MAY 2008

Solid State TECHNOLOGY®

THE INTERNATIONAL MAGAZINE FOR SEMICONDUCTOR MANUFACTURING

32nm Node in Reach by 2009 p. 26

- Silicon Nanowire CMOS p. 34
- SiP Integration of Passives with TSVs p. 38
- Unlocking 45nm OPC p. 48

PennWell®

www.solid-state.com

Despite engineering and cost challenges, 32nm node

IC manufacturing within reach

EXECUTIVE OVERVIEW There are no insurmountable obstacles foreseen for scaling high-volume manufacturing to the 32nm node for logic in 2009. Though this transition will primarily be an extension of 45nm technology, significant challenges remain to be resolved. No disruptive technologies such as new substrates, biaxial stress, or double gates will be needed; instead, 32nm will continue to use silicon, uniaxial stress, and planar transistors. The equipment industry can therefore focus on material innovations and manufacturability, particularly in the transistor region where most of the significant changes are occurring. New materials and integration methods will be needed to achieve a reliable and manufacturable 32nm node technology.

The fourth generation of uniaxial stress inducing films on planar transistors (first generation introduced at the 90nm node) will debut with the 32nm logic technology node. Nitride stressors for both NMOS and PMOS devices should reach >3GPa compressive and over >2GPa tensile stress for blanket film depositions.

Unique integration techniques would enable most of this stress to be transferred to the channel of the device [1,2]. Embedded silicon germanium in closer proximity to the channel with >30% Ge concentration is required to meet the performance needs of the PMOS devices [3,4].

The NMOS device most likely would not use embedded SiC processing, but employ recessed and raised silicon source/drain to overcome the external resistance issues resulting from the narrow dimensions of this technology node [5]. Strain needs to be integrated with the new high-*k* and metal gate stacks.

This transition requires the introduction of a hafnium-based dielectric, replacement of a P^+ poly with a metal of compatible work function and replacement of N^+ poly with a metal of compatible work function.

Introducing strain in the silicon for performance gains is based on carrier mobility and the direction of the current flow in the channel of the device. The "longitudinal" and "transverse" components of the effective mass are modulated by the degree and method of inducing strain on the device channel:

$$m^* = \left[\frac{1}{6} \left(\frac{2}{m_l}\right) + \left(\frac{4}{m_t}\right) \right]^{-1} \tag{1}$$

Reza Arghavani, Hichem M'saad, Applied Materials, Santa Clara, California USA

Since electrons and holes respond differently to this strain, this presents different choices on how to best enhance the electron and hole mobility for the 32nm node generation.

With the application of strain to silicon, the conduction band splits (degeneracy is removed). The difference in energy levels causes the repopulation of electrons to the lower-energy sub-bands with low in-plane effective mass (in the case of tensile strain). Additionally, the back scattering of electrons is reduced giving a higher electron velocity, and electron lifetime increases due to this reduction of back scattering. As a result of these two factors, electron mobility is significantly improved. Also upon the application of strain, the valence band of silicon warps and degeneracy between light-hole and heavy-hole is removed. The majority of holes will now occupy the light-hole band (in the case of compressive strain). As a result of this factor, effective mass is reduced, thus hole mobility increases.

Neither electron nor hole mobilities saturate with continuous increase of stress in the channel of the device up to the 32nm technology node [2]. Today, dielectrics with >1.5GPa tensile and >2.8GPa of compressive stress are routinely used at several steps in the process flow in high-volume manufacturing [1]. Combined with other stressors, the dielectrics introduce >1GPa of uniaxial compressive and tensile stress in the channel of MOSFETs.

Dielectrics with >2GPa tensile and >3.5GPa of compressive stress will be introduced in the 32nm node and will continue to insert a significant amount of uniaxial stress to the channel of an NMOS and PMOS device, respectively. Stress films are used in a variety of processes such as STI, contact etch stop, pre-metal dielectrics, removable films for stress memorization techniques, spacers, and even salicidation (**Table 1**).

