS tunneling current in SDE, ISDE, is the main contribution of discussion of scalability of high-HfAlO with [Al] = 30% as optimized HfAlO in the subsequent channel tunneling (CH) and gate-s/d (SDE) tunneling for some Fig.4 shows the tunneling current density of gate – channel tunneling (CH) and gate-s/d (SDE) tunneling for some typical gate dielectric versus EOT. As is well known , tunneling current in SDE, ISDE, is the main contribution of transistor off-sate leakage Ioff for devices with ultra thin gate dielectric. From Fig.4, our calculations show that tunneling current density in SDE is comparable with that from the channel area and the criterions for Ioff and gate leakage Ig are in same values from ITRS25, the scalability of a dielectric can be obtained by analyzing the gate leakage in channel.

Fig.3: Simulated electron tunneling currents of n-MOS with Al2O3 gate dielectric. The data are from [23]. The tunneling effective mass is found to be 0.28 m0 from overall fitting of all the data. The thickness values from best fitting to the measured data match well with those in [23] from C-V method (in parenthesis).

fraction of Al incorporation into HfO2 can raise the crystallization temperature to ~ 900°C. Therefore, we use HfAlO with [Al] = 30% as optimized HfAlO in the subsequent discussion of scalability of high-κ gate stack.

Fig.4 shows the tunneling current density of gate – channel tunneling (CH) and gate-s/d (SDE) tunneling for some typical gate dielectric versus EOT. As is well known , tunneling current in SDE, ISDE, is the main contribution of transistor off-sate leakage Ioff for devices with ultra thin gate dielectric. From Fig.4, our calculations show that tunneling current density in SDE is comparable with that from the channel. Since the SDE area is smaller than or comparable with the channel area and the criterions for Ioff and gate leakage Ig are in same values from ITRS25, the scalability of a dielectric can be obtained by analyzing the gate leakage in channel.

Scalability: How long HfO2 can be used ?

In the following, the gate leakage is calculated for each CMOS technology generation according to the ITRS 200125. For each generation, the gate leakage is estimated by the gate current value at Vg=VDD with corresponding EOT and operating voltage VDD values provided by ITRSTs25. In Fig.5, we show the results for high performance CMOS technology. It is shown that SiO2 or optimized SiON can meet the ITRS target from leakage current viewpoint, indicating that the scaling of SiO2 in high performance CMOS is probably not limited by its gate leakage and such conventional gate dielectric will be continually explored for high performance CMOS aggressively up to its physical limitation as a dielectric material. However, the situation is different for low power applications, where the gate leakage criterion is more stringent. Fig.6 is the results for low standby power (LSTP) application. Here, an average value of the proposed maximum and minimum EOT from ITRS25 is used for each generation. Al2O3 mole fraction is 30% for HfAO and Si3N4 mole fraction 40% for optimized SiON.

Fig.5: The calculated gate leakage of high performance CMOS. The VDD, EOT and channel length are taken from ITRS200125 for each generation. The calculated high (low) gate leakage for each generation corresponds to the minimum (maximum) EOT proposed in ITRS 2001. Si3N4, mole fraction 40% for optimized SiON.

Fig.6: The calculated gate leakage for low standby power (LSTP) application. Here, an average value of the proposed maximum and minimum EOT from ITRS25 is used for each generation. Al2O3 mole fraction is 30% for HfAO and Si3N4 mole fraction 40% for optimized SiON.

Fig.7: The calculated gate leakage for low standby power applications of HfAO dielectric with different IL layers to improve the interface quality. Physical 3 Å IL of SiO2, optimized SiON, HfSiO4 were presented. A minimum 3 Å SiO2 and SiON ILs are also shown to demonstrate the limit of SiO2-based dielectrics as an IL layer limited by its gate leakage and such conventional gate dielectric will be continually explored for high performance CMOS aggressively up to its physical limitation as a dielectric material. However, the situation is different for low power applications, where the gate leakage criterion is more stringent. Fig.6 is the results for low standby power (LSTP) application. It is shown that SiO2 will approach the scaling limit in 2003 due to its high gate leakage, medium-κ dielectrics (SiON or Al2O3) can only extend the CMOS technology scaling by 1-2 generations. HfO2 and HfAO are demonstrated to be viable and potential candidates for long-term solutions as alternative high-κ gate dielectrics.

In practical applications, a molecular IL layer is generally
needed in order to improve the interface quality and channel mobility. We thus studied several interface materials including SiO₂, SiON and silicates, and their impacts on gate leakage of HfAlO gate stack. Considering the thinnest IL physical thickness of 3 Å (a single molecular layer), Fig. 7 shows that EOT of HfAlO stack with SiO₂ IL cannot be scaled down to below 1 nm. However, using SiON or silicates as IL, HfAlO stack is scalable to 25 nm node LSTP applications by year 2016.

