Challenges of Process Technology in 32nm Technology Node

Wu Hanming[1,†], Wang Guohua[1], Huang Ru[2], and Wang Yangyuan[2]

(1 *Semiconductor Manufacturing International Corporation (SMIC), Beijing* 100176, *China*)
(2 *Peking University, Beijing* 100871, *China*)

Abstract: According to the international technology roadmap for semiconductors (ITRS), 32nm technology node will be introduced around 2009. Scaling of CMOS logic devices from 45 to 32nm node has come across significant barriers. Overcoming these pitch-scaling induced barriers requires integrating the most advanced process technologies into product manufacturing. This paper reviews and discusses new technology applications that could be potentially integrated into 32nm node in the following areas: extension of immersion lithography, mobility enhancement substrate technology, metal/high-k (MHK) gate stack, ultra-shallow junction (USJ) and other strain enhancement engineering methods, including stress proximity effect (SPT), dual stress liner (DSL), stress memorization technique (SMT), high aspect ratio process (HARP) for STI and PMD, embedded SiGe (for pFET) and SiC (for nFET) source/drain (S/D) using selective epitaxial growth (SEG) method, metallization for middle of line (MOL) and back-end of line (BEOL), and ultra low-k (ULK) integration.

Key words: CMOS technology; 32nm technology node; mobility enhancement; metal gate/high k dielectrics; ultra low k dielectrics
EEACC: 2570
CLC number: TN405　　**Document code**: A　　**Article ID**: 0253-4177(2008)09-1637-17

1　Introduction

As predicted by the Moore's law, the past half-century witnessed the rapid progress of the integrated circuit (IC) due to the interaction of technology push and market pull. Currently in the world, state-of-the-art 90nm/65nm IC technology has already entered into mass production. Recently, Intel Corporation has moved on to the leading-edge 45nm technology node and successfully put it into CPU development, marking the industrialization of this node. Simultaneously, companies are on their way to researching and developing the 32nm and beyond CMOS technology.

In China, mass production of the 45nm CMOS IC will be turned into reality as a result of industrialization within one or two years. On the other hand, research and development of 32nm and beyond technology needs the close cooperation of industries, universities, and research institutes with great efforts to realize mass production and form core competencies, thus enhancing our strength and sustainable development in IC technology. In this paper, we will review the key points of 32nm technology node.

With the encroachment of 2009, as shown in Fig. 1, the resolution (half pitch) will approach 50nm, where 32nm technology node will be introduced. However, due to this aggressive scaling, the deficits for device performance will become more severe. Some techniques used in previous generations will still be effective for device performance improvement, such as some strain enhancement methods, including SPT[1], DSL[1], and SMT[2], but more critical techniques will likely become employed, such as double patterning[3~5], direct silicon bonding (DSB)[6~8], hybrid orientation bulk substrate, metal/high-k (MHK) gate stack[9~11], SiC S/D for nFET[12], two-step like SEG SiGe S/D[13,14], millisecond anneals for USJ, new metallization technology, and new ULK materials. By incorporating these creative techniques into the product line, confidence for delivering 32nm technology node on time should be rebuilt. This paper tries to provide a platform for reviewing all possible process technologies used for 32nm node and will be organized as follows: Section 2 is about 32nm possible

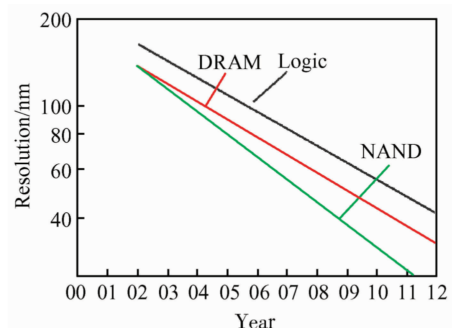

Fig.1　2006 ITRS Roadmap (From Ref.[1])

† Corresponding author. Email: hanming_wu@smics.com
Received 24 June 2008, revised manuscript received 5 July 2008
©2008 Chinese Institute of Electronics

Table 1 Design rules from 65nm to 32nm technology node

Tech node	65nm	45nm	32nm
AA *L/S*	80/100	60/80	45/55
GT *L/S*	60/120	40/100	35/65
CT *L/S*	90/110	60/80	40/60
M1 *L/S*	90/90	60/80	45/55

lithography solutions (i.e. double patterning); in Section 3, hybrid orientation bulk substrate, fabricated by DSB method, will be reviewed; for Section 4, different MHK gate stacks and mechanisms for flat band voltage (V_{fb}) roll-off[15~19] will be introduced; in Section 5, different stress enhancement methods will be stated, but with more focus on SiGe; Section 6 will discuss potential techniques used for USJ; In Section 7, new metallization technology for MOL and BEOL will be introduced; ULK integration will be introduced in Section 8; Conclusions are in Section 9.

2 Possible lithography solutions for 32nm node-double patterning

Design rules for some critical layers at 32nm node are listed in Table 1, where the pitch is 100nm, L means the length of the feature, and S is the space between features. By applying these design rules, one SRAM unit cell size could be as small as $0.125\mu m^2$. Possible lithography solutions for this small pitch have been studied since dry 193nm lithography became relatively mature several years ago[20]. Currently there are five basic lithography horses running for the mass production race: (1) EUV lithography; (2) immersion DP (double patterning); (3) SP/DP; (4) E-beam technology; and (5) High refractive flux immersion. In the view of 32nm node manufacturing, the two most promising approaches are EUV and immersion DP. Extreme ultraviolet lithography (EUV) was believed to be one of them, since the resolution (R) is determined by

$$B = k_1 = \frac{\lambda}{\mathrm{NA}} \quad (1)$$

where λ is the wavelength, NA is the numerical aperture, and k_1 is the measure for system capability. For EUV, λ is ~13.5nm, so the resolution requirement for 32nm node can be met. However, its progress is far behind schedule, and its timely delivery is under serious question right now. ArF ($\lambda \approx 193$nm) immersion lithography with water-based 1.35NA was introduced at 45nm node. At this high NA and with state of art $k_1 \approx 0.35$ (the physical limit for single exposure $k_1 \approx 0.25$)[4], the 50nm resolution requirement for 32nm node was barely reached, but no practical process window was left. Thus, using high-index immersion lithography with achievable 1.55NA is also under serious examination. Right now, lots of work about corresponding photoresists (PR) and lens materials still need to be done, and even if these issues have been taken care of, high-index immersion lithography is a one-generation solution, and its extendibility to next generation will be a problem. Thus, for 32nm node, double patterning becomes a potential bridge between standard water-based immersion lithography and EUV. By sacrificing manufacturing throughput and cost of ownership (CoO), double patterning could effectively reduce k_1 by half, and requirements for 32nm node will all be met. Basically, double patterning splits a pattern into two separate, less dense patterns. Some problems come after this pattern splitting, including overlay and overlay-affected CD[3]. For double patterning at 32nm node, the overlay requirement is somewhere between 1.5 and 3nm, about 10% of the CD.

Different schemes for double patterning have been explored[3,4], including double exposure/double etch (as shown in Fig. 2(a)), single coat/single etch (as shown in Fig. 2(b)), double coat/single etch (as shown in Fig. 2(c)), and spacer self-aligned double patterning (as shown in Fig. 2(d)). Not all of these four are plausible right now. For schemes shown in Figs. 2(b) and 2(c), introduction of new materials is the key concern. The scheme shown in Fig. 2(b) needs a memory-free photoresist, and the scheme shown in Fig. 2(c) needs a photoresist with freezing ability, but these materials are still at research status. The scheme shown in Fig. 2(a) is widely known, but suffers a large number of process steps (two masks, two exposures, and two etches) and high cost. The spacer self-aligned double patterning scheme, shown in Fig. 2(d), is another mature manufacturing scheme, with comparable cost to the one shown in Fig. 2(a). Based on available information, the scheme shown in Fig. 2(a) could be the most promising candidate for double patterning lithography application.

3 DSB hybrid orientation bulk substrate

Holes and electrons are known to have the highest mobility on (110) and (100) silicon, respectively[21]. It is highly desirable to have a hybrid orientation bulk substrate such that each type of carrier has the most optimized crystal orientation to achieve the highest mobility. For 32nm node, such a substrate might be needed to compensate for other stress technologies' performance degradation due to the aggressive pitch scaling. A hybrid orientation technology using direct silicon bonding (DSB) for bulk CMOS is

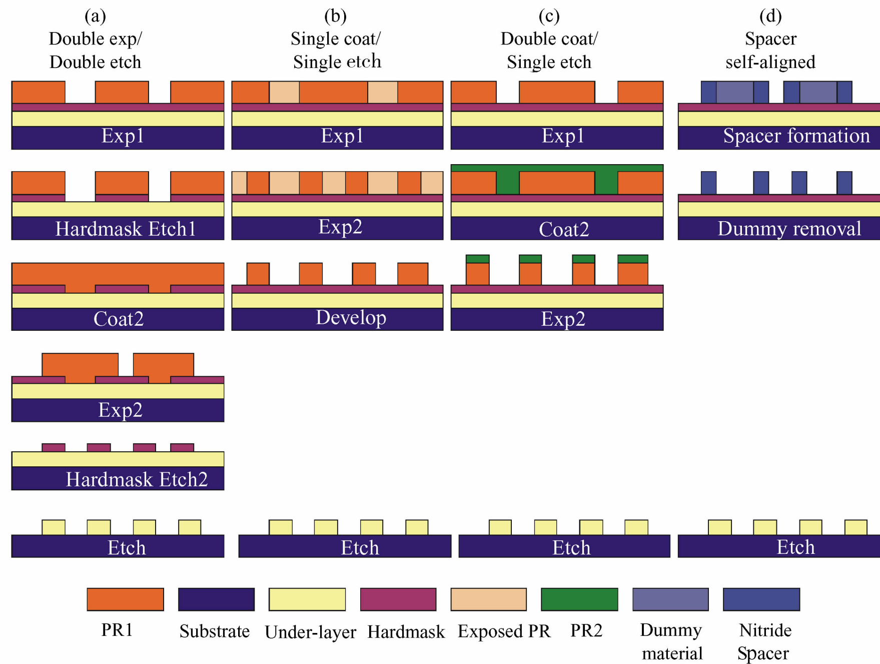

Fig.2　Different schemes of double patterning (From Ref.[4])　(a) Double exposure/double etch; (b) Single coat/single etch; (c) Double coat/single etch; (d) Spacer self-aligned double patterning

demonstrated[6,7]. DSB substrate is prepared by directly bonding a Si film onto a Si substrate with a different crystal orientation. The fabrication process[6] is shown in Fig.3. For now, a structure of a (110) Si film on (100) Si substrate is explored. A cross-section TEM picture with high resolution is shown in Fig. 4[6], from which we can see that great interfacial quality is achieved.