State-of-the-art uniaxial strained silicon technologies in high-volume manufacturing include significant breakthroughs in toolsets, unit-processes, and unique integration approaches. For 32nm, two integration methods are expected to induce strain in the channel of the planar devices: dual stress liner (DSL), which integrates compressive and tensile nitride films [1] combined with tensile stress induced through STI, PMD, and stress memorization, and embedded epitaxial films in the source/drain [2,3]. The epitaxial films include epitaxial SiGe processing with >30% Ge concentration in closer proximity to the channel for PMOS, while tensile stress and a recessed source/drain with raised silicon will be implemented for

Reprinted with revisions to format, from the May 2008 edition of **SOLID STATE TECHNOLOGY**
Copyright 2008 by PennWell Corporation

TABLE 1. Roadmap of stressors				
Lithography node	**45nm**		**32nm**	
Device	Planar		Planar	
Gate length	35-45nm		25-30nm	
Lithography method	193nm immersion lithography or 193nm dry lithography with double exposure/patterning		193nm immersion lithography with double exposure/patterning	
Poly ½ pitch	80nm		55nm	
STI width, AR	120nm, 2.2		110nm, 2.2	
Contact ½ pitch	75nm		50-60nm	
Spacer width	20nm		10-15nm	
Polarity	P	N	P	N
Stress-inducing PECVD nitride Blanket film stress Stress transferred to the channel Mobility gain	Ion bombardment -3.2GPa -440MPa 1.35x	Photon-assisted 1.7GPa 240MPa 1.2x	Ion bombardment -3.5GPa -500MPa 1.5x	In development 2.2GPa 280MPa 1.3x
Source/drain heteroepitaxy Ge% (PMOS), C% (NMOS) Stress transferred to the channel Mobility gain	30% Ge -1.5GPa 2.5x	N/A	35% Ge -1.7GPa 2.7x	N/A
Tensile oxide (STI/PMD) Blanket film stress Stress transferred to the channel Mobility gain	N/A	200MPa 100MPa 1.1x	N/A	200MPa 100MPa 1.1x
Additivity (all stressors) Stress transferred to the channel Mobility gain	-1.9GPa 3.4x	340MPa 1.3x	>|-3.0GPa| ~4.0x	380MPa 1.4x
Other stressors	N/A	Stress memorization, sidewall spacer, source/drain silicide	N/A	Stress memorization, sidewall spacer, source/drain silicide

NMOS. **Figure 1** shows results for NMOS devices with high tensile nitride. **Figure 2** shows a cross section SEM of a PMOS device with both high compressive nitride and embedded SiGe (eSiGe), illustrating the additivity concept [6].

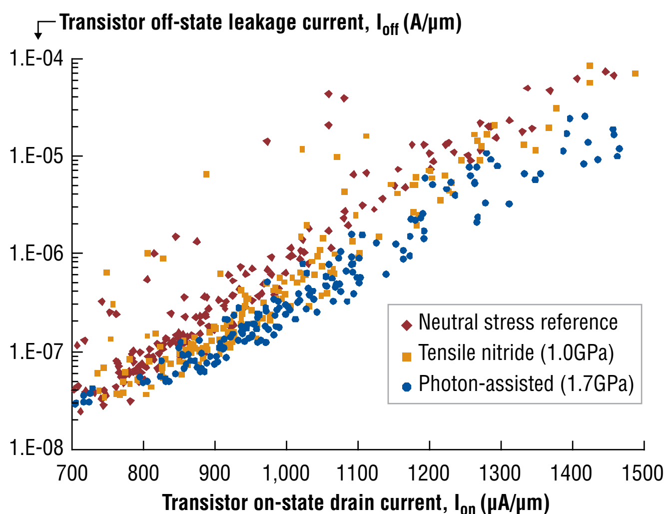

Figure 1. NMOS transistor on-state current increases further with photon-assisted stress-inducing nitride layer.

Transistor parasitic resistances

Current material processes will transfer enough stress to the channel of a MOSFET device to increase electron and hole mobility and meet the 32nm node logic performance goals. The improved mobility gain will transfer into enhanced device performance, given that parasitic external resistances are minimized [4,5]. These external resistances are the result of junctions, salicidation, and contact processing. A variety of processes such as millisecond-annealing are being offered to address the external resistance.