III. How thin EOT can be grown technologically for HfO₂+IL gate dielectric:

There is no difficulty to grow 1-2 nm physical thickness HfO₂ by Atomic Layer Deposition (ALD) with EOT=0.2-0.4 nm. The ultimate limit of equivalent oxide thickness (EOT) of HfO₂ gate dielectric is mainly decided by the thickness and dielectric constant of IL. During high-k dielectric growth, particularly during post-deposition annealing (PDA) and s/d dopant activation annealing at high temperature, oxygen atoms diffuse through the gate electrode and gate dielectric, interact with Si substrate, an interfacial layer of SiOX with lower dielectric constant will be formed. On the other hand, this IL as a buffer can improve mobility degradation due to improvement of surface roughness and reduction of interface trap density, also due to remote scattering (phonon or Columbic) from the high-K dielectric. Therefore the IL is necessary for high performance CMOS transistors and there is no consensus how thin the IL should be. However from the technical point of view, it is not a easy task to control very thin IL and therefore HfO₂+IL with very thin EOT. So far the thinnest reported EOT for HfO₂+IL is 0.6 nm²⁶,²⁷.

There are different methods to reduce oxygen diffusion and suppress the IL growth. One method is to add the third element such as N²⁰ or A¹⁰,²⁸ into HfO₂ to suppress oxygen diffusion in the dielectric. However this method has disadvantage of gate leakage increment. Another method is surface nitridation before growth of HfO₂. However, this method has the disadvantages of higher interface trap density and more serious mobility degradation²¹. We have developed a third method using highly thermal stable HfN metal gate electrode as an oxygen barrier. Using HfN/HfO₂ gate stack without surface nitridation, we will be able to obtain 0.8-0.6 nm gate dielectric after 1000°C annealing, and very low gate leakage current and comparatively high channel mobility.

Experimental:

For MOSCAPs fabrication, after the active area definition with 4000Å field oxide, and a standard DHF-last RCA pre-gate clean process, HfO₂ films were deposited at 400°C using a MCVD cluster tool. Some samples received in-situ surface nitridation (SN) treatment in NH₃ at 700°C prior to CVD HfO₂ deposition. All samples were then subjected to PDA at 700°C in N₂ ambient, followed by PVD deposition of HfN(~50nm)/TaN (as capping layer on HfN, ~100nm). Devices were then rapid thermal annealed (RTA) in N₂ at 900°C-1000°C for 20sec for thermal stability evaluation. For MOSFETs' fabrication, s/d implantsations of phosphorus for n-MOS and BF₂ for p-MOS with a dose of 5x10¹⁵ cm⁻² were performed followed by RTA activation in N₂ at 950°C for 30s. EOT and flat band voltage (Vₐ₈) were simulated by taking into account the quantum mechanical correction.

Results and discussion

MOSCAPs:

Fig. 8 compares measured and simulated CV curves of HfN/HfO₂ devices without SN after various RTA anneal, showing good agreement. The EOT of HfN/HfO₂ is as low as 8.2Å after forming gas anneal (FGA), and slightly increases to 8.8Å/9.1Å after 900°C/1000°C post-metal annealing (PMA).
Insignificant variation of the gate leakage current is observed in these devices after RTA (Fig 9). From the XTEM and CV data, it appears that the IL is not the pure SiO$_2$, and it has a K value of 7–8. The superior thermal stability of HfN/HfO$_2$ gate stack is further demonstrated in Fig.10, where several devices with different EOT are plotted as a function of RTA temperatures. Negligible EOT variations in all HfN/HfO$_2$ devices (without SN) are demonstrated up to 1000°C RTA. For comparison, the EOT of TaN/HfO$_2$ gate stack shows significant increases after 800°C, 900°C and 1000°C RTA, even with SN. The inferior thermal stability of PVD TaN metal gate compared to HfN metal gate might be attributed to its negatively smaller heat of formation and its inferior ability to block oxygen diffusion. Fig.11 depicts gate leakage at $V_{fb}$ – 1V as a function of EOT for HfN/HfO$_2$ without SN. Compared to poly/SiO$_2$ benchmark with the same EOT, HfN/HfO$_2$ shows more than 10$^5$ reduction in gate leakage. The work function ($\Phi_M$) of HfN on HfO$_2$, extracted from plots of $V_b$ versus EOT, after FGA is ~-4.75eV and slightly increases to ~4.8 eV after 1000°C RTA. The small $\Phi_M$ variation after 1000°C RTA is related to the HfN crystallization change.

