INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

2007 EDITION

FRONT END PROCESSES

THE ITRS IS DEVISED AND INTENDED FOR TECHNOLOGY ASSESSMENT ONLY AND IS WITHOUT REGARD TO ANY COMMERCIAL CONSIDERATIONS PERTAINING TO INDIVIDUAL PRODUCTS OR EQUIPMENT.

UMC 1512
UMC v. AICP

TABLE OF CONTENTS

Front End Processes 1
Scope 1
Difficult Challenges 2
The End of Traditional Scaling—The Beginning of New Materials and New Structures 2
Technology Requirements and Potential Solutions 5
Starting Materials 5
Front End Surface Preparation 12
Thermal/Thin Films, Doping, and Etching 19
Front End Etch Processing 37
DRAM Stacked Capacitor 40
DRAM Trench Capacitor 43
Non-volatile Memory (Flash) 45
Phase Change Memory 50
Ferroelectric Random Access Memory (FeRAM) 53
Inter-Focus ITWG Discussion 57
Impact of Future Emerging Research Devices 57
Cross-cut Discussion 58
FEP Metrology Cross-Cut Issues 58
FEP Modeling and Simulation Cross-Cut Issues 58
FEP Environment, Safety, and Health Cross-Cut Issues 59
Conclusion 59
References 60

LIST OF FIGURES

Figure FEP1 Front End Process Chapter Scope 1
Figure FEP2 Starting Materials Potential Solutions 11
Figure FEP3 Front End Surface Preparation Potential Solutions 18
Figure FEP4 Thermal/Thin Films Potential Solutions 31
Figure FEP5 Doping Potential Solutions 36
Figure FEP6 Front End Etch Processing Potential Solutions 39
Figure FEP7 DRAM Stacked Capacitor Potential Solutions 43
Figure FEP8 Minimum Feature Size of NAND Flash Memory 45
Figure FEP9 Minimum Feature Size of NOR Flash Memory 46
Figure FEP10 Flash Memory Interpoly Dielectric Thickness Scaling at 45 nm 46
Figure FEP11 Schematics of STI Isolation Trenches 47
Figure FEP12 Evolution of the STI Aspect Ratio for Flash Memories with the Minimum Feature Size 47
Figure FEP13 Flash Non-Volatile Memory Floating Gate Potential Solutions 49
Figure FEP14 Amorphous / Poly-crystal Phases of a Chalcogenide Alloy, usually $Ge_2Sb_2Te_5$ (GST) 50
Figure FEP15 Resistance Change of GST 51
Figure FEP16 Set/Reset Thermal Cycles to Change the Crystal Phase of the GST Material and to Write/Erase the PCM 51

UMC 1512
UMC v. AICP

LIST OF TABLES

Table FEP1 Front End Processes Difficult Challenges 4
Table FEP2a Starting Materials Technology Requirements—Near-term Years 6
Table FEP2b Starting Materials Technology Requirements—Long-term Years 7
Table FEP3a Front End Surface Preparation Technology Requirements—Near-term Years 14
Table FEP3b Front End Surface Preparation Technology Requirements—Long-term Years 15
Table FEP4a Thermal, Thin Film, Doping and Etching Technology Requirements—Near-term Years 21
Table FEP4b Thermal, Thin Film, Doping and Etching Technology Requirements—Long-term Years 25
Table FEP5a DRAM Stacked Capacitor Technology Requirements—Near-term Years 41
Table FEP5b DRAM Stacked Capacitor Technology Requirements—Long-term Years 42
Table FEP6a DRAM Trench Capacitor Technology Requirements—Near-term Years 44
Table FEP6b DRAM Trench Capacitor Technology Requirements—Long-term Years 44
Table FEP7 FLASH Non-volatile Memory Technology Requirements 48
Table FEP8a Phase Change Memory (PCM) Technology Requirements—Near-term Years 53
Table FEP8b Phase Change Memory (PCM) Technology Requirements—Long-term Years 53
Table FEP9a FeRAM Technology Requirements—Near-term Years 56
Table FEP9b FeRAM Technology Requirements—Long-term Years 56

UMC 1512
UMC v. AICP

FRONT END PROCESSES

SCOPE

The Front End Processes (FEP) Roadmap focuses on future process requirements and potential solutions related to scaled field effect transistors (MOSFETs), DRAM storage capacitors, and non-volatile memory (Flash, Phase-change, and ferroelectric). The purpose of this chapter is to define comprehensive future requirements and potential solutions for the key front end wafer fabrication process technologies and materials associated with these devices. Hence, this Roadmap encompasses the tools, and materials, as well as the unit and integrated processes starting with the silicon wafer substrate and extending through the contact silicidation processes and the deposition of strain layers (pre-metal dielectric deposition and contact etching is covered in the *Interconnect* chapter). The following specific technology areas are covered: *starting materials, surface preparation, thermal/thin films, doping*, and *front end plasma etch* for MOSFETs, as well as processes and materials for *DRAM stack and trench capacitors, Flash memory gate structures, Phase-change memory,* and *FeRAM storage devices.*

A forecast of scaling-driven technology requirements and potential solutions is provided for each technology area. The forecasted requirements tables are model-based unless otherwise noted. The identified potential solutions serve to benchmark known examples of possible solutions, and are intended for other researchers and interested parties. They are not to be considered the only approaches. Indeed, innovative, *novel solutions* are sought, and their need is identified by red colored regions of the requirements tables.

Some FEP-related topics are presented in other sections of this Roadmap. The scaled device performance and structures forecasts that drive FEP requirements are covered in the *Process Integration, Devices, and Structures (PIDS)* chapter. Discussion of issues and potential solutions for shallow trench isolation chemical mechanical polish (CMP) is found in the *Interconnect* chapter because of overlap with interconnect tool issues. The crosscut needs of FEP are covered in the following chapters: *Yield Enhancement, Metrology, Environment, Safety, & Health*, and *Modeling & Simulation*. FEP factory requirements are covered in the *Factory Integration* chapter.

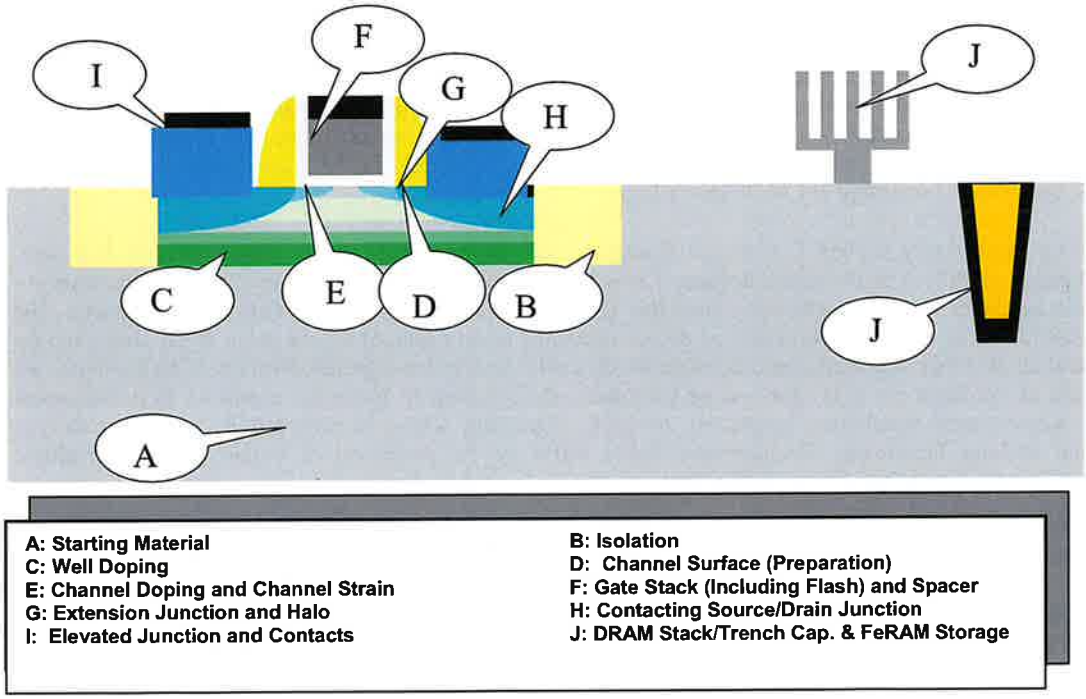


Figure FEP1 Front End Process Chapter Scope

UMC 1512
UMC v. AICP

DIFFICULT CHALLENGES

THE END OF TRADITIONAL SCALING—THE BEGINNING OF NEW MATERIALS AND NEW STRUCTURES

MOSFET scaling has been the primary means by which the semiconductor industry has achieved the historically unprecedented gains in productivity and performance quantified by Moore's Law. These gains have traditionally been paced by the development of new lithography tools, masks, photoresist materials, and critical dimension etch processes. In the past several years it has become clear that despite advances in these crucial process technologies and the resultant ability to produce ever-smaller feature sizes, front end process technologies have not kept pace, and scaled device performance has been compromised. The crux of this problem comes from the fact that the traditional transistor and capacitor formation materials, silicon, silicon dioxide, and polysilicon have been pushed to fundamental material limits and continued scaling has required the introduction of new materials. The current situation can be defined as "material-limited device scaling." In addition, new approaches to device structure, such as non-planar multi-gate devices, will be needed for future performance scaling.

Material-limited device scaling has placed new demands on virtually every front end material and unit process, starting with the silicon wafer substrate and encompassing the fundamental planar CMOS building blocks and memory storage structures. In addition, the end of planar bulk CMOS is becoming visible within the next several years. As a consequence we must be prepared for the emergence of CMOS technology that uses non-conventional MOSFETs or alternatives such as planar fully depleted SOI (FDSOI) devices and dual- or multi-gate devices either in a planar of vertical geometry. An overview of the device alternatives is presented in the *Emerging Research Devices chapter*. Projections for the manufacturing introduction of non-conventional MOSFET devices are 2010 for FDSOI and 2011 for multi-gate. The challenges associated with integration of these diverse new materials and structures is the central theme of the FEP difficult challenges summarized in Table FEP1.

In no area is the issue of material-limited device scaling more clear or urgent than in the MOSFET gate stack. Here, a new gate dielectric material having a higher dielectric constant than SiO_2 is needed. This need was identified in the 1999 ITRS where it was linked to MOSFETs having gate lengths smaller than 65 nm, which were at that time expected to emerge in the year 2005. In the interim, the patterning technology for producing 65 nm gates has accelerated and these have been achieved in 2001. Combined with the extension of silicon oxynitride gate dielectric materials and the introduction of strain-enhanced-mobility channels, the need for high-κ had been delayed, but is now upon us. It is expected that leading manufacturers will start production of high-κ gate dielectrics in 2008. Mobility enhancement and channel-length scaling, which requires accelerated scaling of junctions to control short channel effects, will continue to provide enhanced device performance. In addition, the depletion layers that exist in the doped polysilicon gate material become more limiting as planar devices are scaled into the deep submicron region. Therefore, it is also expected that dual metal gates, having appropriate work functions, will be put in production by leading manufacturers in 2008 to replace the dual doped polysilicon gates, currently the mainstay of CMOS technology.