For the 32nm node, the poly-silicon critical dimension (CD) will be ~30nm with a contacted-gate-pitch of ~120nm. Initial calculations and modeling show that the external resistance of the MOS device would match that of the channel resistance [5]. **Figure 3** shows the components of the total external resistance.

External resistance reduction starts with ultra-shallow junctions. Minimizing both the Schottky barrier height between silicide and silicon, as well as the contact resistance also help. PVD Ni(Pt) for metal deposition together with in-situ clean can significantly reduce the PMOS contact resistance. Low-resistance tungsten contact plugs are deposited with an integrated PVD/CVD liner/barrier followed by an integrated ALD/CVD tungsten nucleation and fill to extend tungsten contacts to 32nm node. Advanced anneal and raised source/drain epitaxial deposition techniques ensure ultra shallow junctions (**Table 2**).

TABLE 2. External resistance components and unit processes

$R_{External}$ Component	Process steps	45nm process		32nm process	
		PMOS	NMOS	PMOS	NMOS
R_C Contact plug resistance	Pre-clean Liner/barrier Nucleation Fill	In-situ reactive dry PVD Ti/ MOCVD TiN ALD W CVD W	In-situ reactive dry PVD Ti/ MOCVD TiN ALD W CVD W	In-situ reactive dry PVD Ti/Low R_C TiN ALD W CVD W	In-situ reactive dry PVD Ti/ Low R_C TiN ALD W CVD W
R_{CSD} Silicide to S/D contact resistance	Pre-clean Metal deposition Epi	In-situ reactive dry NiPt (5%) eSiGe (30%)	In-situ reactive dry NiPt (5%)	In-situ reactive dry NiPt (10%) eSiGe (35%)	In-situ reactive dry NiPt (10%)
$R_{S/D}$ Deep S/D junction	Epi	eSiGe (30%)	N/A	eSiGe (35%)	Raised Si S/D
R_{SDE} S/D extension	Implantation Activation anneal Epi	Beamline Spike anneal eSiGe (30%)	Beamline Spike anneal	Plasma immersion Millisecond anneal eSiGe (35%)	Plasma immersion Millisecond anneal Raised Si S/D
R_{Ov} Accumulation under gate overlap	Implantation Epi	Halo/tip implant eSiGe (30%)	Halo/tip implant	Halo/tip implant eSiGe (35%)	Halo/tip implant

High-*k* and metal-gates

The industry is diverging in two different unique methods of integrating the gate stack for the first time [3,7]. One method features two different dielectrics and one electrode, the so-called dipole field method through a gate first integration. The other method uses two different metal work functions and one dielectric through a damascene integration.

The 90nm node technology saw the debut of the final scaled gate oxynitride at 1.2nm. Although this dielectric had a gate leakage >30A/cm^2 at the desired operating voltages, the 65nm logic node continued with the identical gate stack [8]. To increase the thickness of the gate dielectric (and reduce gate leakage) with no capacitance penalty, oxynitride must be replaced with a higher dielectric constant material [8,9]. This is because the equivalent oxide thickness (EOT) of a dielectric is inversely proportional to its dielectric constant.

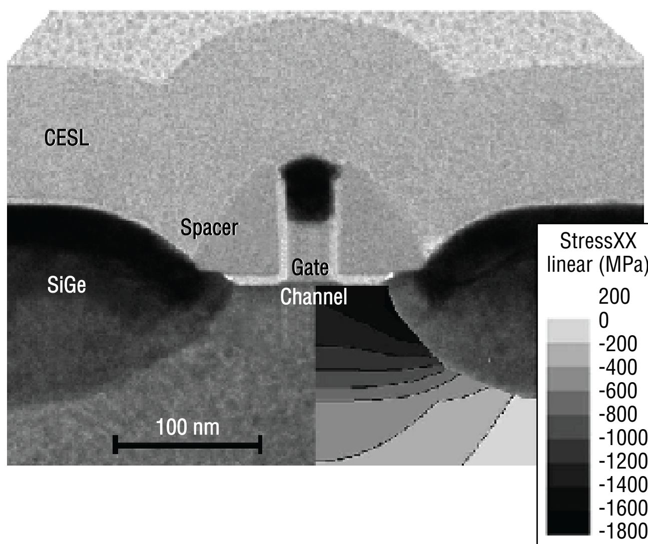

Figure 2. For a 90nm node PMOS transistor with embedded SiGe stressor in the source/drain, and high compressive stress nitride liner, on-state current improvement from the two stressors individually is 21%, while both provide >55% improvement [6].