**Fig.12.** (LogId)-Vg (left) Id-Vd (right) characteristics of a nMOSFET using HfN/HfO$_2$ gate stack (W/O SN) with EOT ~1.18nm.

**Fig.13.** (Log Id)-Vg [top] and Id-Vd [down] curves of pMOSFET with HfN/HfO$_2$ gate stack (no SN treatment) with EOT ~1.28nm.

**MOSFETs:**

Fig.12 show the typical electrical characteristics (Id-Vd & Id-Vg) of the n-MOSFET (EOT=1.18 nm) with subthreshold slope (ss) of ~78mV/dec. Fig.14 compares effective electron mobility ($\mu_e$) between n-MOSFETS devices with and without SN. Mobility is degraded for the SN device despite its larger EOT, which is attributed to the larger interface trap density (Dit) due to nitrogen as confirmed by DCIV interface trap density measurements. Using interface trap capture cross section of 4.4 $A^2$/V-s, Dit is ~3x10$^{10}$/cm$^2$ for fresh SN nMOSFET (EOT=2.04nm), and ~7x10$^{10}$/cm$^2$ for fresh device W/O SN with EOT~1.18nm. Fig.13 show well-behaved Id-Vd, and Id-Vg characteristics of p-MOSFETS with HfN/HfO$_2$ W/O SN (EOT=1.28nm) with effective hole mobility shown in Fig. 15.

**Fig.14.** Effective electron mobility comparison between devices w/ and W/O SN treatment. Mobility is degraded for SN device despite the larger EOT.

**Fig.15.** Effective hole mobility for the pMOSFET with HfN/HfO$_2$ stack with EOT~1.28nm (w/o SN)

**Summary**

Thermally robust high quality HfN/HfO$_2$ gate stack is demonstrated for advanced CMOS applications. EOT of HfN/HfO$_2$ gate stack has been successfully scaled down to less than 10Å with excellent leakage, boron penetration immunity, and long-term reliability even after 1000°C annealing, without using surface nitridation prior to HfO$_2$ deposition. The mobility is improved significantly without surface nitridation. Negligible change in both EOT and the work function of HfN/HfO$_2$ gate stack are observed after 1000°C RTA.

**IV. How high operating voltage can be employed for 10 years reliable operation , assessed by charge trapping and threshold voltage shift.**

It has been widely reported that HfO$_2$ gate dielectric suffers from serious charge trapping phenomena under stress, leading to severe bias temperature instability (BTI) degradations of MOSFETs. In this section, we review our recent results of BTI degradation in HfO$_2$ dielectric, emphasizing on the frequency dependence of the Dynamic BTI degradation in HfO$_2$ dielectric, and its impact on device lifetime.

**Experimental**

n-MOSFETs and p-MOSFETs with HfO$_2$ gate dielectric and HfN/TaN metal gate stack were fabricated with EOT ~1.3 nm for both n- and p- MOSFETs. Threshold voltage evolution under stress was characterized using a computer-controlled HP4156A semiconductor analyzer. The $I_p$-$V_g$ characteristics were monitored periodically during the
stressing under static uniform and dynamic non-uniform stress conditions\(^{14,15}\). The threshold voltage was extracted from the \(I_d-V_g\) measurement by linear extrapolation.

The \(V_{th}\) instability was first investigated under static stress for both n-channel and p-channel MOSFETs at room temperature, with sub-threshold swing (SS) monitored during the stress. For both n- and p- transistors, there is no observable change of the value of SS under stress, indicating that interface state generation plays no significant role and charge trapping in the bulk dielectric is the primary mechanism leading to BTI issues in high-K dielectrics\(^{14,15,16}\).

Under dynamic stress condition\(^3\), the \(V_{th}\) degradation is strongly dependent on the frequency of the dynamic stress, as shown in Fig.16. The BTI degradation under dynamic stress is reduced as compared to that under static stress. This reduction becomes stronger as the stress frequency is increased (up to 1 MHz as demonstrated in this work).

**Fig. 16.** \(V_{th}\) shift time evolution for MOSFET, under static and dynamic stresses of different frequency, for (a) nMOSFET, and (b) pMOSFET. The \(V_{th}\) evolution was shown to have a power law dependence on stress time, having a fast initial stage followed by a slow stage.