Continued transistor performance scaling is expected to require the replacement of planar CMOS devices with non-classical devices including fully depleted planar devices. The introduction of these devices will require the replacement of bulk silicon substrates with ultra-thin, silicon-on-insulator (SOI) substrates and double- or multi-gate devices. The transition from extended bulk CMOS to non-classical device structures is not expected to take place at the same time for all applications and all chip manufacturers. Instead, a scenario is envisioned where a greater diversity of technologies are competitively used at the same point in time—some manufacturers choosing to make the transition to non-classical devices earlier, while others emphasize extensions of bulk technology. This is reflected in the Thermal/Thin Films/Doping and Etching Technology Requirements Table FEP4 by the projection of requirements for multiple approaches in the transition years from 2010 through 2015.

The introduction of new materials is also expected to impose added challenges to the methods used to dope and activate silicon. In addition to the scaling imposed need for producing very shallow highly activated junctions, the limited thermal stability of most high-κ materials is expected to place new boundaries on thermal budgets associated with dopant activation. In a worst-case scenario, the introduction of these materials could have a significant impact on the overall CMOS process architecture.

In the memory area, high-κ materials are now in use for both stacked and trench DRAM capacitors. DRAM stacked capacitors are also now using metal-insulator-metal (MIM) structures with trench capacitors projected to move to MIM by 2010. It is expected that high-κ materials will be required for the floating gate Flash memory interpoly dielectric by 2010 and for tunnel dielectric by 2013. FeRAM will make a significant commercial appearance where ferroelectric and ferromagnetic storage materials would be used. The introduction of these diverse materials into the manufacturing

UMC 1512
UMC v. AICP

mainstream is viewed as important difficult challenges. In addition, phase-change memory (PCM) devices are expected to make a commercial appearance by 2010.

In starting materials , it is expected that alternatives to bulk silicon such as silicon-on-insulator substrates will proliferate. Additionally, various forms of strained silicon technology may be incorporated although these have been and continue to be principally achieved through value-added modifications to the IC manufacturing process. Such bulk alternatives generally imply process architecture changes that impact FEP. Also, an important difficult challenge expected to emerge within this 2007 Roadmap horizon is the potential need for the next generation 450 mm silicon substrate. Such a diameter move is indicated to maintain pace with historic productivity enhancements based on augmented transistor count performance enhancements. However, so-called "More than Moore" approaches, which leverage enhanced design and/or inclusion of non-CMOS content, continue to expand within the industry. Further, higher productivity 300 mm fab approaches are being pursued. The ITRS is actively considering how such approaches will impact overall productivity requirements. Should it be necessary to adopt the next diameter silicon wafer, there are concerns whether the incumbent techniques for wafer preparation can be cost-effectively scaled to the next generation. It is also uncertain whether this substrate will be bulk silicon or SOI and whether strained silicon will be the required active layer material. Therefore, the search for potential substrate alternatives presents an important research need. Based upon historical diameter change cycles, the industry is already several years behind the pace necessary to allow the next generation 450 mm silicon substrate to be ready for device manufacture in the year 2012.

Front end cleaning processes will continue to be impacted by the introduction of new front end materials such as high-κ dielectrics, metal gate electrodes, and mobility-enhanced channel materials. Scaled devices are expected to become increasingly shallow, requiring that cleaning processes become completely benign in terms of substrate material removal and surface roughening. Scaled and new device structures will also become increasingly fragile, limiting the physical aggressiveness of the cleaning processes that may be employed. In addition, these new device structures will require precise cleaning and characterization of vertical surfaces. DRAM stacked and trench storage capacitor structures will show increasing aspect ratios making sidewall contamination removal increasingly difficult. Also, there is a challenge for particle scanning technology to reliably detect particles smaller than 28 nm on a wafer surface for characterization of killer defect density and to enable yield learning.

The persistent need in scaling devices is to control the critical dimension (CD) of the smallest device features, not only to reduce the median, but to also narrow the distribution, at least in absolute terms. With high-κ dielectrics and metal gates going into production, etch processes with sufficient selectivity and damage control for use with these materials have been identified. Continued scaling requires honing and optimizing these processes to allow for CD reduction. Etching uniformity all the way to the wafer edge is a particularly difficult issue. As the CD shrinks, the presence of line edge roughness (LER) is becoming more important to CD control. The LER is at best staying constant as the linewidth shrinks, which makes it a major scaling concern. As non-planar transistors become necessary, etch becomes much more challenging. FinFET configurations bring new constraints to selectivity, anisotropy, and damage control.

UMC 1512
UMC v. AICP

Table FEP1 Front End Processes Difficult Challenges

Difficult Challenges ≥ 22 nm	*Summary of Issues*
Starting Materials	1.5 mm edge exclusion FDSOI Si and buried oxide thickness control SOI defectivity levels Full production of 450 mm wafer size
Surface Preparation	Critical surface particle size below 28 nm not measurable on wafer Ability to achieve clean surfaces while controlling material loss and surface damage
Thermal/Thin Films/Doping/Etch	Introduction of high-κ/metal gate into high performance (HP) and low operating/low standby power (LOP/LSTP) and equivalent oxide thickness (EOT) scaling below 0.8 nm Increasing device performance with strain engineering and applying it to FDSOI and multi-gate technologies Scaling extension junction depths below 10 nm while achieving high dopant activation Achieving manufacturable interfacial contact resistivities below 10^{-7} Ω-cm^2 to meet parasitic series resistance requirements Si thickness and control for FDSOI and Multi-gate Gate critical dimension control for physical gate length < 20 nm Introduction of new channel materials with high interface quality and low processing thermal budget
DRAM	Improvement of oxide etching capability for high aspect ratio (>40) storage node formation in stack capacitor and for oxide hardmask for high aspect ration trench capacitor. Improvement of Si etching capability for high A/R (>90) trench capacitor formation. Continued scaling of stacked and trench capacitor dielectric Teq below 0.5 nm Continued scaling of physcial dielectric thickness (tphys) while maintaining high dielectric constant (>90) and low leakage current of dielectric
Non-volatile Memory	Scaling of IPD Teq to <6Å for NAND and NOR Scaling of tunnel oxide thickness to <8Å for NOR Scaling of STI fill aspect ratio to >9 starting for NAND PCM material conformality of ≥90% PCM minimum operating temperature of 125°C PCM resistivity change and reset current density Integration and scaling of FeRAM ferroelectric materials Continued scaling of FeRAM cell structure
Difficult Challenges < 22 nm	*Summary of Issues*
Starting Materials	1.5 mm edge exclusion FDSOI Si and buried oxide thickness control SOI defectivity Surface particles
Surface Preparation	Surface particles not measurable Ability to achieve clean surfaces while controlling material loss and surface damage Metrology of surfaces that may be horizontally or vertically oriented relative to the chip surface Achievement of statistically significant characterization of surfaces and interfaces that may be horizontally or vertically oriented relative to the chip surface Achievement and maintenance of structural, chemical, and contamination control of surfaces and interfaces that may be horizontally or vertically oriented relative to the chip surface
Thermal/Thin Films/Doping/Etch	Continued scaling of HP multigate device in all aspects: EOT, junctions, mobility enhancement, new channel materials, parasitic series resistance, contact silicidation. Continued EOT scaling below 0.7 nm with appropriate metal gates Gate CD Control
DRAM	Continued scaling of capacitor structures for both stacked and trench type as well as continued scaling of dielectric thickness
Non-volatile Memory	Floating gate Flash technology considered unscalable beyond 22 nm—new Flash NVM technology will be required Continued scaling of phase change memory technology Continued scaling of FeRAM technology

UMC 1512
UMC v. AICP

TECHNOLOGY REQUIREMENTS AND POTENTIAL SOLUTIONS

STARTING MATERIALS

Technology Requirements—Tables FEP2a and FEP2b forecast trends for starting wafers produced by silicon wafer manufacturers that are intended for use in the manufacture of both high density memories such as DRAMs and high-performance MPUs and ASICs. These requirements include parameters common to all wafers plus parameters specific to epitaxial and SOI wafers. Fundamental barriers presently limit the rate of cost-effective improvement in wafer characteristics such as localized light scatterers (LLS) defect densities, site flatness values, and edge exclusion dimensions. These barriers include the capability and throughput limitations of metrology tools, as well as wafer manufacturing cost and yield issues fundamental to the crystal-pulling process and subsequent wafer finishing operations. Accordingly, use of the methodology introduced in the 2005 edition of the ITRS, to display not only the ability of the wafer supplier to meet the parameter trends in Tables FEP2a and FEP2b, but to also display the metrology tool readiness, is continued. The marking system and meanings are shown in the tables for both DRAM and high-performance MPUs. Additionally, a new section accessed by hyperlink provides further relevant information pertaining to metrology. (link to Metrology information)

Wafer Types—For the device types included in the scope of the ITRS, starting materials selection historically involved the choice of either polished Czochralski (CZ) or epitaxial silicon wafers. However, silicon-on-insulator (SOI) wafer usage continues to show strong growth, although the total number of SOI wafers shipped is still small compared to polished and epi wafers. The prospect for SOI wafers to be used in mainstream, high-volume applications is being driven by improved high-frequency logic performance and reduced power consumption. Further opportunity is afforded with enhanced device performance via unique device configurations such as multiple-gate structures, but these require additional development of both the wafer and device processes in order to achieve practical volume production. In some cases, device process flow simplification is also achieved with SOI. Therefore, the selection of wafer type is based strongly on performance versus overall cost per die and should include all aspects for consideration, not just starting wafer price.

Commodity devices such as DRAM are commonly manufactured on lower cost CZ polished wafers. The elimination of "crystal originated pits" (COP) in CZ polished wafers is increasingly required to avoid interference with inline inspections used for defect reduction and yield enhancement. High-performance logic ICs are generally manufactured on more costly epitaxial wafers (compared to polished CZ wafers) because their use has facilitated the achievement of greater-robustness (e.g., soft error immunity and latch-up suppression capability). This latter capability may no longer be as critical due to the implementation of shallow trench isolation (STI) and the development of alternate doping means for achieving latch-up suppression. Additionally, partially depleted SOI has been adopted for certain types of high performance logic ICs.

Annealed wafers were introduced in the early 1990s as an alternative means to provide a silicon wafer with a COP-free surface and are now used for many leading edge device applications. Annealing occurs in either hydrogen ($\leq$ 200 mm wafer diameter) or argon ambients at high temperatures. COPs can also be controlled by appropriately engineered CZ growth methodologies. For the purpose of the Starting Materials tables presented here, annealed wafers and "defect engineered CZ" are both considered forms of polished CZ wafer and have parameter trends noted in the General Characteristics sections.