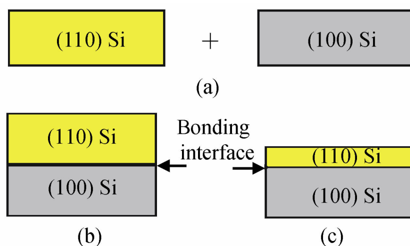

Fig.3　DSB hybrid orientation bulk substrate　(a) (110) bonded wafer and (100) substrate wafer; (b) Direct bonding of the two wafers; (c) Removal of the (110) Si substrate (From Ref.[7])

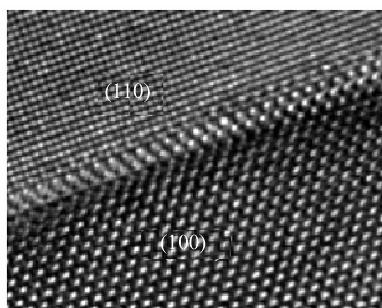

Fig.4　TEM picture for DSB hybrid orientation bulk substrate (From Ref.[7])

There exist two integration schemes for the application of DSB technology[6~8]: the first is to finish crystal orientation conversion by solid phase epitaxy (SPE) before the shallow-trench-isolation (STI), which is named SPE-before-STI, as shown in Fig.5; the second is to form STI before orientation conversion by SPE, which is named STI-before-SPE, as shown in Fig.6. For SPE-before-STI, during the SPE process, defective regions will form between the (110) and (100) regions, as demonstrated in Fig.7 (a). The size of the defective regions are strongly dependent on the bonded film thickness, which is a critical knob for device performance control[8]. These defect regions could be removed by following STI in a self-aligned process, as demonstrated in Fig.7(b). For

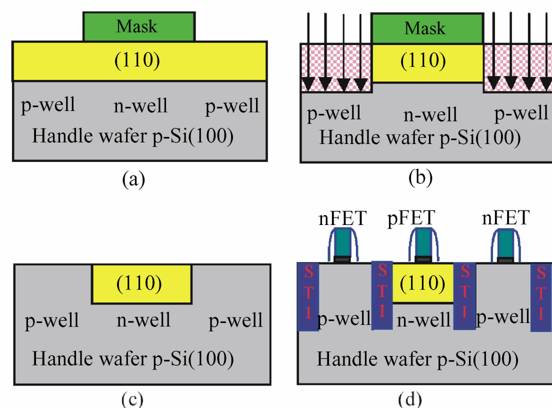

Fig.5　SPE-before-STI (From Ref.[8])　(a) Mask pFET regions; (b) Amorphize nFET regions; (c) nFET regions conversion to (100) by SPE; (d) STI formation

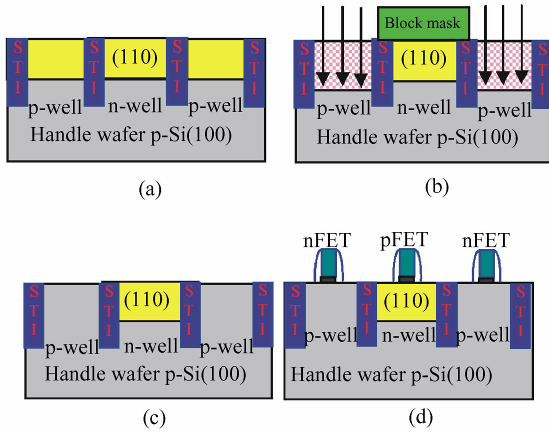

Fig. 6 STI-before-SPE (From Ref. [8]) (a) STI formation; (b) Amorphize nFET regions; (c) nFET regions conversion to (100) by SPE; (d) Process flow continues

STI-before SPE, defects could also form at the STI trench edge, as shown in Fig. 7(c), but these defects can be eliminated by changing the notch orientation between the bonded film and corresponding substrate[8]. By optimizing the DSB technology process, DSB hybrid orientation substrate could be very effective for mobility enhancement, and also all 32nm node ground rules can be met[8].

4 Metal/high-k gate stacks

With the technology node proceeding, further gate oxide scaling cannot meet the tunneling leakage requirement, although corresponding terminal voltages are also lowered. In order to meet the gate leakage requirement, from 45nm node, high-k materials have been introduced without compromising electrical thickness. Using high-k, we could have continuous equivalent oxide thickness (EOT) scaling, but with thicker dielectrics physical thickness ($T_{\text{high-}k}$). This is due to the fact that the EOT of a dielectric is inversely proportional to its dielectric constant[22]:

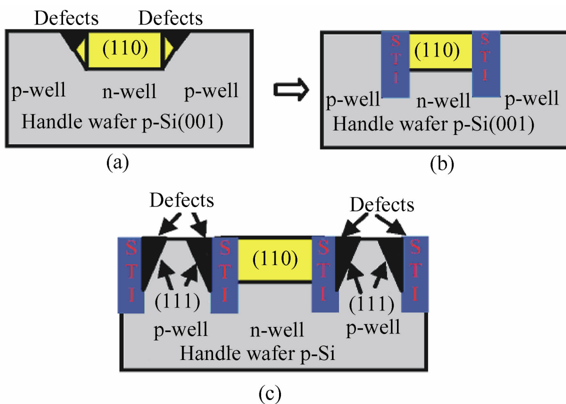

Fig. 7 (a), (b) Defect regions removal using STI in SPE-before-STI scheme; (c) Defects formation in STI-before-SPE scheme (From Ref. [9])

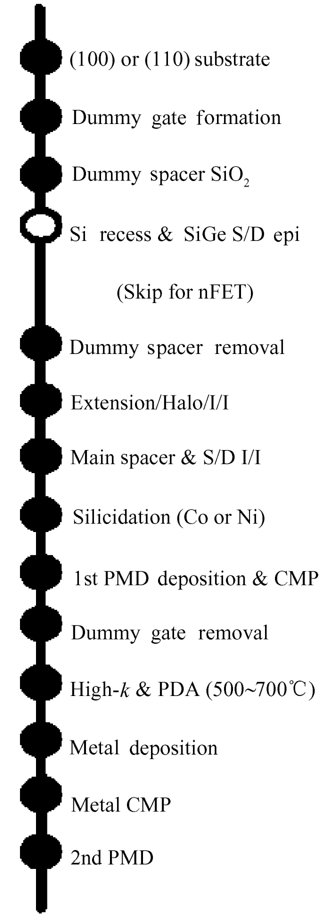

Fig. 8 Gate last process flow for MHK gate stack (From Ref. [26])

$$\text{EOT} = \left(\frac{\varepsilon_{\text{SiO}_2}}{\varepsilon_{\text{high-}k}}\right) T_{\text{high-}k} \quad (2)$$

where $\varepsilon_{\text{high-}k}$ is the high-k permittivity, and $\varepsilon_{\text{SiO}_2}$ is the SiO_2 permittivity. $\varepsilon_{\text{SiO}_2}$ is much smaller than $\varepsilon_{\text{high-}k}$, thus $T_{\text{high-}k}$ is much thicker and a reduction of gate leakage current is achieved. Hafnium-based dielectric (HfO_2, $k=25$; HfSiO/HfSiON, $k=15$) is the most promising high-k candidate. Regardless of the possibility that 32nm low power may not adopt MHK technology for manufacturing, there is almost no doubt that 32nm generic technology needs MHK inevitably for gate leakage reduction. For 32nm node, high-k gate dielectrics with $k>20$ is necessary for EOT continuous scaling <0.9nm[23].

Thermal stability is always a concern for the MHK gate stack[24]. In order to take care of this concern, a gate last process is employed where MHK deposition is done after S/D activation anneals such that MHK is not exposed to high temperatures. Sony, as shown in Fig. 8, proposed a typical flow for this gate last process[25], where SiGe S/D is also integrated. Corresponding MHK schematic configurations are shown in Fig. 9, where a $HfO_2/HfSi_x/TiN/W$ stack is used for nFET and a $HfO_2/TiN/W$ stack is used for

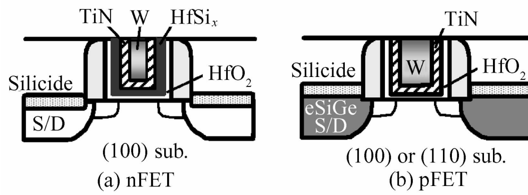

Fig. 9　Schematic diagrams for MHK gate stacks made from gate last process flow (From Ref. [26])　(a) nFET gate stack configuration; (b) pFET gate stack configuration

pFET. In this gate last process sequence, a standard process up to 1st PMD deposition and CMP is performed. Then, the MHK gate trench is formed by dummy gate removal. After the removal, an ALD-HfO_2 layer is deposited, followed by a post-deposition anneal (PDA) at 500～700℃. Thereafter, metal gate depositions are respectively done for nFET and pFET.

By inserting the capping layer or other threshold voltage (V_t) (or flat-band voltage (V_{fb})) tuning techniques[10,11,], MHK gate stacks with high thermal stability can be achieved using a conventional gate first process flow[27～29]. A general gate first process for making these MHK gate stacks is shown in Fig. 10, which could enable us to have full flexibility to fabricate different configurations of MHK gate stacks. In this flow, after the formation of the 1st hard mask on nFET, the 1st high-k and metal gate on pFET can be removed by either dry or wet etching methods dependent on gate material selections. Then, the 2nd high-k and metal gate are deposited on pFET. The 2nd hard mask is deposited and patterned to remove the 2nd high-k and metal gate on nFET. After the separation of high-k and metal gate on the nFET and pFET regions, a poly cap layer is deposited, which is followed by a gate stack dry etch. Thereafter, a normal process sequence will resume.

Usually, MHK gate stack configurations for nFET and pFET are different. A stack of HfO_2/0.1～2nm La or Mg containing cap layer/TiN/Poly-Si for nFET[10] and a stack of HfSiO/AlN cap layer/Poly-Si for pFET[30] were reported by IBM. Similarly, SEMATECH proposed a HfTiSiON[23] (or HfLaSiON[31])/metal stack for nFET and a HfSiO/MAlN (M = Ta, Ti, Mo, W)[11] stack for pFET. The reason for different gate stack configurations is not only to get thermal stability, but also to get different band edge (BE) work functions (WF) for nFET (WF ～ 4eV) and pFET (WF～5.2eV)[32,33] respectively, such that V_t (or V_{fb}) for each type of MOSFET could be at a reasonable value. In order to have an appropriate V_t (or V_{fb}) design, a thorough understanding of the V_{fb} shift is strongly required. A lot of work has been done, and an interface dipole model was proposed[15～19]. In this model, metal gate WF on high-k ($\phi_{m,eff,high\text{-}k}$) is determined not only by intrinsic dipole induced shift ($e\Delta^{intrinsic}_{dipole}$) and extrinsic dipole induced shift ($e\Delta^{extrinsic}_{dipole}$) at the high-$k$/$SiO_2$ interface but also by the metal work function on SiO_2 (ϕ_{m,eff,SiO_2}), as follows[18]:

$$\phi_{m,eff,high\text{-}k} = \phi_{m,eff,SiO_2} + e\Delta^{intrinsic}_{dipole} + e\Delta^{extrinsic}_{dipole} \quad (2)$$

where the values for $e\Delta^{intrinsic}_{dipole}$ and $e\Delta^{extrinsic}_{dipole}$ could be either positive or negative. The sign is determined by the dipole pointing direction. If the dipole points to high-k, the sign for the shift is negative, because it is similar to putting some electrons into the Fermi Sea. Thus, the Fermi level is raised. If the dipole points to SiO_2, the opposite will happen. The value for an intrinsic dipole is determined by the MHK gate stack material selections. For example, if Al_2O_3/HfO_2 is used as high-k, the intrinsic dipole will point to SiO_2, and a positive shift will be expected, but if Y_2O_3/HfO_2 is used as high-k, the intrinsic dipole will point to high-k, and a negative shift will be expected. Extrinsic dipole value is related to oxygen vacancy generation at different situations, including different annealing conditions. A schematic band diagram understanding for this model is shown in Fig. 11. This figure shows that by setting the values of intrinsic and extrinsic dipoles properly, $\phi_{m,eff,high\text{-}k}$ could be tuned to the target value. Although this model explains most cases for V_{fb} roll-off in MHK gate stacks, there are some exceptions in which Fermi level pinning at the metal gate and high-k interface need to be taken into account.