**Lifetime Projections**

The 10-years lifetime projections based on \(\Delta V_{th} = 30 \text{ mV}\) as failure criterion are shown in Figs.17 and 18 for various stress frequencies for n- and p-MOSFETs respectively. As can be seen, device lifetime is significantly enhanced with the increase of stress frequency. The 10-year operation voltages for n- and p-MOSFETs are 0.8 V (and -0.9V) for static BTI, and 1.3 V (and -1.7 V) for dynamic BTI at 1 MHz.

**Fig. 17.** 10-years lifetime projection for n-HfO\(_2\)/MOSFETs, based on \(V_{th}\) shift. \(\Delta V_{th} = 30 \text{ mV}\) is set as the device failure criterion. \(V_{stress} = 1.8 \text{ V}\) for static BTI and 1.3 V for dynamic BTI at 1 MHz.

**Modeling and discussion**

We have developed a model which shows good agreement with all experiment data, particularly the frequency dependence of \(V_{th}\) degradation in n-MOSFETs. The model can be extended to p-MOSFETs. For static stress, following [14,15], it is proposed that pre-existing electron traps in the bulk dielectric are responsible for the \(V_{th}\) shift in n-MOSFETs. Under static positive stress voltage, the electrons are captured (trapping) by the traps distributed in the dielectric. It is observed from Fig.16 that in trapping, \(\Delta V_{th}\) has a power law dependence on time, giving rise to a straight line in \(V_{th}\)-time plot on a log-log scale. This power-law dependence can be obtained by assuming that electron traps have a distribution in the trapping time constant domain, \(N(t_c)\)\(^{15}\). In this work, two-peak (fast and slow) distribution was used. By adjusting the distribution parameters (Table I), good agreement between simulation and the static BTI experiments can be obtained (Fig.19a).

In the case of dynamic BTI, although HfO\(_2\) dielectrics exhibit severe \(V_{th}\) shift after stress, large portion of the \(V_{th}\) shift can be recovered during an electrical passivation phase when the gate to substrate voltage is zero, due to charge de-trapping, as shown in Fig.19(b). In this work, the de-trapping time constant \(t_c\) for every trap has a simple correlation with the trapping time constant \(t_c = 100 t_c\).

**Fig. 18.** 10-years lifetime projection for p-HfO\(_2\)/MOSFETs, based on \(V_{th}\) shift. \(\Delta V_{th} = 30 \text{ mV}\) is set as the device failure criterion. \(V_{stress} = 0.9 \text{ V}\) for static BTI and -1.7 V for dynamic BTI at 1 MHz.

To explain the frequency dependence of \(V_{th}\) shift observed in dynamic BTI, we consider the number of trapped electrons \(\Delta n\) during one cycle in the dynamic BTI experiment, as shown in Fig.20 (a). If \(\Delta n\) grows linearly with \(\Delta t\) in the trapping half-cycle, it leads to an aggregated trapped charge that is...
From the above discussion based on our investigation, HfO₂ based high-k gate dielectric can be used until or after year 2016 for the 25 nm node LSTP applications, with a interfacial layer with dielectric constant higher than that of SiO₂. The recorded EOT up to year 2003 is 0.6 nm and the main obstacle is to grow as thin as possible interfacial layer which can effectively improve the channel mobility and dielectric reliability. How to grow HF-based gate dielectric with EOT = or less than 0.5 nm is under consideration. Charge trapping induced BTI degradation in n-MOSFET is the killer of HfO₂ devices. Dynamic BTI degradation is reduced when the operation frequency is increased. 10-year operation voltage for ultrathin HfO₂ n-MOSFET is 0.8V for static operation and 1.3 V for dynamic operation at 1MHz

This work was supported by Singapore A-STAR research grant R263-000-267-305. We acknowledge many colleagues and students in SNDL/NUS and IME who have contributed to this work however not appeared in the authors list.

Table 1 Time constants used in calculation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta (T_e)$</td>
<td>Distribution function of capture time constant $\tau_c$</td>
<td>$N(\tau_c)$</td>
</tr>
<tr>
<td>$\tau_c$</td>
<td>sum of two log-normal distributions. peak width $\sigma$</td>
<td>$\tau_c$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>40µs, 0.1s</td>
<td>40µs, 0.1s</td>
</tr>
<tr>
<td>$\tau_e$</td>
<td>Emission time constant</td>
<td>100µs $\tau_c$</td>
</tr>
<tr>
<td>$\tau_a$</td>
<td>Activation constant</td>
<td>80µs $\tau_c$</td>
</tr>
<tr>
<td>$\tau_d$</td>
<td>Deactivation constant</td>
<td>5µs $\tau_c$</td>
</tr>
</tbody>
</table>

References:

2. S.J.Lee et al, IEDM 2000, p. 31.