This wide variety of starting materials will likely continue into the foreseeable future and accounts for inclusion of general as well as specific epitaxial and SOI wafer characteristics in Tables FEP2a and FEP2b. Emerging materials that may further augment the variety of starting materials are discussed later in this document.

Parameter Values—The wafer requirements have been selected to ensure that in any given year each parameter value contributes no more than 1% to leading-edge chip yield loss. The values in the tables are generally, but not exclusively, derived from probabilistic yield-defect models. These models take into account leading-edge technology parameters such as critical dimension (CD)—taken as the DRAM half-pitch (i.e., the technology generation)—bit density, transistor density, and chip size. The validity of these derived values is limited by the sometimes questionable accuracy and predictability of the underlying models. With the onset of nanometer device dimensions for both the gate dielectric equivalent oxide thickness and the device physical channel length, compliance with these model-based values can be very costly and, in some cases, requires re-examination. For this reason, detailed re-assessment of the costs incurred versus the value derived from achieving compliance often suggests limiting the scope of these models via appropriate truncation.

UMC 1512
UMC v. AICP

Table FEP2a Starting Materials Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ Pitch (nm) (contacted)	*65*	*57*	*50*	*45*	*40*	*36*	*32*	*28*	*25*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*68*	*59*	*52*	*45*	*40*	*36*	*32*	*28*	*25*
MPU Physical Gate Length (nm)	*25*	*23*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
DRAM Total Chip Area (mm^2)	*93*	*74*	*59*	*93*	*74*	*59*	*93*	*74*	*59*
DRAM Active Transistor Area (mm^2)	*29.6*	*23.1*	*18.2*	*29.1*	*23.1*	*18.3*	*29.1*	*23.1*	*18.3*
MPU High-Performance Total Chip Area(mm^2)	*310*	*246*	*195*	*310*	*246*	*195*	*310*	*246*	*195*
MPU High-Performance Active Transistor Area(mm^2)	*31.7*	*25.1*	*20.0*	*31.7*	*25.1*	*20.0*	*31.7*	*25.1*	*20.0*
*General Characteristics * (99% Chip Yield)*									
Maximum Substrate Diameter (mm)—High-volume Production (>20K wafer starts per month)**	300	300	300	300	300	450	450	450	450
Edge exclusion (mm)	2	2	2	2	2	1.5	1.5	1.5	1.5
Front surface particle size (nm), latex sphere equivalent (A)	≥65	≥65	≥65	≥65	≥65	≥45	≥45	≥45	≥32
Particles (cm–2)	≤0.32	≤0.30	≤0.30	≤ 0.15	≤ 0.15	≤ 0.32	≤ 0.16	≤ 0.16	≤ 0.31
Particles (#/wf)	≤218	≤209	≤205	≤ 105	≤ 105	≤ 498	≤ 249	≤ 249	≤ 492
Site flatness (nm), SFQR 26mm x 8 mm Site Size	≤65	≤57	≤50	≤45	≤40	≤36	≤32	≤28	≤25
Nanotopography, p-v, 2 mm dia. analysis area (I)	≤16	≤14	≤13	≤11	≤10	≤9	≤8	≤7	≤6
*Epitaxial Wafer * (99% Chip Yield)*									
Large structural epi defects (DRAM) (cm^{-2}) (B)	≤0.011	≤ 0.014	≤ 0.017	≤ 0.011	≤ 0.014	≤ 0.017	≤ 0.011	≤ 0.014	≤ 0.017
Large structural epi defects (MPU) (cm^{-2}) (B)	≤0.003	≤ 0.004	≤ 0.005	≤ 0.003	≤ 0.004	≤ 0.005	≤ 0.003	≤ 0.004	≤ 0.005
Small structural epi defects (DRAM) (cm^{-2}) (C)	≤0.022	≤ 0.027	≤ 0.034	≤ 0.022	≤ 0.027	≤ 0.034	≤ 0.022	≤ 0.027	≤ 0.034
Small structural epi defects (MPU) (cm^{-2}) (C)	≤0.006	≤0.008	≤ 0.010	≤ 0.006	≤ 0.008	≤ 0.010	≤ 0.006	≤ 0.008	≤ 0.010
Silicon-On-Insulator Wafer (99% Chip Yield)*									
Edge exclusion (mm) ***	2	2	2	2	2	1.5	1.5	1.5	1.5
Starting silicon layer thickness (Partially Depleted) (tolerance ± 5%, 3σ) (nm) (D)	48-83	44-76	40-60	37-55	34-50	31-45	29-42	27-38	25-35
Starting silicon layer thickness (Fully Depleted) (tolerance ± 5%, 3σ) (nm) (E)				15-28	14-17	14-16	13-16	13-14	12-14
Buried oxide (BOX) thickness (Fully Depleted) (tolerance ± 5%, 3σ) (nm) (F)				26-44	24-40	22-36	18-32	16-28	16-26
D_{LASOI}, Large area SOI wafer defects (DRAM) (cm^{-2}) (G)	≤ 0.011	≤ 0.014	≤ 0.017	≤ 0.011	≤ 0.014	≤ 0.017	≤ 0.011	≤ 0.014	≤ 0.017
D_{LASOI}, Large area SOI wafer defects (MPU) (cm^{-2}) (G)	≤ 0.003	≤ 0.004	≤ 0.005	≤ 0.003	≤ 0.004	≤ 0.005	≤ 0.003	≤ 0.004	≤ 0.005
D_{SASOI}, Small area SOI wafer defects (DRAM) (cm^{-2}) (H)	≤ 0.170	≤ 0.218	≤ 0.276	≤ 0.173	≤ 0.218	≤ 0.274	≤ 0.173	≤ 0.218	≤ 0.274
D_{SASOI}, Small area SOI wafer defects (MPU) (cm^{-2}) (H)	≤ 0.159	≤ 0.200	≤ 0.252	≤ 0.159	≤ 0.200	≤ 0.252	≤ 0.159	≤ 0.200	≤ 0.252

UMC 1512
UMC v. AICP

Table FEP2b Starting Materials Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
DRAM ½ Pitch (nm) (contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*
MPU Physical Gate Length (nm)	*9*	*8*	*7*	*6*	*6*	*5*	*4*
DRAM Total Chip Area (mm^2)	*93*	*74*	*59*	*93*	*74*	*59*	*93*
DRAM Active Transistor Area (mm^2)	*29.1*	*23.1*	*18.3*	*29.1*	*23.1*	*18.3*	*29.1*
MPU High-Performance Total Chip Area(mm^2)	*310*	*246*	*195*	*310*	*246*	*195*	*310*
MPU High-Performance Active Transistor Area(mm^2)	*31.7*	*25.1*	*20.0*	*31.7*	*25.1*	*20.0*	*31.7*
*General Characteristics * (99% Chip Yield)*							
Maximum Substrate Diameter (mm)—High-volume Production (>20K wafer starts per month)**	450	450	450	450	450	450	450
Edge exclusion (mm)	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Front surface particle size (nm), latex sphere equivalent (A)	≥32	≥32	[3]22	[3]22	[3]22	[3]16	[3]16
Particles (cm–2)	≤ 0.16	≤ 0.16	≤ 0.33	≤ 0.17	≤ 0.17	≤ 0.31	≤ 0.16
Particles (#/wf)	≤ 246	≤ 246	≤ 521	≤ 260	≤ 260	≤ 492	≤ 246
Site flatness (nm), SFQR 26mm x 8 mm Site Size	≤23	≤20	≤18	≤16	≤14	≤13	≤11
Nanotopography, p-v, 2 mm dia. analysis area (I)	≤6	≤5	≤4	≤4	≤4	≤3	≤3
*Epitaxial Wafer * (99% Chip Yield)*							
Large structural epi defects (DRAM) (cm^{-2}) (B)	≤ 0.011	≤ 0.014	≤ 0.017	≤ 0.011	≤ 0.014	≤ 0.017	≤ 0.011
Large structural epi defects (MPU) (cm^{-2}) (B)	≤ 0.003	≤ 0.004	≤ 0.005	≤ 0.003	≤ 0.004	≤ 0.005	≤ 0.003
Small structural epi defects (DRAM) (cm^{-2}) (C)	≤ 0.022	≤ 0.027	≤ 0.034	≤ 0.022	≤ 0.027	≤ 0.034	≤ 0.022
Small structural epi defects (MPU) (cm^{-2}) (C)	≤ 0.006	≤ 0.008	≤ 0.010	≤ 0.006	≤ 0.008	≤ 0.010	≤ 0.006
Silicon-On-Insulator Wafer (99% Chip Yield)*							
Edge exclusion (mm) ***	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Starting silicon layer thickness (Partially Depleted) (tolerance ± 5%, 3σ) (nm) (D)	23-32	22-30	21-28	19-26	18-24	18-23	17-21
Starting silicon layer thickness (Fully Depleted) (tolerance ± 5%, 3σ) (nm) (E)	12-13	12-13	12-13	11-12	11-12	11-12	11-12
Buried oxide (BOX) thickness (Fully Depleted) (tolerance ± 5%, 3σ) (nm) (F)	14-22	12-20	10-18	10-16	8-14	8-12	6-12
D_{LASOI}, Large area SOI wafer defects (DRAM) (cm^{-2}) (G)	≤ 0.011	≤ 0.014	≤ 0.017	≤ 0.011	≤ 0.014	≤ 0.017	≤ 0.011
D_{LASOI}, Large area SOI wafer defects (MPU) (cm^{-2}) (G)	≤ 0.003	≤ 0.004	≤ 0.005	≤ 0.003	≤ 0.004	≤ 0.005	≤ 0.003
D_{SASOI}, Small area SOI wafer defects (DRAM) (cm^{-2}) (H)	≤ 0.173	≤ 0.218	≤ 0.274	≤ 0.173	≤ 0.218	≤ 0.274	≤ 0.173
D_{SASOI}, Small area SOI wafer defects (MPU) (cm^{-2}) (H)	≤ 0.159	≤ 0.200	≤ 0.252	≤ 0.159	≤ 0.200	≤ 0.252	≤ 0.159

Meaning and Color Coding of Left Box	***Meaning and Color Coding of Right Box***
Technology Requirements Value and Supplier Manufacturing Capability by Color	**Metrology Readiness Capability by Color**
Manufacturable solutions exist, and are being optimized	*Manufacturable solutions exist, and are being optimized*
Manufacturable solutions are known	*Manufacturable solutions are known*
Interim solutions are known	*Interim solutions are known*
Manufacturable solutions are NOT known	*Manufacturable solutions are NOT known*

UMC 1512
UMC v. AICP

Tables FEP2a and FEP2b notes:

[A] Front surface particle size d_m is the minimum practical size at which particles are measured. Critical front surface particle size $d_c = K_1F$, $[K_1=1]$ (where F is the DRAM half-pitch and is used to calculate required particle densities at the given technology generation).