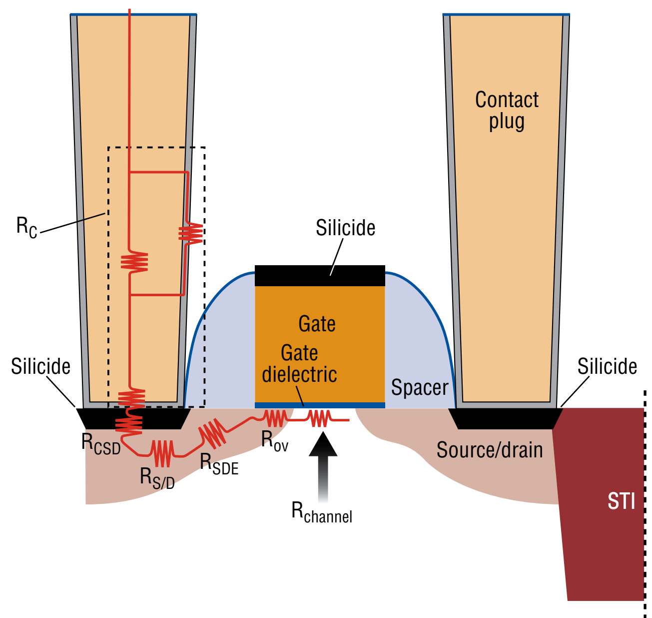

Figure 3. Components of external resistance in an MOS device.

High-*k*/metal-gate implementation requires three new materials (**Fig. 4**). The damascene integration requires a material with high dielectric constant for the dielectric, a metal whose work function is the same as poly NMOS (4.2eV) and another metal whose work function is the same as poly PMOS (5.2eV) [9]. The dipole field method requires a dielectric that can tune the V_{th} the same as poly NMOS [10], a dielectric that can tune the V_{th} the same as poly PMOS [11], and a mid-gap metal.

Atomically smooth interfaces are essential in high-*k* and metal-gate processing as most metal-based high-*k* dielectric materials such

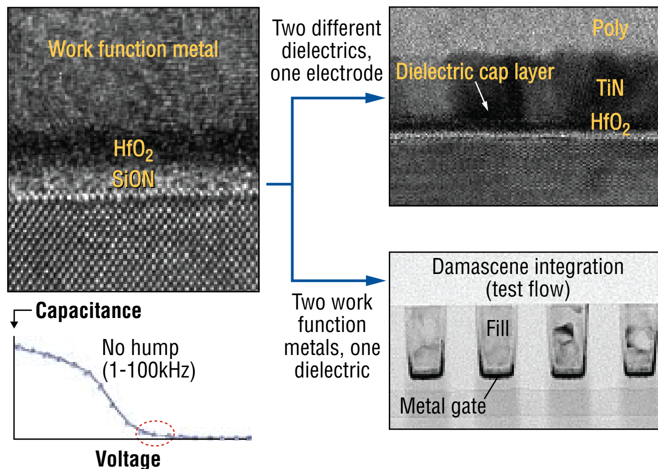

Figure 4. Two methods of high-k/metal-gate integration include: dipole field using two different dielectrics and one electrode through a gate-first integration, and two different metal work functions but one dielectric through a damascene integration.

as HfO_2 degrade the channel mobility [8,9,12]. The interface could be controlled by inserting an ultra-thin layer of oxynitride (SiON) prior to high-k material deposition. High mobility is expected from an integrated gate dielectric stack of SiON interface and HfO_2 high-k/metal-gate with an atomically smooth interface.

Interconnect challenges

Reduction of RC delay remains the key driver in lowering the effective interlayer dielectric constant for reduced capacitance between the wire lines [13,14]. While beneficial, the introduction of copper wires at the 130nm node to reduce the line resistivity in combination with SiOF interlayer dielectrics proved inadequate for significant signal delay improvement at 90nm node. Thus, the 90nm node debuted a SiCOH-based low-k dielectric with an effective bulk k~3.1, integrated together with copper wires via a dual-damascene process.