5　Different carrier mobility enhancement technologies

Starting from 45nm technology node, different stress technologies for mobility enhancement have been widely employed. Various stress silicon technologies have been developed from 90nm technology node since 2001[2,13,14,34～38]. The mobility improvement budgets with different kinds of stress technologies are compared in Table 2. Of all the available stress technologies, the most effective mobility enhancement for pMOS is selective epi-growth (SEG) SiGe S/D. Since the data in the table is from 65/45nm, we are aware that some mobility enhancement may be even better in 32nm node as the critical size shrinks down.

For the stress liner technique, the improvement of the device performance is determined by the liner stress level, thickness of the liner, and liner proximity to the channel. By using DSL, the device performance for nFET and pFET can be improved together. A

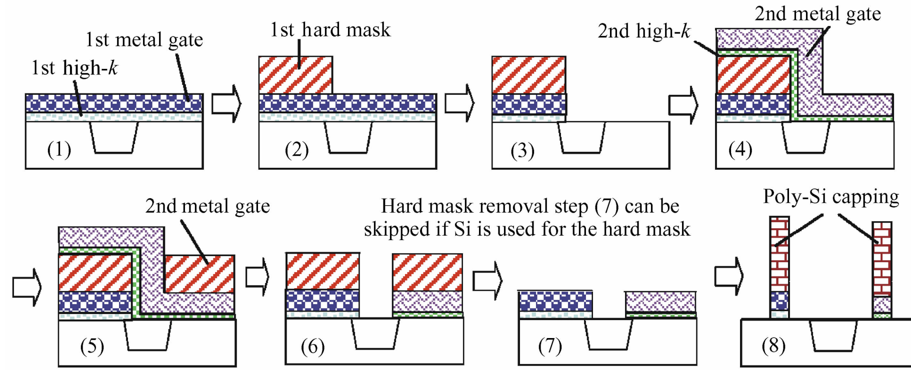

Fig. 10 A general gate first process flow (From Ref. [28]) (1) 1st metal gate and high-k deposition; (2) 1st hard mask formation; (3) 1st metal gate and high-k removal on PFET regions; (4) 2nd metal gate and high-k deposition; (5) 2nd hard mask deposition and patterning; (6) 2nd metal gate and high-k selective removal; (7) Hard mask removal; (8) MHK gate stack formation

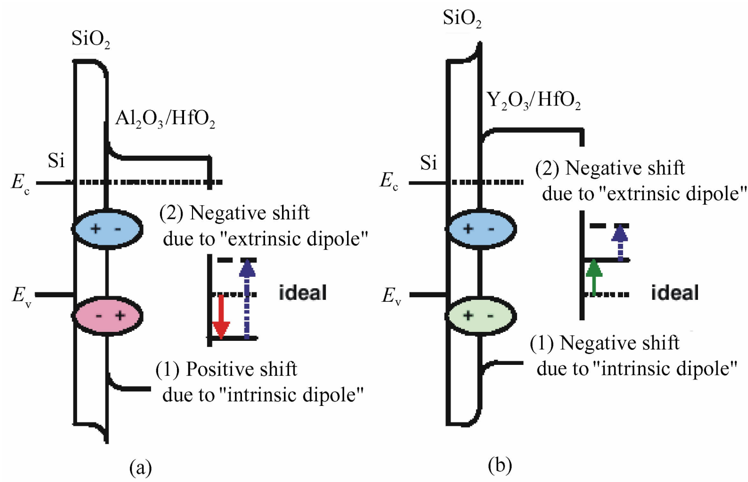

Fig. 11 V_t/V_{fb} tuning mechanism schematic understanding (From Ref. [19]) (a) Configuration for negative extrinsic dipole and positive intrinsic dipole at high-k/SiO_2 interface; (b) Configuration for negative extrinsic dipole and negative intrinsic dipole at high-k/SiO_2 interface

schematic demonstration for DSL is shown in Fig. 12 (a)[2], where a tensile nitride liner is used for nFET and a compressive liner is used for pFET. In Fig. 12 (b), a schematic diagram for SPT is shown. This figure shows that stress liners are deposited after the removal of the spacer 2. By doing so, the liners become more proximate to the channel, and more stress can be introduced into the channel, further enhancing the device performance[1]. For pFET with SPT, drive current improvements of 20% for isolated and 28% for nested devices are reported. SMT is another advanced stress technique for nFET performance improvement[3]. The process flow for SMT is schematically demonstrated in Fig. 13. After nFET poly pre-amorphization implantation (PAI), a high tensile (>1.5GPa) nitride liner is deposited, as shown in Fig. 13

Table 2 Comparison of improvement budget among some stress technologies

	nMOS	pMOS
Stress memorization technique	7%～15%	none
Single tensile nitride layer	10%～19%	none
Dual stress liner	10%～11%	18%～20%
Stress proximity technique	none	20%
eSiGe	none	20%～65%

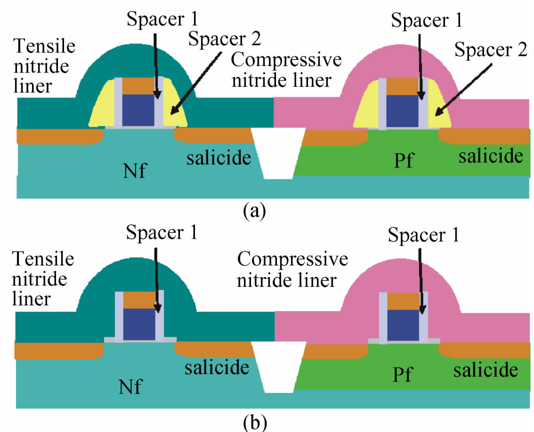

Fig. 12 (a) Schematic diagram for DSL; (b) Schematic diagram for SPT with DSL (From Ref. [2])

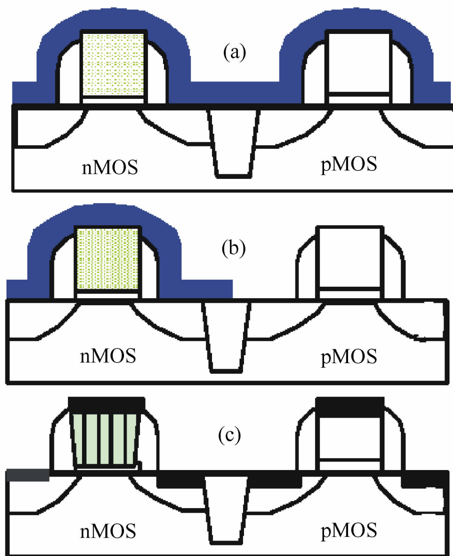

Fig. 13　Schematic process flow demonstration for SMT (From Ref. [3])　(a) High tensile stress liner deposition on both nFET and pFET; (b) Selective removal of stress liner on pFET; (c) After S/D activation anneal, remove stress liner left on nFET

(a). In order to alleviate the stress induced performance degradation on pFET, the tensile stress liner on pFET is selectively removed, as shown in Fig. 13(b). After S/D activation anneal, the stress liner on nFET is also stripped, as shown in Fig. 13(c), but after this activation anneal process, the stress is already transferred and memorized in the channel. As reported, more than a 15% current drivability improvement for nFET can be achieved using SMT.

There are some other stress improvement techniques, such as STI and PMD (pre-metal dielectric) gap-fill processes[39,40]. Right now, for 65nm node, high-density plasma chemical vapor deposition (HDP-CVD) is used for STI and PMD gap-fill. However, as geometries scale to 45nm and 32nm nodes, gap-fill process becomes more challenging due to the inherent high aspect ratio (AR) for STI and PMD. Thus, a high AR gap-fill process (HARP) is proposed. HARP could provide void-free gap-fill for STI $<$ 30nm ($>$ 10 : 1 AR) and PMD $<$ 10nm ($>$ 6 : 1 AR), which can meet the requirements for 32nm node. But for STI gap-fill, HARP is strongly dependent on the profile of an RIE etched STI trench, and it is required that the STI sidewall slope should be $> 87^\circ$ and no double slope exists. The gap-fill film used for HARP is sub-atmospheric chemical vapour deposition (SACVD) oxide, which is O_3/TEOS based. HARP film, in nature, is tensile (for as deposited film, the stress level $\sim$ 200MPa)[39]. Due the two-dimensional stress geometry effects[40], the device performance for both nFET and pFET could be enhanced at the same time. It is reported that compared with HDP-CVD STI and PMD gap-fill, HARP gap-fill for STI and PMD can enhance drive current 11% and 18% for nFET and pFET, respectively[39,40]. One additional benefit for this HARP process is that it is plasma free, thus the damage to the underlying films is minimized.

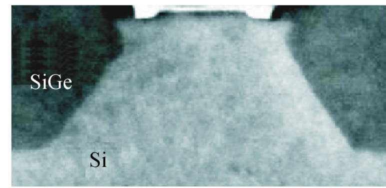

Fig. 14　TEM picture for shape SiGe S/D (From Ref. [44])

SiGe S/D is an effective way to introduce compressive uni-axial strain into the channel for pFET[13,14,41,42]. For 32nm node, this technique will continue to be employed. The profile of SiGe S/D will strongly affect the strain level introduced into the channel[13,14,43,44]. SiGe S/D with shape has been reported[43]. A cross-section TEM picture for this Σ shape SiGe S/D is shown in Fig. 14. This shape is realized using RIE recess etching first, as shown in Fig. 15 (a), then Si implantation is employed to amorphize the recessed bottom, as shown in Fig. 15(b). Thereafter, the wet etching method with strong selectivity along the (111) Si crystal plane is used, thus Σ shape SiGe S/D is formed. For 32nm technology node, more strain in the channel is needed. It is likely that Σ shape SiGe S/D cannot fulfill this requirement, so a two-step like SiGe S/D profile is proposed[13,14]. Figure 16 is a cross-section TEM picture for such a SiGe S/D profile. The process flow to make such a two-step like SiGe S/D profile is shown in Fig. 17, from which it can be seen that after the offset spacer, a shallower step is first formed, then, with the help of a dummy spacer, a deeper step is formed. Using this two-step like SiGe S/D profile, a 100% drive current improvement for pFET was reported[14].