Particle density n is calculated as:

$n = n_c * (d_m/d_c)^2$

where nc is the particle density extracted from the conventional Maly Yield Equation[1]

{$Y = \exp[-(D_P R_P) A_{eff}]$, where A_{eff} is the effective chip area $A_{eff}=2.5*F^2T+(1-aF^2T/A_{chip})A_{chip}*0.18$, "a" is the DRAM cell fill factor (see Table FEP5a and b) and T = number of transistors or bits/chip per technology generation}.

[B] Large structural epi defects (large area defects >1 μm LSE signal) modeled at 99% yield where $Y = \exp(D_{LAD} R_{LAD} A_{chip})$.[1] *where $R_{LAD} = 1$ and A_{chip} applies to DRAM and high-performance MPU as appropriate.*

[C] Small structural epi defects (≤1μm LSE signal) modeled at 99% yield where $Y = \exp(-D_{SF} R_{SF} A_{chip})$[1] *where $R_{SF} = 0.5$ and A_{chip} applies to DRAM and high-performance MPU as appropriate. Starting Materials uses the DRAM at production and the MPU high-performance MPU areas.*

[D] The silicon final device layer thickness (partially depleted) is obtained by 2 × MPU physical gate length (with a range in nominal values of ± 25%). Range of target value refers to the center point measurement with uniformity to indicate within-wafer maximum positive or negative % deviation from the center value. In order to account for silicon consumption during device processing, starting wafer thickness values are obtained by adding 10 nm to the lower range value and 20nm to the upper range value. It is assumed that silicon consumption will be more tightly controlled after 2009 so table values are derived by adding 10 nm to both the upper and lower range values.

[E] The silicon final device thickness (fully depleted) is obtained by 0.35 × MPU physical gate length 2010 to 2011 and 0.3 × MPU physical gate length at 2012 and thereafter (with a range in nominal values of ± 25%). Range of target value refers to the center point measurement with uniformity to indicate within-wafer maximum positive or negative % deviation from the center value. In order to account for silicon consumption during device processing, starting wafer thickness table values are derived by adding 10 nm to both the upper and lower range values.

[F] The BOX thickness for fully depleted devices is taken as the 2 × MPU physical gate length.

[G] Large area SOI (LASOI) wafer defects with yield of 99%; $Y = \exp(-D_{LASOI} R_{LASOI} A_{chip})$,[1] *$D_{LASOI}$ = LASOI defect density, R_{LASOI} =1.0 (best present estimate). Sources of LASOI may include missing top silicon and / or BOX or bond defects.*

[H] Small area SOI (SASOI) wafer defects with yield of 99%; $Y = \exp(-D_{SASOI} R_{SASOI} A_{eff})$,[1] *$D_{SASOI}$ =SASOI defect density, R_{SASOI} = 0.2 (best present estimate). Sources of SASOI can include COPs, metal silicides, or local SiO_2 islands in the top silicon layer. These SASOI defects may also be detected by localized light scattering (LLS) measurements.*[2, 3, 4]

[I] Peak-to-valley threshold (nm), 2 mm diameter analysis area. Maximum p-v is empirically taken as F/4 (F = DRAM half-pitch)

Model Limitations—These model-based parameter requirements do not include effects of distribution of parameter values intrinsic to the wafer manufacturing process where either of two statistical distributions commonly apply. Parameter values distributed symmetrically around a central or mean value, such as thickness, can often be described by the familiar normal distribution. The values of zero-bounded parameters (such as site flatness, particle density, and surface metal concentration) can usually be approximated by a lognormal distribution, in which the logarithms of the parameter values are normally distributed. The latter distribution is skewed with a long tail at the upper end of the distribution. Validation of the yield models remains elusive despite the experience of more than forty years of IC manufacturing.

The ideal methodology for management of material-contributed yield loss would be to allocate loss by defect type such that these defects do not contribute more than 1% to the overall IC fabrication yield loss. Yield loss for a particular defect is equal to the integral of the product of (1) the probability of failure due to a given value of the parameter (as established by the appropriate yield model) and (2) the fraction of wafers having that value (as established by the normal or lognormal distribution function). By applying this methodology, one could determine acceptable product distributions. Successful implementation of a distributional specification requires that the silicon supplier's process is sufficiently well understood, under control, and capable of meeting the IC user requirements. Until these ideals can be achieved, however, Poisson Distribution yield models based on the best available information are used and parameter limits assigned based on a 99% yield requirement for that parameter. It is further assumed that the yield loss from any individual wafer parameter does not significantly contribute to the yield loss from any other parameters, that is, that the defect yield impacts are statistically independent. Where validation data are available, this empirical approximation has been shown to result in requirement values nearly equal to the limit values obtained from the aforementioned methodology using parameter distributions.

Cost of Ownership (CoO)—As the acceptance values for many parameters approach metrology limits, enhanced cooperation between wafer suppliers and IC manufacturers is essential for establishing and maintaining acceptable product distributions and costs. Further development and validation of IC yield/defect models is required. However, it is essential to balance the "best wafer possible" against the CoO opportunity of not driving wafer requirements to the detection limit defined by acceptable metrology practice, but instead to some less stringent value consistent with achieving high IC yield. For example, the surface metal and particle contamination requirements for starting wafers are less stringent than the pre-gate values given in the *Surface Preparation* section (see Tables FEP3a and FEP3b) because it

UMC 1512
UMC v. AICP

is assumed that a minimum cleaning efficiency of 50% (actually 95% has been reported for surface iron removal) results during IC processing steps such as the pre-gate clean. It is also noted that the chemical nature of the surface requested by the IC manufacturer (hydrophilic versus hydrophobic) and the wafer-carrier interaction during shipment as well as the humidity in the storage areas are important in affecting the subsequent adsorption of impurities and particles on the wafer surface. Further emphasis on the CoO has been ascertained by developing a model examining the viability of a 100% wafer inspection to a particular parameter (i.e., site flatness). This model considers the wafer supplier's additional cost of ensuring 100% compliance to the IC manufacturer's specification relative to the potential loss associated with processing a die with a high probability of failing if a 100% inspection is not done. The relevant worksheets employing this methodology are available as links to this chapter so that each IC manufacturer can analyze the trade-off appropriate for their wafer specifications and product family of interest.

Wafer Parameter Selection—Both the chemical nature and the physical structure of the wafer front surface are of critical concern. Parameters related to the former are not included in Tables FEP2a and FEP2b due to lack of appropriate model-based definitions. Chemical defects include metal and organic particles and surface chemical residues. These defects are equally significant for all wafer types, although there is some concern that the detrimental effects of surface metals may be magnified in ultra-thin SOI films when the metals diffuse into a small silicon volume. Organic contamination strongly depends on environmental conditions during wafer storage and transportation, and accordingly is not included in Tables FEP2a and FEP2b.

With the adoption of double-side polished wafers, attention is also being given to particles on the back surface of the wafer to improve both the chemical and physical characteristics. The polished back surface more readily exhibits microscopic contamination and wafer handling damage. As a result, back-surface cleanliness requirements may emerge and drive the need for more stringent robotic handler standards. However, based on a Starting Materials IC Users Survey, site flatness degradation due to the presence of back-side particles does not currently appear to be of significance and has not been included in this edition of the ITRS. In addition, any back-surface treatments (e.g., extrinsic gettering and oxide back seal) may degrade the quality of both the polished back- and front-surfaces and are generally not compatible with standard wafer manufacturing approaches at diameters > 200 mm.

Important physical characteristics of the wafer front surface include wafer topography, structural defects and surface defects. Wafer topography encompasses various wafer shape categories that are classified according to their spatial frequency as site flatness, surface waviness, nanotopography or surface micro-roughness. Front surface site flatness and nanotopography are believed to be the most critical of the topographic parameters and are addressed in this ITRS revision. Back surface topography also has begun to receive attention recently, particularly in view of possible wafer interactions with stepper chucks, but the technology for quantifying this interaction is still not sufficiently mature to include such parameters in the tables at this time. Near-edge wafer geometry is also emerging as a potential yield-limiting attribute for silicon wafers. Often referred to as edge roll-off (ERO), it encompasses a variety of angularly and radially varying features in the region of the wafer surface between the substantially flat major central region of the wafer and the edge profile (the intentionally rounded outer periphery of the wafer). No industry consensus on metrics has yet been reached and therefore no trend values for future technology generations are established.

Structural defects include grown-in microdefects, such as COPs and bulk microdefects (BMDs). Methods of COP control have been discussed above. With advanced silicon manufacturing techniques, BMDs can be controlled independently of the interstitial oxygen concentration. In addition, current fab thermal cycles use lower temperatures and shorter times, and are not suitable to produce high levels of BMD for intrinsic gettering. As a result, in applications for which the customer is depending on BMD for gettering, a careful discussion of options with the silicon supplier is required

Other starting material requirements are expressed in terms of specific types of surface defects for different wafer types. Recent data suggest that certain devices (such as DRAM) produced on polished wafers may be sensitive to very shallow small scratches and pits. Epitaxial and SOI materials appear to exhibit fewer surface defects of this type. On the other hand, epitaxial and SOI wafer defects include large structural defects (arbitrarily defined as > 1 μm) and small structural defects (< 1 μm). Epitaxial wafers are subject to grown-in crystallographic defects such as stacking faults, and large defects created by particles on the substrate. Such defects must be controlled to maximize yields when using epitaxial wafers. Several defects are unique to SOI wafer. Large area defects are of the greatest concern to yield, and include voids in the SOI layer and large defects of the SOI/BOX interface. These large defects are judged to have a serious effect on chip yield and are assigned a kill rate of 100%. Smaller defects, such as COPs, metal silicides or local SiO_2 islands in the top silicon layer (measured in tens of nanometers to tenths of microns) are believed to have a less severe impact on device performance and thus the allowable density is calculated based on a lower kill rate. The development of laser scanning and other instrumentation to count, size, and determine the composition and morphology of these defects is a critical metrology challenge. While threshold size for detection continues to enjoy improvements, compositional and

UMC 1512
UMC v. AICP

morphological segregation remains insufficient. As such, the removal and prevention of surface defects continues to be a state-of-the-art challenge for silicon wafer technology.

The dependence of gate dielectric integrity and other yield detractors on crystal growth parameters as well as the related role of point defects and agglomerates have been extensively documented. The resulting defect density (D_o) parameter has served effectively as a measure of material quality for several device generations. However, for devices with EOT < 2 nm, this parameter is no longer an indicator of device yield and performance and is accordingly been not included in Tables FEP2a and FEP2b as a requirement. It should be noted, however, that starting material cleanliness requirements might change if pre- and post-gate surface preparation methods are modified when high-κ gate dielectric materials are introduced (see *Surface Preparation* section).