The components of interconnect delay defined by the product of line resistance and capacitance can be approximated as [14]:

$$\tau = RC = 2pk\varepsilon_0 \left(\frac{4L^2}{P^2} + \frac{L^2}{T^2} \right) \qquad (2)$$

where ρ in this equation is the metal resistivity (bulk copper and the barrier), k is the dielectric constant (effective dielectric constant from all the dielectrics between the copper lines), ε_0 is the permittivity in vacuum, L is the metal wire length, P is the metal wire pitch, and T is the metal wire thickness. Each successive generation of technology nodes has the option of increasing the metal aspect ratio to reduce the line resistance. This, however, increases the line capacitance and needs to be accounted for in the balance for overall RC delay.

At the 32nm node, Metal 1 and Metal 2 pitches are estimated to be near 100nm with aspect ratios of 1.7 and 2.3, respectively. To continue improving the signal delay compared to that of the 45nm node, an effective dielectric constant of ~2.5 (requiring a low-k bulk dielectric constant of k~2.2) is ideal. PECVD-based low-k films have already been developed with k~2.2 and are currently undergoing reliability and packaging tests.

Dielectrics, etch and cleans. PECVD-based low-k (bulk k~2.2) with UV-assisted curing technology has been developed for the 32nm technology node. Integration of this film with Cu at the 32nm technology node, however, faces a few key challenges (**Table 3**).

TABLE 3. Interconnect process flow to enable 32nm technology node

Cross-sectional schematic	Process steps	45nm process	32nm process
TiN metal HM Low k Low-k etch stop	• PECVD dielectric etch stop deposition • Low k deposition (+ dielectric cap) • (PVD metal) hard mask deposition	• PECVD SiCN (k=4.0) etch stop • PECVD p-SiOC (k=2.5) ULK with UV cure • PVD TiN metal hard mask	• PECVD SiCN (k=3.5) etch stop • PECVD p-SiOC (k=2.2) ULK with UV cure • PVD TiN metal hard mask
	• Hard mask pattern and etch • Dual Damascene dielectric etch and post-etch clean	• Track, 193nm immersion scanner (or dry 193nm with double patterning) • Metal etch, low damage ash • Low damage dielectric etch and post-etch wet clean	• Track, 193nm immersion litho with double exposure/patterning • Metal etch, low damage ash • Low damage dielectric etch and post-etch wet clean
	• PVD barrier and PVD Cu seed deposition • ECP Cu gapfill	• PVD Ta(N) barrier and Cu seed • ECP Cu gapfill	• Reduced thickness PVD Ta(N) barrier and Cu seed • ECP Cu gapfill
	• Cu and barrier CMP, post-CMP clean • PECVD dielectric barrier deposition	• 3-step slurry CMP (copper bulk, copper clear, barrier clear) • Post-CMP cleans, vapor dry • PECVD SiCN (k=4.0) Cu barrier	• 3-step slurry CMP (copper bulk, copper clear, barrier clear) • Post-CMP cleans, vapor dry • PECVD SiCN (k=3.5) Cu barrier

Dielectric barriers must be developed that are compatible with PECVD-based bulk dielectric of k~2.2. Ultra-thin carbon-based barriers with k~4.0 have been in production for 45nm node products. Further dielectric constant reduction is possible by optimizing the carbon content of the film and reduced film thicknesses would make such a family of films suitable for the 32nm node.

Etch process integration of the 32nm dielectric stack would start with the main trench etch, followed by an over-etch [15]. The over-etch stops at the SiCN-based etch stop layer over Cu lines. The main etch chemistry is a combination of argon and oxygen ions to sputter-etch the film. Flourine-based chemistries are then followed to dissociate the Si-O and CH_3 compounds of the low-k dielectric. To maintain CD uniformity and protect the sidewalls, the via etch processes use various polymer-containing chemistries with the etch terminating on top of the etch stop layer.

The remaining residues from the post-etch step are removed via an oxygen containing ash step. If not fully optimized at the 32nm node, this step can cause damage to a few atomic layers of the low-k dielectric resulting in reliability and line leakage issues and poor adhesion to the Cu barrier.