Although the SiGe S/D profile is a critical factor in control, the $Si_{1-x}Ge_x$ SEG process is also very important[45,46]. Ge composition, x, can be controlled during this growth process. Normally for 45nm node, x is around 0.25, but for 32nm node, an x value around 0.3 seems necessary, since a greater x introduces more strain into the channel. Generally, a SiGe/Si heterostructure is at a metastable state and tends to relax itself during heat treatment by forming defects in the form of misfit dislocations, stacking faults, and so on. These defects could exist in SiGe S/D or at the SiGe/Si substrate interface. These defects could incur detrimental problems, such as enhancing dopant diffusion, degrading current drivability due to strain relaxation, and so on. Thus, a high quality SEG SiGe layer

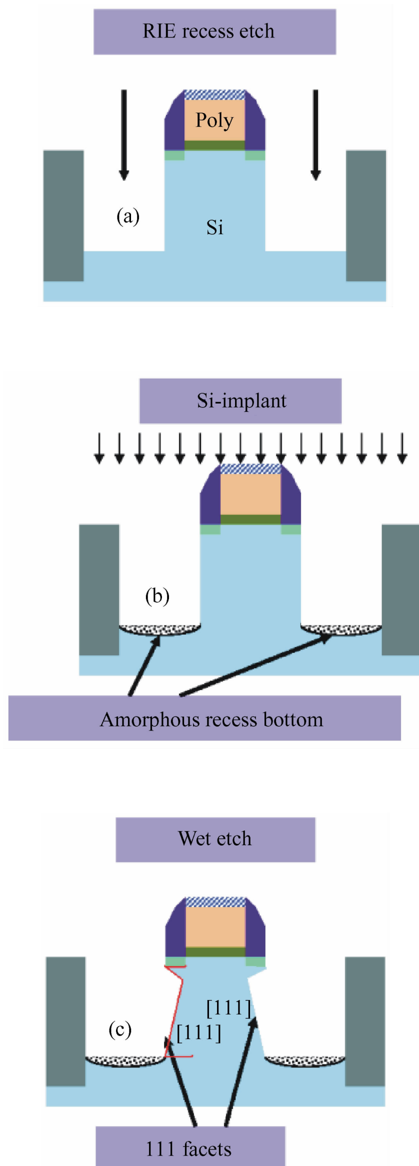

Fig. 15 Σ shape SiGe S/D formation process (a) RIE recess etch; (b) Implantation to amorphize recessed bottom; (c) Wet etching method with strong selectivity along (111) crystal plane used

is necessary to minimize these defects and improve the introduced strain in the channel. This can be real-

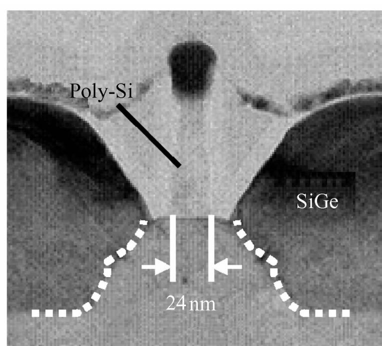

Fig. 16 TEM picture for two-step like SiGe S/D (From Ref. [14])

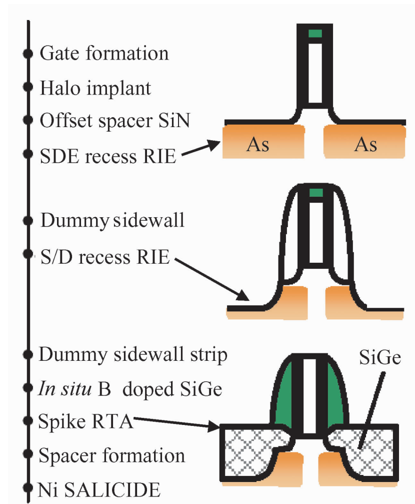

Fig. 17 Process flow for fabricating two-step like SiGe S/D (From Ref. [14])

ized by taking care of a few points. One of them is *in situ* low temperature wet pre-cleaning, using diluted HF before SiGe SEG[46]. This pre-cleaning removes interfacial impurities such as O and C, and could also recover the plasma damage induced by RIE. The quality of the SEG SiGe layer is also dependent on the recessed shape, specifically the recessed crystal plane. It was reported that SiGe epitaxial morphology is degraded on Si (110) substrate, but significantly improved on (100) substrate[46]. Good selectivity control of the epitaxial growth process is another aspect that needs to be taken care of. Optimizing the flow rate of HCl, which is used in the growth process, could ensure this selectivity.[46] One more thing that could be controlled during this SiGe SEG process is the Ge profile in SiGe S/D. This could be realized by changing the Ge containing precursor flow ratio, and it is reported that a graded Ge composition in SiGe S/D with an optimized Ge profile could introduce more strain into the channel, and an extra 15% drive current enhancement over a non-graded SiGe S/D was reported[47]. One more benefit for this kind of graded Ge profile is to inhibit dopant diffusion and provide a broader heat process window[48]. Another thing could be taken care of during this epi growth process is germanosilicide[49]. Usually a sacrificial Si only cap layer on top of SiGe is used to suppress Ge out diffusion during the silicidation process, and this cap layer could be realized by *in situ* growth right after SiGe growth. In a word, successful SiGe SEG control is the decisive factor for whether this stress technique is effective or not.

Another potential application for SiGe is for nFET, which is called reverse embedded SiGe (Rev. e-SiGe)[50]. A TEM demonstration picture for this structure is shown in Fig. 18. The SiGe island is com-

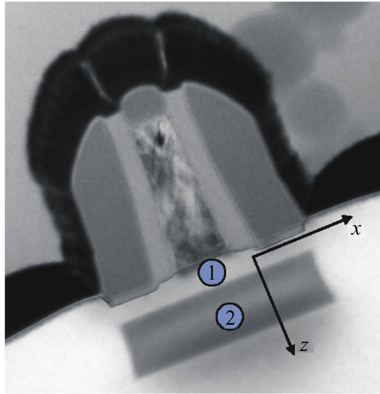

Fig. 18 TEM picture for nFET with Rev. e-SiGe (From Ref. [51]) Region 1 is Si; region 2 is SiGe.

pressively embedded under the Si channel, and the channel, thus, will be under tensile strain. The strain level in the channel is modulated by a few parameters[50], such as the thickness of the Si channel and the thickness and width of the SiGe island. The fabrication process for this Rev. e-SiGe is schematically demonstrated in Fig. 19. As shown in Fig. 19(a), substrate with epitaxial SiGe/Si is first prepared, and then, STI, gate patterning, and spacer formation follow, as shown in Fig. 19(b). Figure 19(c) shows RIE etched S/D regions, which are followed by Si SEG, as shown in Fig. 19(d). nFET performance with this Rev. e-SiGe can be greatly improved, with 40% mobility enhancement and 15% drive current improvement[50].

For 32nm node, in order to maintain the necessary strain level in the channel, other methods need to be incorporated, including SiC S/D[12] for nFET. A straightforward approach to fabricate SiC S/D is to use RIE to recess S/D, and thereafter, SiC gas-phase SEG follows. Similar to SiGe S/D, the SiC SEG process is also very critical. The process flow is demonstrated in Fig. 20. Another approach is to directly dope S/D regions using the cluster ion implantation technique or the PAI integrated C implantation technique[12]. The cluster ion implantation technique gets its name from the new implantation species it uses.

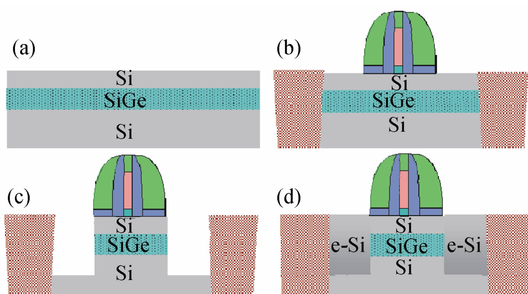

Fig. 19 Fabrication process for nFET with Rev. e-SiGe (From Ref. [51]) (a) Epitaxial SiGe/Si growth; (b) STI formation, gate patterning and spacer formation; (c) S/D RIE; (d) Si SEG

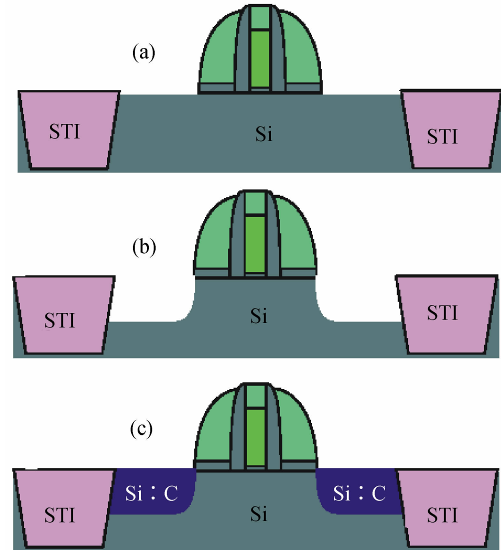

Fig. 20 SiC S/D formation using SEG (From Ref. [13]) (a) nMOS substrate; (b) S/D RIE; (c) SiC S/D SEG

These species are in the form of a large ion cluster. Usually, they are $B_{18}H_{22}$, C_7H_7, and so on. After the implantation process, solid phase epitaxial growth (SPE) follows. The process flow is shown in Fig. 21. For SiC S/D, even a relatively small C concentration (1～2at%) at substitutional sites can induce a high stress level in the channel, ～ 0.7GPa.

6 Ultra-shallow junction (USJ)

With the scaling of the entire transistor, the S/D junction will also become shallower. For 32nm technology node, the USJ depth, X_j, should be in the range of 12～15nm with sheet resistance (R_s) $<1k\Omega/\square$[51]. Usually, a shallow junction is achieved by low energy and high current implantation with Si or Ge PAI. Additional carbon implantation is also frequently used to get USJ, which could change the Si material properties such that dopant diffusion during subsequent anneal steps could slow down[52]. However, low energy USJ implantation will lower the manufacturing throughput. In order to compensate for this, the cluster ion implantation technique may be used. Due to the large size of the species used in cluster ion implantation, the implantation process is self-amorphized, and thus, no more PAI is needed. Furthermore, much higher energy could be used during this

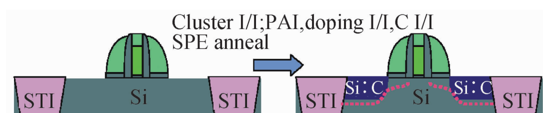

Fig. 21 SiC S/D formation by SPE using cluster ion implantation or PAI integrated implantation technique (From Ref. [13])

cluster ion implantation process. Taking these into account, the cluster ion implantation technique could enable us to have higher productivity and a broader process window.