Metrology for SOI wafers is a significant challenge. Optical metrology tools operating at visible wavelengths do not have the same capabilities for characterizing SOI wafers as they have with polished or epitaxial wafers. Interference effects arising from multiple reflections from the Si and BOX layers fundamentally alter the response of these tools compared to polished and epitaxial wafers, generally degrading the measurement capability. Recently developed UV/DUV wavelength optical tools have been shown to help alleviate these difficulties at least for top silicon layers thicker than 10 nm, because of the much shorter optical absorption depth at these wavelengths. Metrology methods for many of the SOI defect categories call for destructive chemical etching that decorates, but does not uniquely distinguish, various types of crystal defects. These various defects may not all have the same origin, size, or impact on the device yield and, therefore, may exhibit different kill rates. Additionally, decorative defect etching on SOI wafers with very thin top silicon layers is very difficult because of the extremely small etch removals that must be used to avoid completely etching away the entire layer under inspection. Non-destructive and fast-turn around methods are also needed for the measurement of electrical properties and structural defects in SOI materials. Finally, the metrology issues for the various strained silicon configurations (spatially varying strain levels and Si:Ge composition, threading dislocations and associated defects as well as unique surface roughness issues) will require significant efforts (see the *Emerging Materials* section below).

Layer thickness and uniformity are included in Tables FEP2a and FEP2b for SOI wafers. For such wafers, the broad variety of today's IC applications requires a considerable range of Si-device layer and buried oxide (BOX) layer thicknesses. A number of SOI wafer fabrication approaches have entered production to serve this range of SOI applications. In some cases this includes strained SOI (sSOI), which has the same layer structure as conventional SOI, except the Si film is under biaxial tensile strain that increases the electron mobility, and to much lesser extent, the hole mobility. Strained Si is discussed in more detail in the *Emerging Materials* section of this chapter. The tables give incoming silicon thickness for both partially depleted (PD) and fully depleted (FD) devices. While the PD thickness values are extended through 2020, it is expected that around 2012, the actual application will be multi-gate devices. In the first order, these PD values are consistent with expected silicon thickness values for such multi-gate devices. Also, in order to be consistent with actual manufacturing practices within the industry, FD thickness values prior to 2010 have been removed from Table FEP2a.

Potential Solutions—Figure FEP2 lists the most significant starting materials challenges and shows potential solutions that have been identified, along with the necessary timing for development of these solutions and their transfer into high-volume production. In alignment with Tables FEP2a and FEP2b, Figure FEP2 reflects the requirements of leading edge DRAMs and high-performance MPUs, built on 300 mm (or larger) diameter wafers. It should be noted however that application of 200 mm wafers beyond the 90 nm technology generation is occurring and requires double-side polishing to achieve the necessary flatness and nanotopography levels. Implementation of this wafer type will require additional investments from both the wafer suppliers and users.

UMC 1512
UMC v. AICP

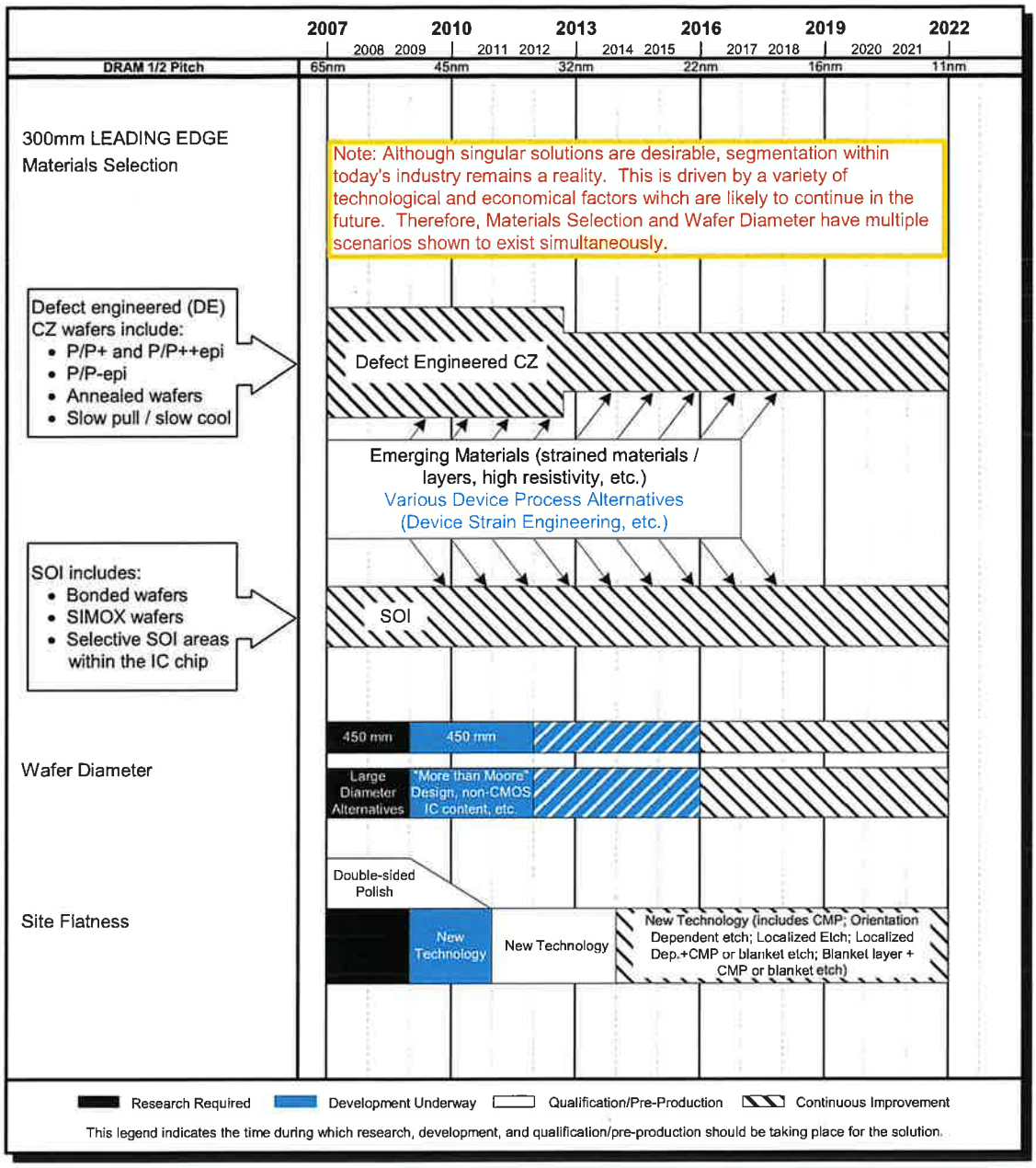


Figure FEP2 Starting Materials Potential Solutions

Material Selection—The materials selection category is divided into two sections—defect engineered CZ and SOI wafers. The type of material chosen depends strongly on the IC application and cost performance optimization. The former is typically utilized for cost-sensitive applications while the latter is used for performance-sensitive applications. As noted in Figure FEP2, potential solutions are diverging, which will result in a greater challenge to available resources.

Emerging Materials—The utilization of materials solutions, Emerging Materials, that augment other methods to meet ITRS targets remain critically important to the future of the silicon industry. For the 2007 ITRS, three distinct categories of Emerging Materials have been identified: 1) thermal management solutions, 2) mobility enhancement solutions, and 3) system-on-chip solutions. Examples of emerging materials that could potentially provide thermal managements solutions (i.e., improve heat dissipation properties) for future microelectronics applications include Si-on-Diamond and Si-on-

UMC 1512
UMC v. AICP

insulator with the insulator being a material of higher thermal conductivity than SiO_2, for example Al_2O_3 (alumina) or silicon nitride. In addition to concerns regarding heat dissipation, future microelectronic systems will feature transistor channels that have greater mobility than that of Si. Among those emerging materials potential solutions targeted at enhancing channel mobility are strained Si, germanium (relaxed and strained), and carbon nanotubes. Lastly, the ability to integrate new functionality into traditional CMOS logic architecture can be enabled by emerging materials innovations as well. High resistivity Si substrates and monolithic optical interconnection on Si are potential system-on-chip solutions. These emerging material topics, although potentially providing technical solutions to critical challenges facing future microelectronics, lack the maturity to include detailed specifications in Tables FEP2a and FEP2b for this year's ITRS revision. However, these topics will continue to be tracked and the Emerging Materials sub-committee of the ITRS Starting Materials team has assembled a detailed set of notes and references for the reader online.

Wafer diameter—Productivity enhancement has historically been achieved partially by wafer diameter migration. The transition from 200 mm to 300 mm occurred at a time when the industry was facing serious economic challenges. This substantially delayed the onset of high-volume manufacturing for that diameter versus the expected timing based on the historical cycle. This has already influenced the timing of the move from 300 mm to 450 mm. While there recently have been some increased activities within the industry to consider draft "standards" for mechanical Si wafers used in precursory handling operations, the industry remains considerably behind schedule for achieving diameter migration in 2012, should that be necessary. Issues related to 450 mm silicon wafer introduction have been compiled separately in a 450 mm position paper available online.

Site Flatness—The industry made a substantial gain in site flatness process capability by going to double-sided polish for 300 mm wafers. Incremental improvements on this basic gain are expected to satisfy IC manufacturers' requirements to approximately the 45 nm technology generation. Continued improvement beyond that point may require the implementation of new flatness-improvement technologies, including those discussed in Figure FEP2 and its accompanying text. However, next generation lithography may strongly impact the actual flatness requirements.

Front End Surface Preparation

The requirements for front-end surface preparation have reached a technological crossroads. The models used to determine the pre-gate clean metrics that have historically driven the ITRS projections for advanced technologies, including critical particle counts and surface metals, have become outdated. Consequently, due to the lack of new forward-looking models, many of the pre-gate clean metrics are nearly constant now and in the near future. Wafer cleaning and surface preparation challenges will evolve as demanded by new technology requirements, driven by the need to effectively clean the ultra-shallow junctions (USJs) and to meet the challenges of new integration schemes, new materials, and new transistor structures. The use of ceria-based slurries for CMP; epitaxial SiGe for raised source/drains and new materials for capacitors; the ability to remove high-dose implanted resist; the removal of small particles without impact to materials; and the need for increasing aspect ratios for contacts will all drive the investigation of new techniques and chemistries for cleaning and drying.

Technology requirements for surface preparation are shown in Tables FEP3a and FEP3b; more details are available in the *supplemental material link.* Also included are the difficult challenges for surface preparation in Table FEP3 . Although Hf-based materials appear ready to be implemented at the 45 and 32 nm technology generations, quantifying front-end surface preparation needs continues to be problematic due to the lack of data associated with future dielectric and gate electrode materials and their properties. Metal gate materials and their integration schemes are still under development; however, the metals that are used in a dual metal CMOS device must still be cleaned to a level that does not affect device performance.