A wet chemistry post-ash step marks the final clean process. This step removes the remaining inorganic residues and the copper-oxide together with cleaning the surfaces, thus preparing the surface of Cu and sidewalls of the trench and via for the Cu barrier deposition. In general, the wet clean chemistries use fluoride-based solvents. **Figure 5** shows a cross-sectional image of a multilevel integrated Cu/low-k interconnect [16].

Metallization. The electrical resistivity (ρ) of Cu is partly determined by intrinsic components such as impurities and the internal grain structure of lines. The main contribution to line resistivity, however, arises from extrinsic geometry factors such as sidewall scattering. At the 32nm node, the resistivity of Cu increases dramatically due to sidewall scattering of electrons [17]. With constant line height, resistivity of ~2.5μΩ-cm at 100nm linewidth increases to nearly 2.9μΩ-cm at 40nm linewidth [17]. Thus the key concern is the extendibility of Ta/TaN/Cu [18] to the 32nm technology node with regard to filling the targeted <100nm pitch trenches without voids and also meeting the net resistivity targets. Void-free Ta/TaN barrier together with copper seed and electroplating have already been developed that meet the 32nm node targets [19].

The toolsets required for interconnect metallization are highly ionized Physical Vapor Deposition (PVD) for conformal Ta(N) barrier deposition and low overhang Cu seed deposition. Electrochemical Plating (ECP) is used to fill the damascene lines with Cu.

Planarization. Cu CMP at the 32nm node, similar to that of 45nm technology, starts with bulk Cu polish [20]. This is followed by an over-polish. The process must then remove the Ta and TaN barrier.

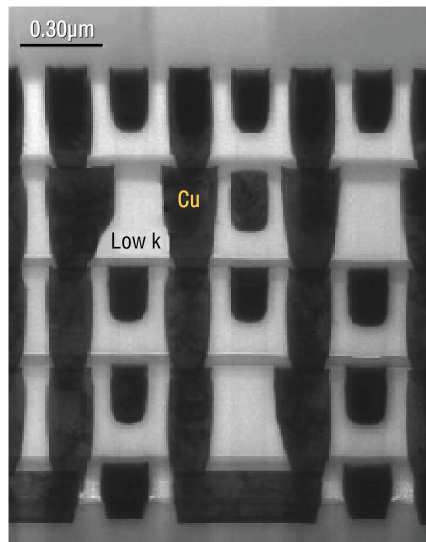

Figure 5. Cross-sectional image of multi-level Cu/low-k interconnect (k=2.5), with pitch=200nm [16]. ©2005 IEEE

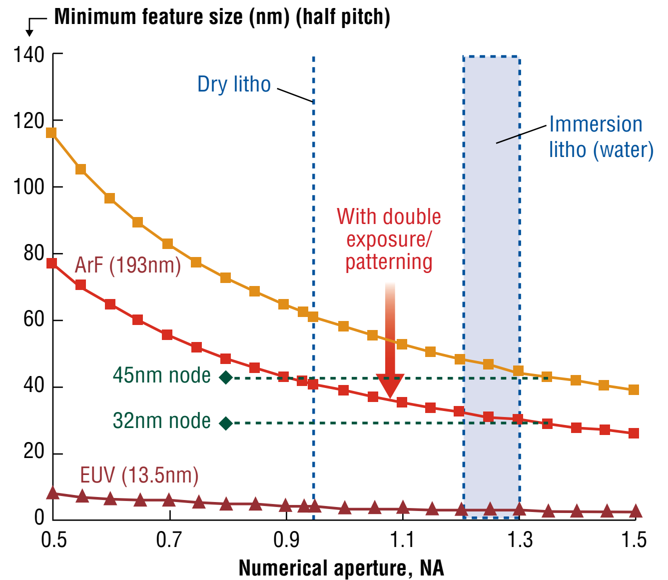

Figure 6. Minimum feature size for 32nm technology node can be achieved by combining 193nm (ArF) immersion lithography with double-exposure or double-patterning.