For 32nm node, in order to meet the USJ requirements, not only new implantation techniques will be used, but also new anneal techniques (laser anneal and flash anneal) will be employed. These anneals are also called millisecond anneals, and their peak temperature is in the range 1300 ~ 1400℃, which is Si sub-melt temperature. Although they are called millisecond anneals, the peak temperature dwell time could range from milliseconds to nanoseconds. For 32nm node, the peak temperature dwell time is likely to be in the microsecond range. By using millisecond anneal techniques, dopant diffusion will be suppressed, but the activation rate will be dramatically increased. Millisecond anneal compatibility with an MHK gate stack was also studied[51]. Good MHK gate stack reliability was achieved, but the carrier mobility was still degraded by the anneal due to the interface defects generation. The degradation could be recovered by an optimized post-metallization anneal. The compatibility of millisecond anneal with SiGe S/D was also studied[53]. It was reported that relaxation-free SiGe after millisecond anneal was preserved, and a 10% drive current increment was achieved.

Si recess (Si loss) under S/D extensions can change the USJ profile and, thus, increase the extension resistance and degrade the device performance. The recess could occur during some process steps, such as wet or dry cleans, photoresist strip, activation anneals, and so on. For 32nm node, Si loss per step should be smaller than 0.03nm[52]. In order to achieve this, low temperature and low concentration chemicals should be used for wet cleans, and photoresist stripping with non-fluorine plasma chemistries is also necessary.

7 New metallization technology for MOL and BEOL

Interconnect technology has played an important role for continuous scaling of CMOS-based microelectronics products. For 32nm node, the contact stud diameter is expected to be between 40 and 50nm[54]. In this case, if W is still used as the filling material for contact studs, the stud resistance is expected to double the number of 45nm node. Replacing W with copper is believed to reduce the mean resistance. However, compared with other external device resistances, the impact of switching to Cu is expected to be less than 3%[54]. Electroplated rhodium (Rh) is proposed as an-

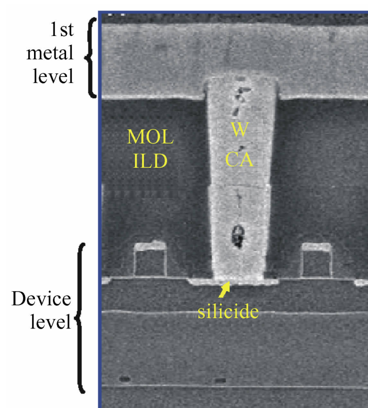

Fig. 22 Cross-sectional SEM picture for W plug (From Ref. [56])

other solution for replacing W contact studs[55]. At 32nm node, it is difficult for the traditional CVD W filling process to fill the vias with high AR (diameter ~ 40nm, height ~ 240nm) because center voids and seams will form, as shown in Fig. 22. Electroplating of Rh has similar super-conformal filling capability for nanometer sized features with high AR as Cu. Figure 23 is the TEM picture for void-free electroplated Rh filled vias. The successful fill for this aggressive dimension demonstrates the extendibility of the Rh fill process to 32nm node or even beyond. Another main advantage of using Rh as a fill material is that Rh does not diffuse in Si, and therefore, no thick liner is needed. PVD TaN or Ti with an ALD ruthenium (Ru) seed layer could be used as the liner[55]. Although Rh has a higher resistance than Cu, but a lower resistance than W, the low resistance of the Cu filled contact stud will be offset by the necessarily thick and robust liner stack. Using optimized annealing conditions, the contact stud resistance using Rh could be reduced to 1.5 or 3X lower[55] than current CVD W contact studs. However, Rh is a noble metal with exceptional chemical resistance. Thus, the development of the CMP process, with aggressive abrasives

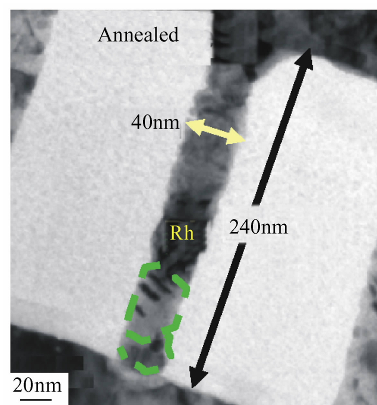

Fig. 23 Cross-sectional TEM picture for electroplated Rh plug (From Ref. [56])

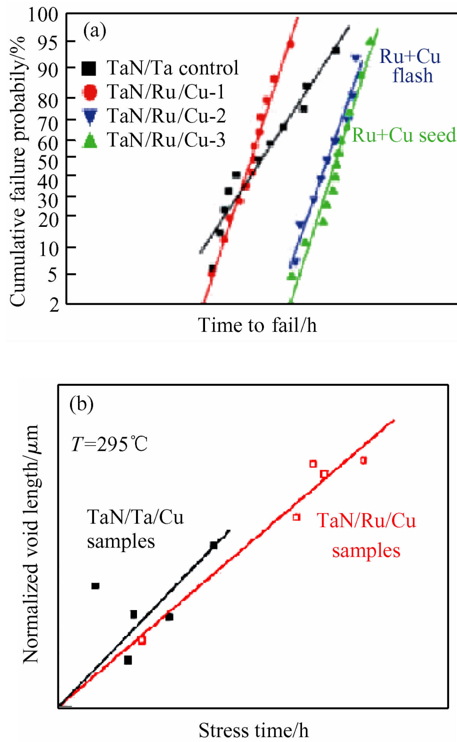

Fig.24 (a) EM test data for baseline liner and Ru liner ones; (b) Void growth rate comparison for different liners (From Ref.[57])

and oxidizers, is the key challenge for the integration of Rh.

Electroplated Cu with a PVD TaN/Ta liner and a Cu seed layer has been successfully implemented in many BEOL schemes for years. With the scaling of the technology nodes, scaling of the thickness of the liner and the seed layer is also required to maximize Cu volume and reduce corresponding resistance. At this point, Ru was proposed as a promising substitute liner material for BEOL Cu integration[56,57]. Cu has good adhesion to Ru, and the Ru-Cu system is thermally stable and reported to be immiscible and also has lower resistance. It was reported that using an ALD TaN/ ALD Ru liner (followed by a PVD Cu seed layer or other options), thinner liner thickness is needed and more room for Cu could be spared, and a 4% performance gain is achieved[56]. Electromigration (EM) and stress-induced voids (SIV) test results for the PVD TaN/ CVD Ru liner, compared with PVD TaN/Ta baseline liner control, are shown in Figs. 24 (a) and 24(b), which indicate that good reliability for the Ru liner could be achieved.

At 32nm node, due to the increase of the current density, narrow Cu line reliability performance will become a key concern[58]. In order to enhance the EM performance, self-aligned barriers (SAB) on top of Cu lines are proposed as an effective solution for 32nm node or even beyond[59~61]. Cu line surface

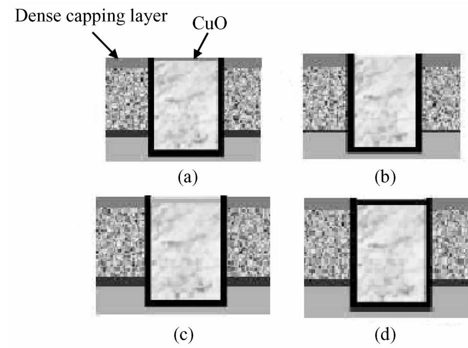

Fig.25 (a) Post-CMP Cu metal level; (b) CuO removal and surface activation; (c) Si riched precursor; (d) NH_3 plasma treatment (From Ref.[60])

treatments and selective deposition on top of Cu lines are the two distinguished SAB approaches. The Cu surface treatment or modification approach (referred as the CuSiN process) is developed for its direct compatibility with existing PE-CVD processes. The integration process for this approach is shown in Fig. 25[59]. First, native Cu oxide is removed and the surface is activated using H_2-based cleaning plasma, as shown in Fig. 25(b). Second, Cu silicidation using Si-based precursor is performed, as shown in Fig. 25(c). Finally, nitridation with NH_3 plasma is done to achieve the CuSiN SAB, as shown in Fig. 25(d). This final nitridation step is very important, since it could not only reduce further Si diffusion into copper under thermal or electrical stress, but also increase the CuSiN barrier efficiency against Cu diffusion and oxidation. There are two ways to realize the approach of SAB selective deposition on top of the Cu line, and the general process flow for these two ways is shown in Fig. 26[59]. One way is the W-CVD process, for which deposition selectivity is the main concern. The other way is to use Co-based ternary compound alloys (e. g. CoWP, CoWB, CoWPB, etc.). This electroless deposition process also consists of Cu oxide removal and Cu surface activation, as shown in Fig. 26(a). This surface activation is done by a seed solution immersion. This solution is Pd rich. Pd will undergo an exchange reaction with oxidized Cu sites, thus more Cu is recessed. After seeding with Pd, the CoWP capping layer could be applied by an electroless immersion

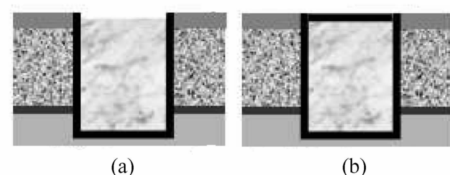

Fig.26 (a) CuO removal and surface activation; (b) SAB deposition (From Ref.[60])

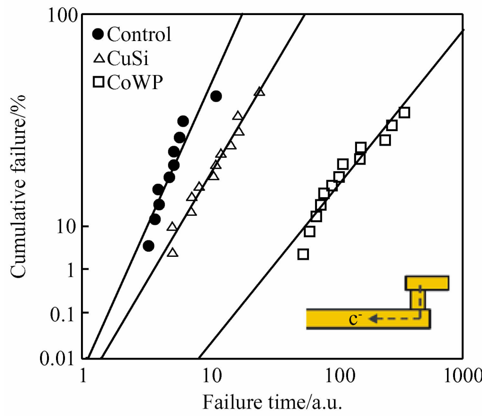

Fig. 27 EM test for different capping layers (From Ref. [62])

bath. CoWP nodules will nucleate on the Pd seed and the barrier grows auto-catalytically. The deposition rate, compound composition, and uniformity are dependent on solution parameters. Using these SAB capping layers, the adhesion between the interface and Cu line is strengthened, thus the EM performance is enhanced, as shown in Fig. 27[61]. However, the stress-migration (SM) could act differently, as shown in Fig. 28[61]. From this figure, it can be seen that for the $CuSi_x$ capping layer, the SM failure rate could be as high as 18%, while the CoWP capping layer suppresses the SM failure rate. Thus, optimization of the SAB capping layer chemistry is necessary.