Particulate contamination, both on the front and back surfaces of the wafer, will continue to be a concern at increasingly demanding levels. Control of particle levels without damaging structures or etching material is seen as a formidable challenge. The Maly equation, with its use of a Poisson distribution, continues to be used to determine the allowable defect density of front surface particles based on yield. The "killer defect" size, the critical particle diameter, continues to decrease based on the generation. With die sizes increasing while feature size is decreasing, the model has diverged; no new models are forthcoming due to the lack of data for particles less than 30 nm in diameter. It is anticipated that these future models will be derived by the Yield Enhancement TWG, with more emphasis placed on the yield requirements, to ensure compatibility with the rest of the *Yield Enhancement Roadmap.*

Historically, the critical particle counts have been based on DRAM technology and have fluctuated with changes in DRAM half-pitch (contacted), cell area factor, functions per chip (Gbits), and chip size. The consensus of the Surface Preparation sub-TWG team is that the critical particle count levels employed by device manufacturers should not fluctuate in a corresponding manner. Once a minimum level is reached, specifications should not increase with a change

UMC 1512
UMC v. AICP

in the technology. It should also be noted that the use of DRAM as a technology generator is also under review since the Flash minimum dimensions have accelerated ahead of DRAM by two years.

Back surface and bevel edge defects and particles are being more thoroughly investigated with respect to their effect on yield. Now that equipment is commercially available to detect back surface and edge defects, more data on yield should be forthcoming. However, with an understanding that the lowest level of back surface particles is desirable, there is little data and few models available that can link the size or density of back surface particles to yield on the front surface of the wafer. Although consensus is still limited as to the appropriate specifications for back surface particles based on particle size, it is agreed that particles must not be very large (>50 microns) as they will impact lithographic depth of focus. See the table footnotes for further explanation.

In 2007, the Surface Preparation Sub-TWG team acknowledged that advanced processes did not always present the only valid technology challenges. The 99% yield per process step used by the Maly equation is not sufficient for applications for the medical and automotive industry where a 99.99999% yield (so-called "zero ppm") is required. Although the killer defect is a much larger size for these older technologies (which may lag advanced CMOS by ten years), the surface preparation challenges that must be overcome to meet this yield are as equally daunting and require numbers for defect density far below those in the roadmap. This is a topic to be resolved in future ITRS editions.

Control of particulate contamination will become more challenging as the need to minimize oxide and silicon loss, with zero structural damage, becomes more critical. The oxide and silicon loss requirements have generated the most discussion of any item on the Surface Preparation Roadmap. On planar CMOS devices, this requirement is most critical at the post-gate steps. The pre-gate clean requirement is currently not as critical, but that may change in the future as device structures evolve to MUGFETs and FinFETs. For the post-gate clean steps, the values listed are an average of multiple ash/clean steps on non-damaged, polysilicon and oxide blanket test wafers, not the measurement of a single clean step on device wafers. The numbers also represent the total silicon and oxide loss that may be tolerated for the combination of all post-gate, post-implant mask cleaning steps in which the extension areas are exposed. This requirement will be different from company to company because certain device types will have a significantly different number of post gate-etch cleans and the USJ profiles will be device-specific. DRAM devices may have only four cleans, but system-on-a-chip (SoC) RF and analog devices may have double and triple gates and could be subjected to 12 or more resist strips, not including those associated with possible reworking of the photoresist layers. These devices certainly necessitate the material loss values listed in the table. The presence and type of capping layer will also severely impact the USJ profile and, subsequently, the amount of allowable material loss.

The introduction of silicon-on-insulator (SOI) by 2010 and the implementation of raised source/drains may affect the allowable levels of metal contamination as there is evidence metals may build up at the buried oxide layer interface. It is not yet clear how this will affect allowable metals level, and it has not been accounted for in these tables. Like the critical particle specifications, the metal requirements in previous roadmaps have fluctuated with changes in technology as set forth by the PIDS team. A similar decision was made that the requirements for allowable metal levels will not increase with a change in the technology.

Interface control is expected to become increasingly critical as devices begin to be fabricated with deposited gate dielectric materials and epitaxial Si and SiGe for strained channel formation. Deposited high-κ gate dielectrics may require an oxidized or nitrided surface before deposition, whereas epitaxial Si will require an oxide-free surface. The requirements for surface preparation implemented before high-κ deposition will probably require lower carbon and oxygen requirements than have been used for traditional SiO_2 furnace and rapid thermal steps. High-κ gate dielectrics may also lead to a loosening of requirements for metal contaminant control as gates become physically thicker. The appropriate levels will need to be defined and modeled through processing on a stable baseline; they are currently still under investigation. After gate formation, post-etch cleans must be introduced that are compatible with high-κ dielectrics and metal gate electrodes. This includes preventing corrosion or oxidation, CD loss, and loss or roughness of the metal gate. New MPU and DRAM materials coupled with tightening material budgets will increase the need for highly selective etching chemistries and processes, and these must be introduced without deleterious ESH effects.

There is universal understanding that watermarks and drying-related defects cannot be tolerated on a cleaned surface. Therefore the line item for watermarks, which showed "0" across the table, is deleted in the 2007 roadmap. Yet, drying high aspect ratio structures remains an ongoing issue.

UMC 1512
UMC v. AICP

Table FEP3a Front End Surface Preparation Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*	*Driver*
DRAM ½ Pitch (nm) (contacted)	*65*	*57*	*50*	*45*	*40*	*36*	*32*	*28*	*25*	*D ½*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*68*	*59*	*52*	*45*	*40*	*36*	*32*	*28*	*25*	*M*
MPU Physical Gate Length (nm)	*25*	*23*	*20*	*18*	*16*	*14*	*13*	*11*	*10*	*M*
Wafer diameter (nm)	**300**	**300**	**300**	**300**	**300**	**450**	**450**	**450**	**450**	*D ½, M*
Wafer edge exclusion (mm)	**2**	**2**	**2**	**2**	**2**	**1.5**	**1.5**	**1.5**	**1.5**	*D ½, M*
Front surface particles										
Killer defect density, D_pR_p (#/cm^2) [A]	**0.11**	**0.14**	**0.17**	**0.11**	**0.14**	**0.17**	**0.11**	**0.14**	**0.17**	*D ½*
Critical particle diameter, d_c (nm) [B]	**32.5**	**28.3**	**25.0**	**22.5**	**20.0**	**17.9**	**15.9**	**14.2**	**12.6**	*D ½*
Critical particle count, D_{pw} (#/wafer) [C]	**75.4**	**75.4**	**75.4**	**74.7**	**74.7**	**270.6**	**170.5**	**170.5**	**170.5**	*D ½*
Back surface particle diameter: lithography and measurement tools (µm) [D]	**0.12**	**0.12**	**0.1**	**0.1**	**0.1**	**0.1**	**NA**	**NA**	**NA**	*D ½*
Back surface particles: lithography and measurement tools (#/wafer) [E]	**200**	**200**	**200**	**200**	**200**	**200**	**NA**	**NA**	**NA**	*D ½*
Back surface particle diameter: all other tools (µm) [D]	**0.16**	**0.16**	**0.14**	**0.14**	**0.14**	**0.14**	**NA**	**NA**	**NA**	*D ½*
Back surface particles: all other tools (#/wafer) [E]	**200**	**200**	**200**	**200**	**200**	**200**	**NA**	**NA**	**NA**	*D ½*
Critical GOI surface metals (10^{10} atoms/cm^2) [F]	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	*MPU*
Critical other surface metals (10^{10} atoms/cm^2) [F]	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*MPU*
Mobile ions (10^{10} atoms/cm^2) [G]	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	*MPU*
Surface carbon (10^{13} atoms/cm^2) [H]	**1.2**	**1**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	
Surface oxygen (10^{13} atoms/cm^2) [I]	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	*D ½, M*
Surface roughness LVGX, RMS (Å) [J]	**4**	**4**	**4**	**2**	**2**	**2**	**2**	**2**	**2**	
Silicon and oxide loss (Å) on polysilicon blanket test wafers per LDD clean step—DRAM [K]	**1.5**	**1.2**	**1.2**	◆ **0.9**	◆ **0.9**	◆ **0.9**	◆ **0.6**	◆ **0.6**	◆ **0.6**	*M*
Silicon and oxide loss (Å) on polysilicon blanket test wafers per LDD clean step—Microprocessor/SoC/Analog [L]	**0.5**	**0.4**	**0.4**	◆ **0.3**	◆ **0.3**	◆ **0.3**	◆ **0.2**	◆ **0.2**	◆ **0.2**	*M*

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

UMC 1512
UMC v. AICP

Table FEP3b Front End Surface Preparation Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*	*Driver*
DRAM ½ Pitch (nm) (contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*D ½*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*M*
MPU Physical Gate Length (nm)	*9*	*8*	*7*	*6*	*6*	*5*	*4*	*M*
Wafer diameter (mm)	**450**	**450**	**450**	**450**	**450**	**450**	**450**	*D ½, M*
Wafer edge exclusion (mm)	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**	*D ½, M*
Front surface particles								
Killer defect density, D_pR_p (#/cm^2) [A]	**0.11**	**0.14**	**0.17**	**0.11**	**0.14**	**0.17**	**0.11**	*D ½*
Critical particle diameter, d_c (nm) [B]	**11.3**	**10.0**	**8.9**	**8.0**	**7.1**	**6.3**	**5.6**	*D ½*
Critical particle count, D_{pw} (#/wafer) [C]	**170.5**	**170.5**	**170.5**	**170.5**	**170.5**	**170.5**	**170.5**	*D ½*
Back surface particle diameter: lithography and measurement tools (µm) [D]	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	*D ½*
Back surface particles: lithography and measurement tools (#/wafer) [E]	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	*D ½*
Back surface particle diameter: all other tools (µm) [D]	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	*D ½*
Back surface particles: all other tools (#/wafer) [E]	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	*D ½*
Critical GOI surface metals (10^{10} atoms/cm^2) [F]	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	*MPU*
Critical other surface metals (10^{10} atoms/cm^2) [F]	**1**	**1**	**1**	**1**	**1**	**1**	**1**	*MPU*
Mobile ions (10^{10} atoms/cm^2) [G]	**2**	**2**	**2**	**2**	**2**	**2**	**2**	*MPU*
Surface carbon (10^{13} atoms/cm^2) [H]	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	
Surface oxygen (10^{13} atoms/cm^2) [I]	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	*D ½, M*
Surface roughness LVGX, RMS (Å) [J]	**2**	**2**	**2**	**2**	**2**	**2**	**2**	
Silicon and oxide loss (Å) on polysilicon blanket test wafers per LDD clean step—DRAM [K]	◆**0.6**	◆**0.6**	◆**0.6**	◆**0.6**	◆**0.6**	◆**0.6**	◆**0.6**	*M*
Silicon and oxide loss (Å) on polysilicon blanket test wafers per LDD clean step—Microprocessor/SoC/Analog [L]	◆**0.2**	◆**0.2**	◆**0.2**	◆**0.2**	◆**0.2**	◆**0.2**	◆**0.2**	*M*

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

Tables FEP3a and FEP3b Notes:

[A] Killer defect density is calculated from the formula for 99% yield, $Y=0.99=\exp[-D_pR_pA_{eff}]$. A_{eff} is the effective chip area, D_p is the defect density, and R_p is a defect kill factor indicating the probability that a given defect will kill the device. The product D_pR_p is the density of device-killing defects on the wafer. R_p is dependent on numerous things including the size and shape of the particle, the composition of the particle, and specifics of the device layout. In previous years, R_p was assumed to be 0.2 for any particle > the critical particle size, d_c. For DRAM, $A_{eff}= 2.5F^2T+(1-aF^2T/A_{chip})*0.6A_{chip}$. where F is the minimum feature size, a is the cell fill factor, T is the number of DRAM bits (transistors) per chip, and A_{chip} is the DRAM chip size. For MPUs, $A_{eff}=aT(GL)^2$, where GL is the gate length. Because A_{eff} can increase or decrease with each successive technology generation, D_pR_p does not always decrease over time.