A capping layer is deposited to protect the low-k ~2.2 material from scratches and cracking during the CMP step. To further protect the dielectric, CMP downforce must be significantly reduced. The removal rate with low downforce can be improved by enhanced rotation speed in the tool.

Lithography challenges

The 32nm node targets an isolation half-pitch of ~90nm and physical gate length of 30nm. Poly half-pitch is roughly targeted at 50nm and contact-to-gate half-pitch at 60nm. Metal 1 is targeted at a half-pitch of 50nm with an aspect ratio approaching 2. The minimum feature size, or the half-pitch, that can be patterned by optical lithography is governed by the Rayleigh Equation:

$$d = k_1 \frac{\lambda_0}{NA} \quad (3)$$

where λ_0 is the illumination wavelength, NA is the numerical aperture of the lens, and k_1 is a constant that measures the "degree of difficulty" of patterning.

In conventional 193nm ArF dry lithography, a k_1 value >0.3 ensures optimal imaging. Increasing NA by using high-index fluids (immersion lithography) or high-index lens elements extends 193nm ArF to smaller feature patterning (**Fig. 6**). Another method to extend 193nm ArF is to use a lower k_1 factor, which can be achieved by adjusting the process integration.

Figure 7 illustrates generic process flows for double-exposure and double-patterning integration, which can both provide k_1 <0.2 capability. Whereas conventional lithography can result in pitch P_0 with lines $P_0/2$, these

techniques can halve the pitch and resolve lines equivalent to $P_0/4$. Double-exposure uses two resist exposures to define the layout prior to resist development. In double-patterning, the photoresist is coated twice and the exposure and development steps are repeated to pattern the underlying hard mask.

Two lithography options currently used for 45nm technology node manufacturing [3,21,22] are double-exposure with dry 193nm

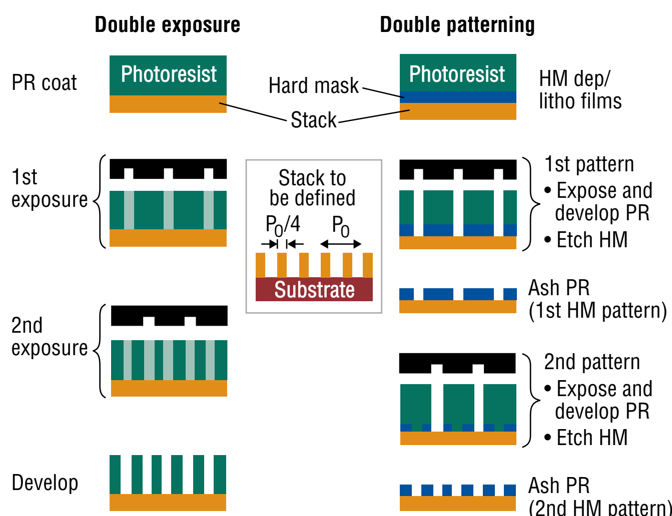

Figure 7. General process flows for double-exposure and double-patterning schemes.

ArF and a single exposure water-based 193nm ArF immersion lithography. Neither wet single-exposure nor dry double-exposure might meet the printing requirements of the 32nm technology node. Immersion lithography combined with either dual-exposure or double-patterning is able to meet these requirements.

Conclusions

The 32nm technology node for logic will likely remain on planar CMOS transistors. It would use high-*k* and metal-gates with the fourth generation of stressor films. For the first time, external resistance would rival the transistor channel resistance, and innovation in salicidation, junctions, and contact are required to reduce the overall external resistances. Cu/low-*k* interconnect with bulk *k*~2.2 and a lower-*k* etch-stop layer would further reduce the interconnect capacitance to meet the 32nm node back-end of the line performance requirements. ■

Acknowledgments

The authors would like to thank senior vice president Tom St. Dennis and the engineering teams of Applied Materials' Silicon System Group for their dedication and support of this program.