8 Ultra low-*k* integration

Device dimension shrinkage at each generation will raise many challenges, such as interconnect RC delay and power consumption. Device performance requirements continue to drive the needs for new BEOL materials and processes. For 32nm CMOS manufacturing in BEOL, the first challenge is how to fabricate the ultra low k ($k<2.5$) dielectrics and lower k dielectric barriers[62]. The second is how to make thinner low resistivity Cu barriers and more conformal seeds with faster gap-fill plating chemistries. The third challenge is how to cope with low pressure and shear force CMP with low defect slurries and post-CMP cleans. Regarding these issues, some R&D work has been carried on in the following areas: mechanically robust porous ELK materials; low k dielectric barriers or self-aligned barriers; iALD barriers and seeds; faster and void-free Cu fills and plates on highly resistive seeds; and low damage CMPs with low defectivity post-CMP cleaning. All these process developments are based on process uniformity and repeatability to ensure the future scaling for manufacturing[63].

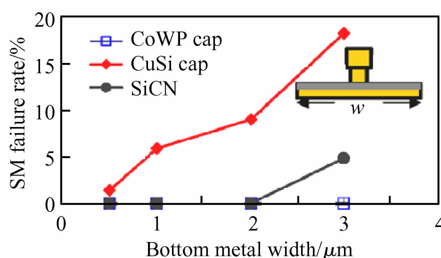

Fig. 28 SM test for different capping layers (From Ref. [62])

Table 3 ULK ($k=2.3$ and $k=2.5$) properties (From Ref. [67])

	$k=2.3$	$k=2.5$
k@100kHz	2.3	2.5
L_c@1MV/cm(A/cm^2)	7×10^{-10}	3×10^{-9}
BV/(MV/cm)	5.9	6.0
RI@632.8nm	1.323	1.341
$E_{reduced}$/GPa	3.5	6
Hardness/GPa	0.7	1
Stress/MPa	37	52
Porosity/%	30	26
Pore radius/nm	1.2	1.2

For 32nm node, the bulk dielectric constant k should be in the range of 2.1 ~ 2.4, and the effective dielectric constant k_{eff} should be in the range of 2.5~2.8 after integration[1]. There are many ways to produce such ultra low-k materials. One way is to use porous organic spin-on dielectrics (SOD)[64,65]. But the main approach is to use plasma enhanced chemical vapor deposition (PECVD) porogen SiOCH[66,67]. The first step of this approach is to co-deposit two precursors in plasma; a matrix precursor (to generate the SiOCH skeleton) and a porogen precursor (to generate an organic sacrificial phase). The second step is UV assisted post-treatment in a separate chamber to remove the sacrificial inclusions and generate the porosity. The porosity can be tuned by the porogen amount loaded into the co-deposited hybrid film. The UV post-treatment (or cure) needs to be optimized to obtain good balance between porosity (limited film shrinkage) and skeleton mechanical properties. UV cure could increase Si-O chemical bonds and reduce Si-H/Si-CH_3 bonds in low-k film[68,69]. Without degradation of low k value, the film elastic modulus (E) will increase with the increment of Si—O bonds. Actually, not only a UV cure was proposed for low-k film property improvement, but laser spike anneal and electron beam (EB) cure[71] were also proposed as potential solutions. The reason for selecting UV cure is due to its simplicity, less damage, and lower cost. Properties of two ULK ($k=2.5$ and $k=2.3$) made from this UV cure approach are summarized in Table 3.

There are many integration schemes. One of them is the single damascene (SD) scheme, but the

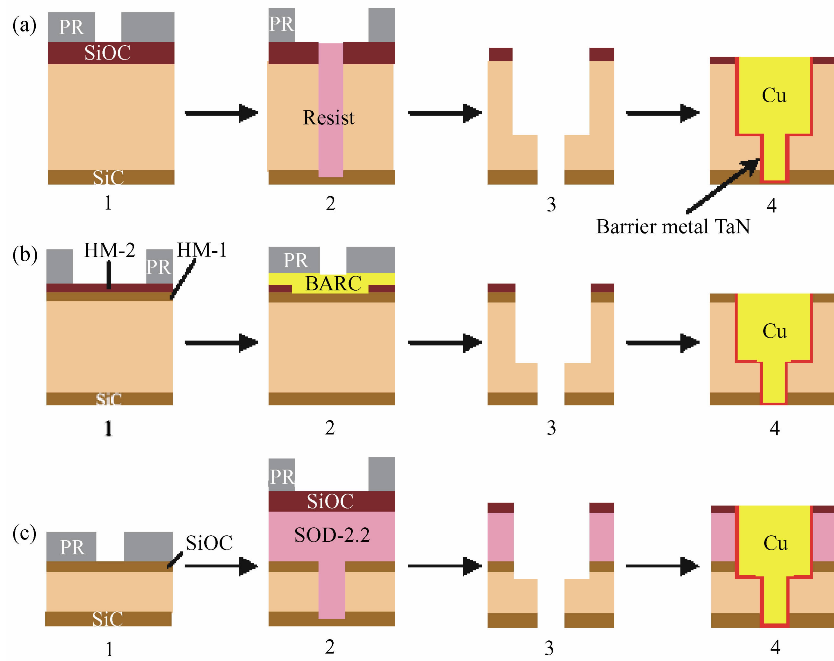

Fig. 29 Different DD integration schemes (From Ref. [66]) (a) Typical via first flow; (b) Typical trench first flow; (c) Flow for via-fill hybrid ULK structure

most widely used is the dual damascene (DD) scheme due to the advantages of fewer steps and lower cost. Even for DD, there exist many different process sequences, including via first, trench first, and so on[65]. A typical via first flow is shown in Fig. 29(a), where via forms before the trench. A typical trench first flow with hard mask approach is shown in Fig. 29(b). Using this hard mask approach, ULK resist poisoning and PR ashing damage can be minimized. In this approach, the ULK could be either uniform or hybrid. A hybrid ULK structure composed of Polyarylene ether (PAr) and SiOC was reported[64,72]. The hard masks used in this approach could be double or even triple layers, and here, in the picture, only flow with double-layer hard masks is shown. Another approach with a via-filled hybrid ULK structure is shown in Fig. 29 (c). In this approach, via in dense low-k forms first, and the trench forms after SOD deposition.

During ULK patterning or metal barrier pre-cleaning, properties of the sidewall and top ULK in the range of a few nanometers from the surface will be modified. Depending on low-k properties, some modification can have carbon depleted at the surface, as shown in Fig. 30(a)[65]. Thus, the surface will become hydrophilic and ready to absorb moisture[58]. This could lead to k value and reliability degradation. But some modification could induce positive effects on low-k surface, making it denser, and good reliability could be achieved. This kind of modification can serve as a pore sealing strategy. By optimizing this sealing strategy, carbon depletion at the surface could be negligible, as shown in Fig. 30(b). This sealing modification can be realized at the last step of the plasma process, by inserting some special gas, such as CH_4, NH_3 and He/H_2[73]. Electron energy loss spectroscopy (EELS) analysis is used to study the sealing

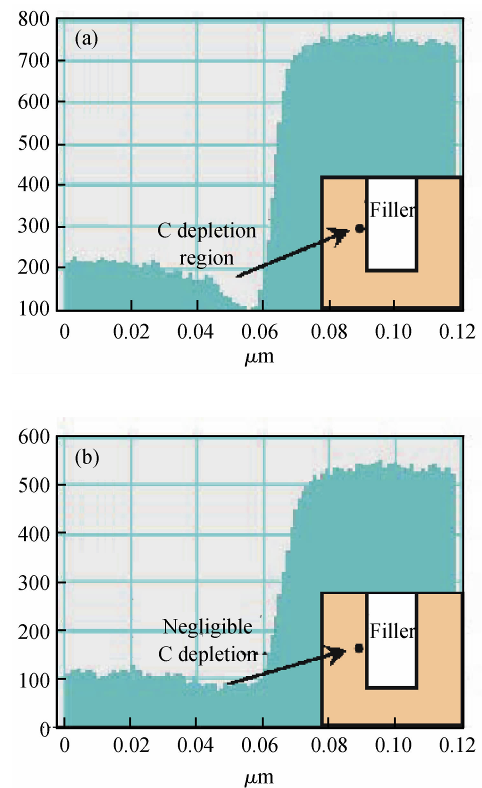

Fig. 30 TEM EELS of carbon depletion at sidewall (From Ref. [66]) (a) Modification without pore sealing; (b) With pore sealing

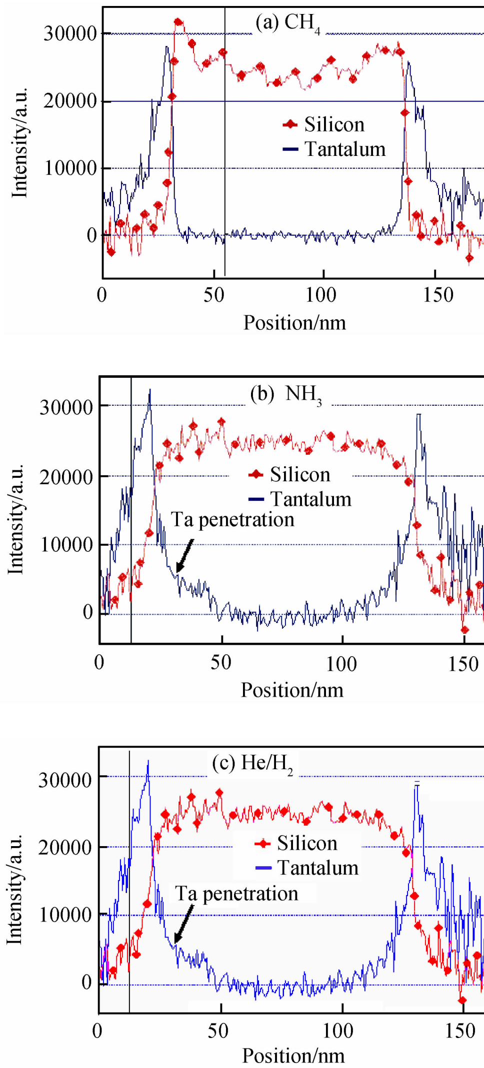

Fig. 31 EELS analysis of pore sealing effects for different chemistries (From Ref. [74])

effects, as shown in Fig. 31. Tantalum (Ta) penetration into low-k sidewalls is found for He/H_2 sealing chemistry, but for CH_4 and NH_3 sealing, no penetration is found. The reliability results for CH_4 and NH_3 sealing are also compared, as shown in Fig. 32. It is found that CH_4 chemistry has better sealing effects since it can increase carbon concentration in low-k and preserve the hydrophobic property of low-k.