[B] Critical particle diameter, d_c, is defined by Yield Enhancement as ½ of the metal ½-pitch dimension. This should be considered an "effective" particle diameter as most particulate contamination is irregular in shape.

[C] An example is provided which assumes that the kill factor, R_p, is 0.2 for all particles larger than the critical particle size. This is the assumption made in previous versions of the roadmap, but is not universally valid and is included only for purposes of an example calculation. Particles/wafer is calculated using $[R_p*3.14159*(\text{wafer radius-edge exclusion})^2]$. To convert from particles/wafer at the critical particle size to particles/wafer at an alterative size, a suggested conversion formula is: $D_{alternate}=D_{critical}*(d_{critical}/d_{alternate})^2$. The calculated values actually vary with DRAM technology, but the critical particle count levels employed by device manufacturers should not fluctuate in a corresponding manner. Once a minimum level is reached, the values do not increase with a change in the technology.

UMC 1512
UMC v. AICP

[D] and [E] While there are some experimental models and empirical data available- and future tables may use these models, there is as yet no industry agreement regarding the number or sizes of back surface particles which could be deleterious to semiconductor processing consequently, the back surface contact specs are based on present day realistic expectations and future aggressive lithographic improvements. In the past, arguments have been made that back surface particles affect device yield mainly at the lithographic steps by causing the front surface of the wafer to move out of the focal plane leading to critical dimension variations. However, it is not clear how the limited back surface contact achievable with pin chucks interacts with back surface particle density to cause front surface flatness variations. In addition, it is also not clear how lithographic depth-of-focus (DOF) will change from year to year as this is not specified in the lithography roadmap. Aggressive specs for litho/measurement tool may necessitate edge-grip or edge-contact handling only.

It is not possible to measure absolute levels of back surface particles on in-process wafers due to large variations in back surface finish and films. A generally accepted practice is to process wafers with the polished front surface down in order to assess back surface particle adders for a particular process or operation.

Back surface particle metrics are for wafers with 3mm edge exclusions. These metrics reflect the TOTAL number of touches in any given front end of line process tool.

[F] In past roadmaps, metal contamination targets have been based on an empirically derived model determining failure due to metal contamination as a function of gate oxide thickness. However, the oxides used in the experiments from which this model was derived were far thicker than gate oxide thicknesses used today. More recent data suggest an updated approach is appropriate. The metals are empirically grouped into three classes[5, 6]: (a) Mobile metals which may be easily removed such as Na and K and may be modeled by taking the flat-band shift of a capacitance-voltage (CV) test less than or equal to 50 mV; (b) metals which dissolve in silicon or form silicides such as Ni, Cu, Cr, Co, Hf, and Pt; and (c) major gate-oxide-integrity (GOI) killers such as Ca, Ba, and Sr. Metals such as Fe may fall into both classes (b) and (c). Targets for mobile ions are based on allowable threshold voltage shift from a CV test. Current targets for GOI killers and other metals are based on empirical data.[7] There may be reason to indicate less stringent targets because effects should scale with respect to physical dielectric thickness (not EOT) that will increase upon introduction of high-κ gate dielectrics. However, in the absence of data to corroborate such, as well as that of physical dielectric thickness, the targets are left constant for future years. Finally, the introduction of SOI may also affect the allowable levels of metal contamination as there is evidence that some metals may build up while others diffuse through the buried oxide layer interface. It is not yet clear how this will affect allowable metals level and has not been accounted for in these tables. Another factor to be considered in future years will be the spatial distribution of localized contamination as opposed to the average contamination per wafer.

[G] The model for mobile ions, D_i, calculates the number of ions that will create a threshold voltage shift that is within a portion of the Allowable Threshold Voltage Variability (ATVV). ATVV was specified in Row 15 of Table 28a in the 1999 ITRS, but is no longer specified. For the purposes of the mobile ion model in 2003, it is assumed that the ATVV is 3% of the Nominal Power Supply Voltage for Low Operating Power or Low Standby Power Technology (see PIDS chapter). The portion of ATVV allocated to mobile ions is assumed to be 5%. Therefore, $D_i=1/q(C_{gate}*ATVV*0.05)$, where C_{gate} is computed for an electrically equivalent SiO_2 gate dielectric thickness and q is the charge of an electron. This model reduces to $D_i = ((3.9*8.85)/1.6)*(0.05*ATVV/EOT)*10^9$, where ATVV is in mV and EOT is in nm (also from Low Operating Power or Low Standby Power Technology Requirements Table in the PIDS chapter) , and the oxide dielectric constant is 3.9. Note that the year-to-year value for D_i does not always decrease because D_i is not only proportional to ATVV, but is also inversely proportional to EOT.

[H] Residual carbon resulting from organic contamination after surface preparation. Surface Carbon at the 180 nm corresponded to 10% carbon atom coverage of a bare silicon wafer (7.3E+13 atoms/cm^2). Surface carbon for subsequent generations was scaled linearly with the ratio of CD (½ DRAM ½ pitch) to 180 nm. D_c=(CD/180)(7.3E+13)

[I] Surface oxygen requirements at <1E+12 atoms/cm^2 are driven by the needs of pre-epitaxial cleaning. Epitaxial deposition of Si and SiGe is used for some devices, now, and will find more widespread use with the implementation of strained silicon channel technology. While some level of oxide can be removed in situ, prior to epitaxial deposition, the trend towards lower deposition temperature will preclude the use of a higher temperature hydrogen pre-bake. Surface oxygen concentrations up to <1E+13 atoms/cm^2 are acceptable for processes such as pre-silicide cleaning. Current pre-gate cleaning does not require an oxide-free surface, but the pre-gate surface should be either fully passivated by a continuous oxide layer or have <1E+13/cm^2, as an intermediate level of oxygen will be unstable. It is unclear whether high-κ gate dielectrics will require oxide-free or oxide-passivated surfaces prior to deposition.

[J] In the 2001 ITRS, it was assumed channel mobility cannot be degraded by >10% due to surface preparation induced surface roughness. It was further claimed that current technologies were successfully manufactured with AFM based determination of 2 Å RMS of surface micro-roughness. Where this is still approximately true for surface preparation induced, i.e. additive roughness, it is more direct to simply measure roughness on product immediately after the low voltage gate oxidation (LVGX) pre-clean. In this case, the total surface micro-roughness takes into account starting substrate roughness, plus the additional micro roughness induced by pre-cleans and strips of initial oxidation, any implant screen oxidations, dummy or sacrificial oxidations the first portions of the high voltage gate oxidation (dual gate flows) and any additional roughness brought about by plasma nitridations. With this taken into account, product has recently been successfully built with 4 Å RMS surface micro-roughness. This may in part be explained by TCAD indications that show carrier mobility being mainly affected by spatial frequencies smaller than those that are typically sampled by AFM micro-roughness metrology tools. The reason the required number is not scaling over time near term is based off comparisons to the PIDS roadmap where carrier mobility is held constant over this same time period.

[K] The values for silicon and oxide loss are driven by requirements of DRAM devices and represent the total silicon and oxide loss that may be tolerated for the combination of all post-gate, post-implant mask cleaning steps in the portion of the flow where source/drain extensions are fabricated, typically four in a DRAM process flow. Specific values are relative to silicon and oxide loss measured optically on blanket polysilicon or oxide test wafers. Actual consumption on product will vary driven by damage from plasma etch/ash, ion implantation, and dopant concentrations. Decreasing values are in response to the requirement to control negative impact on drive currents (Ids). If the silicon under the source/drain extensions is recessed, this changes the junction profile increasing the source/drain extension resistance and decreasing drive currents. By not consuming the oxide, assuming similar processing, this reduces the ability of subsequent processes to further oxidize and consume silicon. Less oxidized silicon equates to less silicon recess under the source/drain extensions. It is not yet possible to express a rigorous model connecting this metric with table parameters such as implant mask levels, junction depth, and critical dimension. IC manufacturers are currently targeting silicon loss to be 1.5 Å per cleaning step for 65 nm technologies and 1.2 Å per cleaning step for 57 nm technologies. Other technology values are extrapolated or interpolated from those two values. It is not clear what will be required or possible in the longer term years, so the value is set at 0.9 Å in 2010 and held constant after that.