References

1. H. S. Yang et al., "Dual stress liner for high performance sub-45nm gate length SOI CMOS manufacturing," in *IEDM Tech. Dig.*, Dec. 2004, pp. 1075-1078.
2. R. Arghavani et al., "Strain engineering push to the 32nm logic technology node," in *Semiconductor Fabtech*, 32nd Edition, 2007.
3. K. Mistry et al., "A 45nm logic technology with high-*k*+metal gate transistor, strained silicon, 9Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging," in *IEDM Tech. Dig.,* Dec. 2007, pp. 247-250.
4. J. W. Sleight et al., "Challenges and opportunities for high performance 32nm CMOS technology," in *IEDM Tech. Dig.*, Dec. 2006, pp. 697-700.
5. R. Arghavani et al., "External resistance: a paradigm shift in approaching strain engineering," in *Semiconductor Fabtech*, 35th Edition, 2007.
6. L. Washington et al., "pMOSFET with 200% mobility enhancement induced by multiple stressors,"*IEEE Electron Device Lett.*, vol. 26, no. 6, June 2006, pp. 511-513.
7. M. Chudzik et al., "High-performance high-*k*/metal gates for 45nm CMOS and beyond with gate-first processing," in *VLSI Tech. Dig.*, 2007, pp. 194-195.
8. R. Chau et al., "Advanced CMOS transistors in the nanotechnology era for high-performance, low-power logic applications," in *ICSICT Tech. Dig.*, 2004, pp. 26-30.
9. E. P. Gusev, V. Narayanan, and M. M. Frank, "Advanced high-*k* dielectric stacks with polySi and metal gates: Recent progress and current challenges," in *IBM J. Res. & Dev.*, vol. 90, no. 4/5, Jul./Sept. 2006, pp. 387.
10. V. Narayanan 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," in *VLSI Tech. Dig.*, 2006, pp. 178-179.
11. H. N. Alshareef et al., "Metal gate work function engineering using AlNx interfacial layers," *Appl. Phys. Lett.*, vol. 88, 2006, pp. 112-114.
12. M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-*k* insulator: The role of remote phonon scattering." *J. Appl. Phys.*, vol. 90, no. 9, 2001, pp. 4587-4608.
13. S. Sankaran et al., "A 45nm CMOS node Cu/low-*k*/ultra low-*k* PECVD SiCOH (*k*=2.4) BEOL technology," in *IEDM Tech. Dig.*, 2006.
14. M. T. Bohr, "Interconnect scaling - The real limiter to high performance ULSI," in *IEDM Tech. Dig.*, 1995, pp. 241-244.
15. R. Cheung and D. Hoffman, "Physics and process drive etch performance at 45nm," in *Solid State Technology*, March 2005, pp. 27.
16. R. Fox et al., "High Performance k=2.5 ULK Backend Solution Using an Improved TFHM Architecture, Extendible to the 45nm Technology Node," in *IEDM Tech. Dig.*, 2005.
17. J. Van Olmen et al., "Cu resistivity scaling limits for 20nm copper damascene lines," in *IITC Tech. Dig.*, 2007, pp. 49-51.
18. D. Edelstein et al., "A High Performance Liner for Copper Damascene Interconnects," in *IITC Tech. Dig.*, 2001, pp. 9-11.
19. J. Van Olmen et al., "Integration of 50nm half pitch single damascene copper trenches in BDIIx by means of double patterning 193nm immersion lithography on metal hardmask," in *AMC Tech. Dig.*, 2007, pp. 5.
20. L. Karuppiah et al., "Overview of CMP Process Control Strategies," in *CMP-MIC Tech. Dig.*, February 2006, pp. 45-54.
21. S. Narasimha et al., "High Performance 45-nm SOI Technology with Enhanced Strain, Porous Low-*k* BEOL, and Immersion Lithography," in *IEDM Tech. Dig.* 2006, pp. 1-4.
22. H.-Y. Chen et al., "Novel 20nm Hybrid SOI/Bulk CMOS Technology with 0.183μm^2 6T-SRAM Cell by Immersion Lithography," in *VLSI Tech. Dig.*, 2005, pp. 16-17.

Reza Arghavani is a Fellow at Applied Materials, currently focused on developing thin film technologies to enable the sub-45nm node for logic/non-volatile memory technologies. He graduated from the U. of California at Los Angeles with a PhD in physics.

Hichem M'saad is general manager of Applied Materials' Dielectric Systems and CMP Group. He earned his MS in materials science and engineering from Cornell U. and PhD in materials science at MIT.

APPLIED MATERIALS®