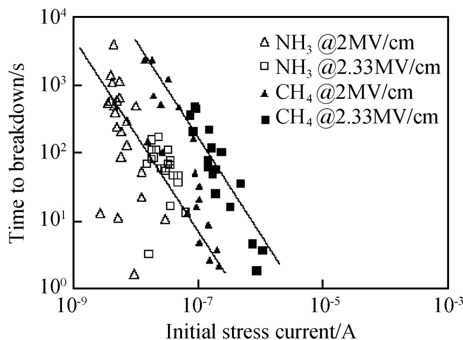

Fig. 32 Reliability results comparison for NH_3 and CH_4 (From Ref. [74])

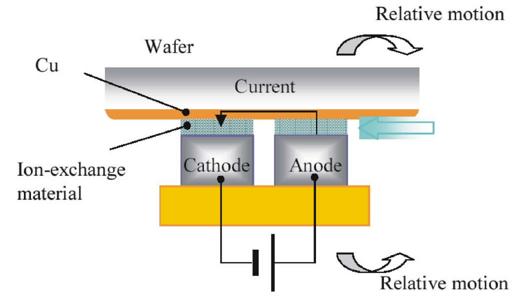

Fig. 33 Mechanism for ECMP (From Ref. [76])

Some other plasma pore sealing technology using organosilane (OPT) was also reported, and perhaps better results can be expected[72].

For 32nm node, the requirements for Cu chemical mechanical polishing process (CMP) on fragile ULK become more stringent. In order to meet the stringent design rules and compatibility with lithography depth of focus (DOF) requirement, better topography with a lower pressure process is necessary. The traditional slurry approach will become more and more complex in order to achieve a total planarization with dishing and erosion below 20nm. An electrochemical mechanical polishing process (ECMP) could be introduced at this point as a potential solution[74]. ECMP uses an applied electric field, instead of an oxidizer to change metal copper to copper ions, as shown in Fig. 33[75]. These ions will react with the agents on the electrolyte, and a passivation layer is created, which could be removed at a very low pressure (0.3psi). Right now, the ECMP approach could provide direct polishing all the way to the low-k barrier. AFM scanning results for conventional slurry process (CONV) and ECMP on features with different pitches (S1～S8) are shown in Fig. 34. This figure shows that the erosion for ECMP is below 20nm on all features, while for CONV, the erosion could be as high as 60nm. By employing this new ECMP approach, not only can more process windows for the lithography process be achieved, but also lower sheet resistance can be achieved. However, there are still some other challenges for ECMP appli-

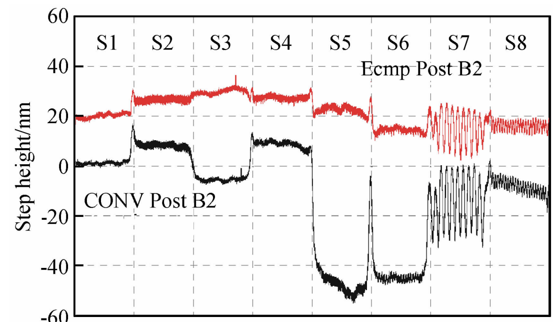

Fig. 34 AFM scanning results for CONV and ECMP at different features (From Ref. [75])

cations in 32nm node for both economic reasons and technical concerns about effective barrier layers removal.

9 Conclusion: challenges from 45nm to 32nm node

Overall, 32nm node will continue to enjoy some main technologies used in previous generations, such as different kinds of strain engineering, although some adjustments are needed. For liner stress technology, liners with higher stress and thinner thickness may be necessary due to the further pitch scaling of 32nm node. For STI and PMD SACVD applications, the HARP process will become much more difficult, thus, the etched profile and gap-fill process need to be optimized. For SiGe S/D, in order to maintain similar strain level in the channel as 45nm node, S/D profile optimization and Ge composition increment need extra care. Other stress enhancement technologies, such as DSB hybrid orientation bulk substrate, SiC S/D and Rev. e-SiGe, may be introduced for the first time at 32nm node. For USJ at 32nm node, millisecond anneal will continue to be effective for dopant activation, although more stringent requirements, such as peak temperature dwell time, MHK gate stack reliability, and so on, will be applied. The cluster ion implantation technique for USJ may also be newly introduced to 32nm node. For MOL and BEOL, new metallization technology, such as Rh and Ru, maybe employed for the first time, and new ULK materials will be also introduced.

The most fundamental changes made for the transition from 45nm to 32nm node are the application of immersion lithography double patterning and the wide introduction of MHK gate stacks. Although by 32nm node, lots of experience about immersion lithography will have been accumulated, many problems about double patterning, like overlay and CD related issues, will still be ahead. For 32nm low power technology, the possibility of using soft plasma gate oxide nitridation still exists due to the integration process simplicity as well as for economic purposes. It is believed that using MHK is inevitable for 32nm generic technology. For MHK gate stacks, until now, different stack configurations have been reported. However, no consensus on the gate stack material selections has been acheived. Other problems related to MHK gate stacks, such as V_t (or V_{fb}) roll-off, device reliability, carrier mobility degradation and so on, are still under serious study. It is believed that still more efforts need to be put into these two aspects.

From 45nm to 32nm technology node, there are still many other challenges. Silicide to silicon contact resistance is one of them. Due to the reduced contact length, this contact resistance will increase. It is expected that a 4% performance penalty will be suffered due to this interface resistance increase. One possible solution for this is to use some new materials, such as ErSi or YbSi[54]. Another possible solution is to use dopant segregation implantation technology[76]. Another concern for the contact stud is the gate to stud capacitance as the pitch scales. A possible solution is gate height reduction[54]. Although much work needs to be done for the transition from 45nm to 32nm node, our confidence in timely 32nm node delivery has become stronger. Actually, realization of a 32nm low power foundry technology with 0.15μm^2 6-T SRAM has been reported[77]. A Bright future for 32nm node is before us. We should grasp opportunities and meet the challenges. Aiming at factual demand and doing research work systematically, we are to acquire discovery and invention. The first 30 years of the 21st century is an import transition period, particularly in China.

References

[1] Chen X, Fang S, Gao W, et al. Stress proximity technique for performance improvement with dual stress liner at 45nm technology and beyond. Symp VLSI Tech, 2006:60

[2] Chen C H, Lee T L, Hou T H, et al. Stress memorization technique (SMT) by selectively strained-nitride capping for sub-65nm high-performance strained-Si device application. Symp VLSI Tech, 2004:56

[3] Peters L. 32nm marked by litho, transistor changes. Semiconductor International, January 2008

[4] Hand A. Double patterning wrings more from immersion lithography. Semiconductor International, February 2007

[5] Peters L. Double patterning leads race for 32nm. Semiconductor International, October 2007

[6] Sung C Y, Yin H, Ng H Y, et al. High performance CMOS bulk technology using direct silicon bond (DSB) mixed crystal orientation substrates. IEDM Tech Dig, 2005:225

[7] Yin H, Sung C Y, Ng H, et al. Direct silicon bonded (DSB) substrate solid phase epitaxy (SPE) integration scheme study for high performance bulk CMOS. IEDM Tech Dig, 2006:75

[8] Yin H, Sung C Y, Saenger K L, et al. Scalability of direct silicon bonded (DSB) technology for 32nm node and beyond. Symp VLSI Tech, 2007:222

[9] Chudzik M, Doris B, Mo R, et al. High-performance high-*k*/metal gates for 45nm CMOS and beyond with gate-first processing. Symp VLSI Tech, 2007:194

[10] Narayanan V, Paruchuri K, Bojarczuk N A, et al. Band-edge high-performance high-*k*/metal gate n-MOSFETs using cap layers containing group IIA and IIIB elements with gate-first processing for 45nm and beyond. Symp VLSI Tech, 2006:224

[11] Wen H C, Song S C, Park C S, et al. Gate first metal-aluminum-nitride PMOS electrodes for 32nm low standby power applications. Symp VLSI Tech, 2007:160

[12] Liu Y, Gluschenkov Ó, Li J, et al. Strained Si channel MOSFETs with embedded silicon carbon formed by solid phase epitax-

y. Symp VLSI Tech, 2007:44

[13] Yasutake N, Azuma A, Ishida T, et al. A high performance pMOSFET with two-step recessed SiGe-S/D stucture for 32nm node and beyond. ESSDERC, 2006:77

[14] Yasutake N, Azuma A, Ishida T, et al. Record-high performance 32nm node pMOSFET with advanced two-step recessed SiGe-S/D and stress liner technology. Symp VLSI Tech, 2007:48

[15] Hobbs C, Fonseca L, Dhandapani V, et al. Fermi level pinning at the polySi/metal oxide interface. Symp VLSI Tech, 2003:9

[16] Kadoshima M, Ogawa A, Ota H, et al. Two different mechanisms for determining effective work function on high-k. Symp VLSI Tech, 2006:226

[17] Iwamoto K, Ogawa A, Kamimuta Y, et al. Re-examination of flat-band voltage shift for high-k MOS devices. Symp VLSI Tech, 2007:70

[18] Kamimuta Y, Iwamoto K, Nunoshige Y, et al. Comprehensive study of VFB shift in high-k CMOS (dipole formation, Fermi-level pinning and oxygen vacancy effect). IEDM Tech Dig, 2007:341

[19] Song S C, Park C S, Price J, et al. Mechanism of V_{fb} roll-off with high work function metal gate and low temperature oxygen incorporation to achieve PMOS band edge work function. IEDM Tech Dig, 2007:337

[20] Wang Yangyuan, Kang Jinfeng. Development and challenges of lithography for ULSI. Chinese Journal of Semiconductors, 2002, 23(3): 225 (in Chinese)[王阳元,康晋锋. 硅集成电路光刻技术的发展与挑战. 半导体学报,2002, 23(3): 225]

[21] Yang M, Ieong M, Shi L, et al. High performance CMOS fabricated on hybrid substrate with different crystal orientations. IEDM Tech Dig, 2003:453

[22] Arghavani R, Miner G, Agustin M, et al. High-k/metal gates prepare for high-volume manufacturing. Semiconductor International, November 2007

[23] Sivasubramani P, Kirsch P D, Huang J, et al. Aggressively scaled high-k gate dielectric with excellent performance and high temperature stability for 32nm and beyond. IEDM Tech Dig, 2007:543

[24] Ota H, Hirano A, Watanabe Y, et al. Intrinsic origin of electron mobility reduction in high-k MOSFETs (from remote phonon to bottom interface dipole scattering). IEDM Tech Dig, 2007:65

[25] Tateshita Y, Wang J, Nagano K, et al. High-performance and low-power CMOS device technologies featuring metal/high-k gate stacks with uniaxial strained silicon channels on (100) and (110) substrates. IEDM Tech Dig, 2006:63

[26] Sivasubramani P, Böscke T S, Huang J, et al. Dipole moment model explaining nFET V_t tuning utilizing La, Sc, Er, and Sr doped HfSiON dielectrics. Symp VLSI Tech, 2007:68

[27] Song S C, Zhang Z B, Hussain M M, et al. Highly manufacturable 45nm LSTP CMOSFETs using novel dual high-k and dual metal gate CMOS integration. Symp VLSI Tech, 2006:16

[28] Harris H R, Kalra P, Majhi P, et al. Band-engineered low PMOS VT with high-k/metal gates featured in a dual channel CMOS integration scheme. Symp VLSI Tech, 2007:154

[29] Mise N, Morooka T, Eimori T, et al. Single metal/dual high-k gate stack with low V_{th} and precise gate profile control for highly manufacturable aggressively scaled CMISFETs. IEDM Tech Dig, 2007:527

[30] Chudzik M, Doris B, Mo R, et al. High-performance high-k/metal gates for 45nm CMOS and beyond with gate-first processing. Symp VLSI Tech, 2007:194

[31] Kirsch P D, Quevedo-Lopez M A, Krishnan S A, et al. Band edge n-MOSFETs with high-k/metal gate stacks scaled to EOT=0.9nm with excellent carrier mobility and high temperature stability. IEDM Tech Dig, 2006:626

[32] Wang X P, Shen C, Li M F, et al. Dual metal gates with band-edge work functions on novel HfLaO high-k gate dielectric. Symp VLSI Tech, 2006:12

[33] Harris H R, Alshareef H, Wen H C, et al. Simplified manufacturable band edge metal gate solution for NMOS without a cap ping. IEDM Tech Dig, 2006:633

[34] Ge C H, Lin C C, Ko C H, et al. Process-strained Si CMOS technology featuring 3D strain engineering. IEDM Tech Dig, 2003:73

[35] Chan V, Rengarajan R , Rovedo N, et al. High speed 45nm gate length CMOSFETs integrated into a 90nm bulk technology incorporating strain engineering. IEDM Tech Dig, 2003:77

[36] Singh D V, Sleight J W, Hergenrother J M, et al. Stress memorization in high performance FDSOI devices with ultra-thin silicon channels and 25nm gate lengths. IEDM Tech Dig, 2005:511

[37] Yin H, Ren Z, Chen H, et al. Integration of local stress technique with strained-Si directly on insulator substrates. Symp VLSI Tech, 2006:12

[38] Yang H S, Malik R, Narasimha S, et al. Dual stress liner for high performance sub-45nm gate length SOI CMOS manufacturing. IEDM Tech Dig, 2004:1075

[39] Ching C, Whitesell H, Venkataraman S, et al. Improving electrical performance using SACVD oxide films. Semiconductor International, April 2008

[40] Cam C L, Guyader F, de Buttet C, et al. A low cost drive current enhancement technique using shallow trench isolation induced stress for 45-nm node. Symp VLSI Tech, 2006:82

[41] Okamoto S, Miyashita K, Yasutake N, et al. Study on high performance (110) PFETs with embedded SiGe. IEDM Tech Dig, 2007:277

[42] Pan J W, Liu P W, Chang T Y, et al. Mobility and strain effects on 〈110〉/(110) SiGe channel pMOSFETs for high current enhancement. IEDM Tech Dig, 2006:227

[43] Ohta H, Kim Y, Shimamune Y, et al. High performance 30nm gate bulk CMOS for 45nm node with Σ-shaped SiGe-SD. IEDM Tech Dig, 2005:247

[44] Luo Z, Rovedo N, Ong S, et al. High performance transistors featured in an aggressively scaled 45nm bulk CMOS technology. Symp VLSI Tech, 2007:16

[45] Kim Y S, Shimamune Y, Fukuda M, et al. Suppression of defect formation and their impact on short channel effects and drivability of pMOSFET with SiGe source/drain. IEDM Tech Dig, 2006:871

[46] Shimamune Y, Fukuda M, Koiizuka M, et al. Technology breakthrough of low temperature, low defect, and low cost SiGe selective epitaxial growth (L3 SiGe SEG) process for 45nm node and beyond. Symp VLSI Tech, 2007:116

[47] Han J P, Utomo H, Teo L W, et al. Novel enhanced stressor with graded embedded SiGe source/drain for high performance CMOS devices. IEDM Tech Dig, 2006:59

[48] Tsai C H, Lan B C, Lin Y H, et al. Unique ultra shallow junction scheme with conventional activation process. Symp VLSI Tech, 2006:188

[49] Chen Y W, Chang Y L, Chen Y C, et al. Formation of Ni(Pt) germanosilicide using a sacrificial Si cap layer. IITC, 2007:37

[50] Donaton R A, Chidambarrao D, Johnson J, et al. Design and fabrication of MOSFETs with a reverse embedded SiGe (Rev. e-SiGe) structure. IEDM Tech Dig, 2006:178

[51] Kalra P, Majhi P, Heh D, et al. Impact of flash annealing on performance and reliability of high-k/metal-gate MOSFETs for sub-45nm CMOS. IEDM Tech Dig, 2007:353

[52] Dejule R. Trends in shallow junction engineering. Semiconductor International, April 2008

[53] Yu M H, Li J H, Lin H H, et al. Relaxation-free strained SiGe with super anneal for 32nm high performance PMOS and beyond.

IEDM Tech Dig，2006：867
[54] Sleight J W，Lauer I，Dokumaci O，et al. Challenges and opportunities for high performance 32 nm CMOS technology. IEDM Tech Dig，2006：1
[55] Shao I，Cotte J M，Haran B，et al. An alternative low resistance MOL technology with electroplated rhodium as contact plugs for 32nm CMOS and beyond. IITC，2007：102
[56] Rossnagel S M，Wisnieff R，Edelstein D，et al. Interconnect issues post 45nm. IEDM Tech Dig，2005：1
[57] Yang C C，Spooner T，Ponoth S，et al. Physical，electrical，and reliability characterization of Ru for Cu interconnects. IITC，2006：187
[58] Arnal V，Farcy A，Aimadeddine M，et al. Materials and processes for high signal propagation performance and reliable 32nm node BEOL. IITC，2007：1
[59] Gosset L G，Chhun S，Guillan J，et al. Self aligned barrier approach：overview on process，module integration and interconnect performance improvement challenges. IITC，2006：84
[60] Kawahara N，Tagami M，Withers B，et al. A novel CoWP cap integration for porous low-W/Cu interconnects with NH_3 plasma treatment and low-*k* top（LKT）dielectric structure. IITC，2006：152
[61] Yeh M S，Chang H L，Shih C H，et al. Effect of Cu line capping process on stress migration reliability. IITC，2006：113
[62] Wang Yangyuan，Kang Jinfeng. Development of ULSI interconnect integration technology. Chinese Journal of Semiconductors，2002，23(11)：1121（in Chinese）[王阳元，康晋锋. 超深亚微米集成电路中的互连问题-低 *k* 介质与铜的互连集成技术. 半导体学报，2002，23(11)：1121]
[63] Havemann B. 5th International Novellus/Fudan Cu Interconnect Symposium，Shanghai，China，May 28—29，2008
[64] Usui T，Tsumura K，Nasu H，et al. High performance ultra low-*k*（$k=2.0/k_{\rm eff}=2.4$）/Cu dual-damascene interconnect technology with self-formed $MnSi_xO_y$ barrier layer for 32nm-node. IITC，2006：216
[65] Su Y N，Shieh J H，Tsai J S，et al. Integration of Cu and extra low-*k* dielectric（$k=2.5\sim2.2$）for 65/45/32nm generations. IEDM Tech Dig，2005：74
[66] Aimadeddine M，Jousseaume V，Arnal V，et al. Robust integration of an ULK SiOCH dielectric（$k=2.3$）for high performance 32nm node BEOL. IITC，2007：175
[67] Ikeda A，Travaly Y，Humbert A，et al. Low-*k* properties and integration processes enabling reliable interconnect scaling to the 32nm technology node. IITC，2006：42
[68] Yoneda K，Kato M，Kondo S，et al. Impacts of UV cure for reliable porous PECVD SiOC integration. IITC，2005：84
[69] Nakao S I，Ushio J，Ohno T，et al. UV/EB cure mechanism for porous PECVD/SOD low-*k* SiCOH materials. IITC，2006：66
[70] Volksen W，Dubois G，Kellock A，et al. Laser spike annealing：a novel post-porosity treatment for significant toughening of low-*k* organosilicates. IITC，2006：146
[71] Miyajima H，Fujita K，Nakata R，et al. The application of simultaneous ebeam cure method for 65nm node Cu low-*k* technology with hybrid（PAE/MSX）structure. IITC，2004：222
[72] Arakawa S，Mizuno I，Ohoka Y，et al. Breakthrough integration of 32nm-node Cu ultra low-*k* SiOC（$k=2.0$）interconnects by using advanced pore-sealing and low-*k* hard mask technologies. IITC，2006：210
[73] Aimadeddine M，Arnal V，Roy D，et al. Effect of CH_4 plasma on porous dielectric modification & pore sealing for advanced interconnect technology nodes. IITC，2006：81
[74] Mellier M，Berger T，Duru R，et al. Full copper electrochemical mechanical planarization（Ecmp）as a technology enabler for the 45 and 32nm nodes. IITC，2007：70
[75] Wada Y，Noji I，Kobata I，et al. The enabling solution of Cu/low-*k* planarization technology. IITC，2005：57
[76] Onoda H，Miyashita K，Nakayama T，et al. 0.7V SRAM technology with stress-enhanced dopant segregated Schottky（DSS）source/drain transistors for 32nm node. Symp VLSI Tech，2007：76
[77] Wu S Y，Chou C W，Lin C Y，et al. A 32nm CMOS low power SoC platform technology for foundry applications with functional high density SRAM. IEDM Tech Dig，2007：263

32nm CMOS 工艺技术挑战

吴汉明[1,†] 王国华[1] 黄 如[2] 王阳元[2]

（1 中芯国际半导体制造公司，北京 100176）
（2 北京大学，北京 100871）

摘要： 根据国际半导体技术发展蓝图(international technology roadmap for semiconductor，ITRS)，CMOS技术将于2009年进入32nm技术节点.然而，在CMOS逻辑器件从45nm向32nm节点按比例缩小的过程中却遇到了很多难题.为了跨越尺寸缩小所带来的这些障碍，要求把最先进的工艺技术整合到产品制造过程中.文中总结并讨论了可能被引入到32nm节点的新的技术应用，涉及如下几个方面：浸入式光刻的延伸技术、迁移率增强衬底技术、金属栅/高介电常数栅介质(metal/high-*k*，MHK)栅结构、超浅结(ultra-shallow junction，USJ)以及其他应变增强工程的方法，包括应力邻近效应(stress proximity effect，SPT)、双重应力衬里技术(dual stress liner，DSL)、应变记忆技术(stress memorization technique，SMT)、STI和PMD的高深宽比工艺(high aspect ratio process，HARP)、采用选择外延生长(selective epitaxial growth，SEG)的嵌入SiGe(pFET)和SiC(nFET)源漏技术、中端(middle of line，MOL)和后端工艺(back-end of line，BEOL)中的金属化以及超低 *k* 介质(ultra low-*k*，ULK)集成等问题.

关键词： CMOS技术；32nm技术节点；迁移率增强；金属栅/高 *k* 栅介质；超低 *k* 介质
EEACC： 2570
中图分类号： TN405 **文献标识码：** A **文章编号：** 0253-4177(2008)09-1637-17

† 通信作者.Email：hanming_wu@smics.com
2008-06-24 收到，2008-07-05 定稿
©2008 中国电子学会