UMC 1512
UMC v. AICP

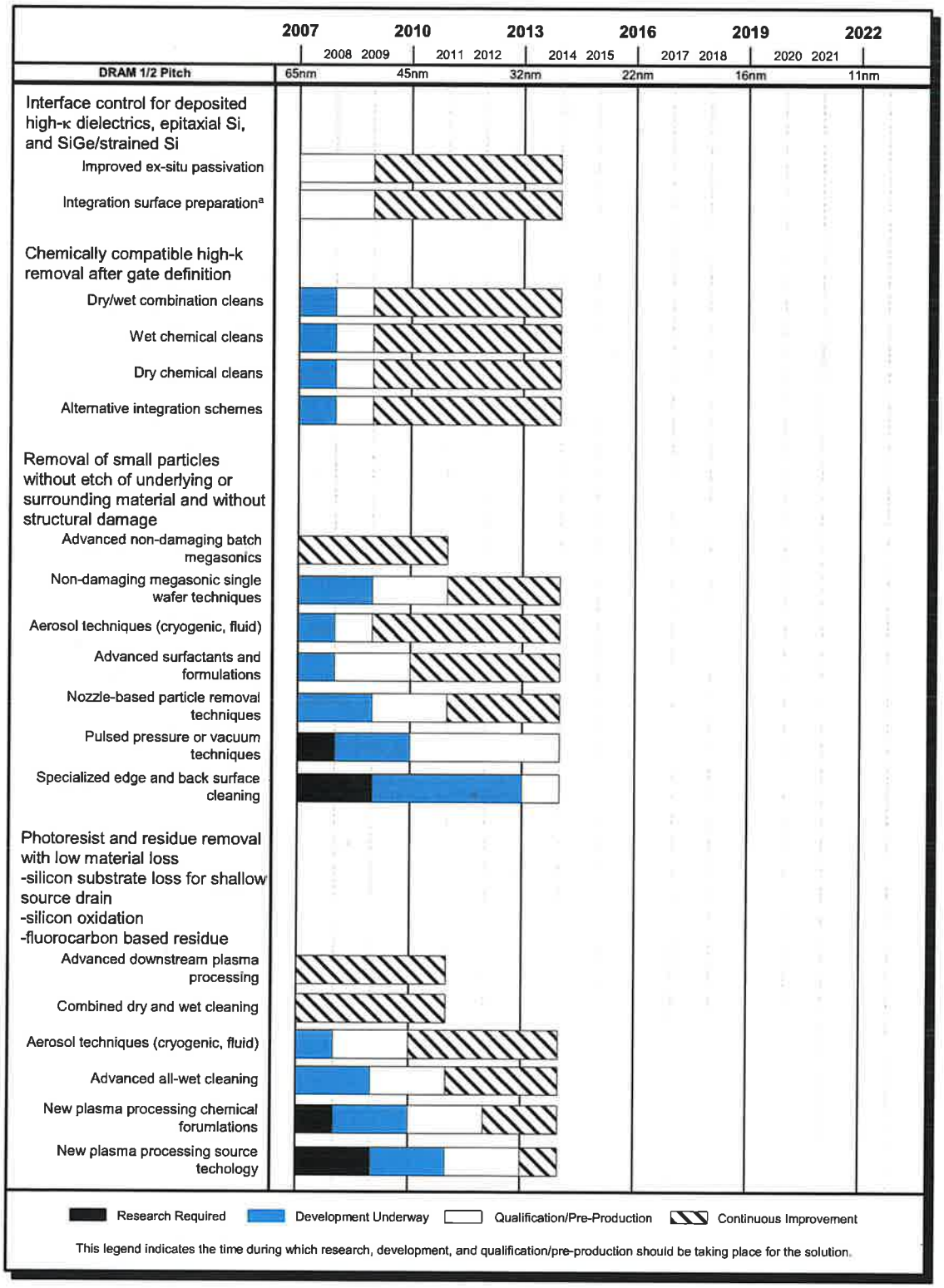
[L] The values for silicon and oxide loss are driven by requirements of system-on-a-chip (SoC) RF and analog devices and represent the total silicon and oxide loss that may be tolerated for the combination of all post-gate, post-implant mask cleaning steps in the portion of the flow where source/drain extensions are fabricated. This could be as many as twelve in an analog process flow. Specific values are relative to silicon and oxide loss measured optically on blanket polysilicon or oxide test wafers. Actual consumption on product will vary driven by damage from plasma etch/ash, ion implantation, and dopant concentrations. Decreasing values are in response to the requirement to control negative impact on drive currents (Ids). If the silicon under the source/drain extensions is recessed, this changes the junction profile increasing the source/drain extension resistance and decreasing drive currents. By not consuming the oxide, assuming similar processing, this reduces the ability of subsequent processes to further oxidize and consume silicon. Less oxidized silicon equates to less silicon recess under the source/drain extensions. It is not yet possible to express a rigorous model connecting this metric with table parameters such as implant mask levels, junction depth, and critical dimension. IC manufacturers are currently targeting silicon loss to be 0.5 Å per cleaning step for 65 nm technologies and 0.4 Å per cleaning step for 57 nm technologies. Other technology values are extrapolated or interpolated from those two values. It is not clear what will be required or possible in the longer term years, so the value is set at 0.3 Å in 2010 and held constant after that.

Surface preparation challenges along with potential solutions are shown in Figure FEP3. Wet chemical critical cleaning is still favored because many inherent properties of liquid solutions facilitate the removal of metals (high solubility in liquid chemistries) and particles (zeta-potential control, shear stress, and efficient energy transfer by megasonics). The need for other techniques will arise, however, to provide interfacial control for advanced gates as well as non-etching, damage-free particle removal. At this time, the development of broadly effective and non-damaging particle and residue removal using liquid and non-liquid techniques is well underway. Chemical cleaning is being augmented by a combination of physical and chemical methods, such as megasonic agitation, nozzle-based cleaning, and the use of surfactants. Single wafer cleaning, both "wet" and "dry" (including cryogenic aerosol), is expected to be increasingly implemented due to process integration and cycle time concerns, but it remains unclear when its use will become widespread in front end of line processes. Pre-gate cleaning of single wafers is gaining favor as a viable alternative to traditional batch cleaning, as the evolution to 300 mm processing in advanced lines has led to more widespread usage of single wafer thermal and deposition processes for gate stack formation and the need for tighter coupling of the production processes.

Other techniques for cleaning, such as laser, electrostatics, and other novel processes, are experiencing a high level of research and development and, if implemented, most likely will be on a single wafer system. The chemistries used for cleaning will continue to evolve. Dilute chemistries, especially dilute RCA style cleans, have shown feasibility and are still used in production, although not necessarily as traditional RCA1-RCA2 cleans. For example, a dilute HCl rinse has replaced SC2 in many applications. Ultra-dilute SC-1 or NH_4OH in water alone are being researched. Because they can minimize attack of the oxide and silicon surfaces, dilute chemistries have gained greater acceptance in most advanced fabs. Ozonated water processes are being implemented as replacements for some sulfuric acid-based resist strips and post-cleans. To address the materials loss issues associated with stripping and cleaning implanted photoresist, there is an increasing trend towards non-ashing processes. However, plasma stripping and cleaning processes are also evolving; new gas formulations and new source technologies are demonstrating promising results. The presence of metal in the gate stack is driving interest and research with solvent chemistries instead of traditional SPM. Wet-only processes for resist stripping may not be feasible for all situations, and research is ongoing to enable non-ashing resist removal with technologies such as cryogenic aerosol treatments. Material capability issues will need to be characterized and understood. Finally the emergence of immersion lithography and new resist formulations for advanced lithography will raise additional new cleans challenges.

Potential solutions are indicated for only the near-term years (through 2013, for 32 nm) as it is unclear what challenges will arise for surface preparation at 22 nm. Cleaning and measuring vertical surfaces as well as SiGe and III-V surfaces will be a new challenge. As in the past, current and future surface preparation processes are expected to be the subject of continuous improvement efforts.

UMC 1512
UMC v. AICP



a) Integrated Surface Preparation techniques include various techniques that can be coupled to the deposition chamber and allow processing to continue with minimal exposure to the atmosphere. This includes, but is not limited to, UV-based cleaning, gas phase techniques, and single wafer wet techniques.

Figure FEP3 Front End Surface Preparation Potential Solutions

UMC 1512
UMC v. AICP

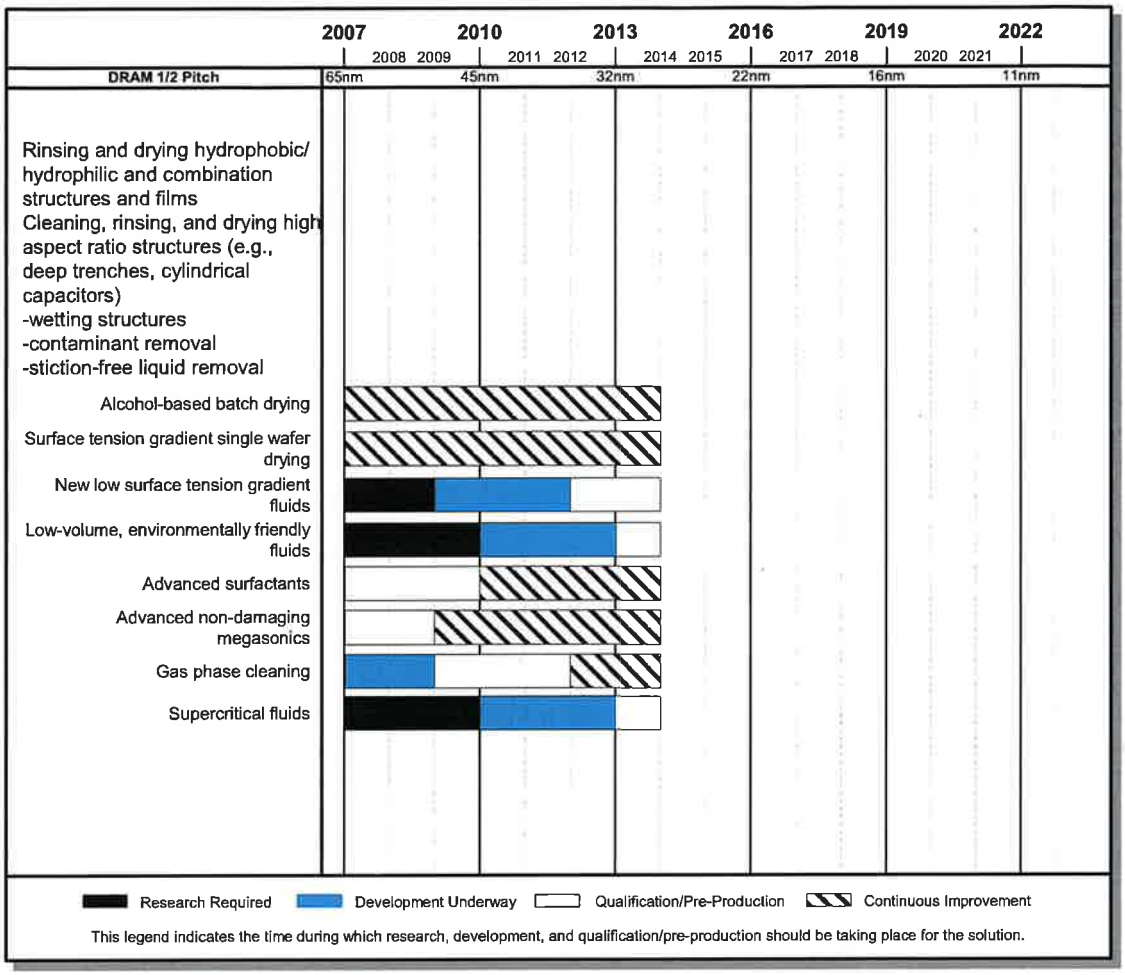


Figure FEP3 Front End Surface Preparation Potential Solutions (continued)

Other technologies, such as ESH and Yield Enhancement, overlap surface preparation. Reduced chemical usage, chemical and water recycling, and alternative processes using less harmful chemistries can offer ESH and CoO benefits. Efforts in chemical and water usage reduction should continue. Automated process monitoring and control can also reduce CoO; their increased use is expected particularly for 300 mm and larger wafer sizes for which the cost of monitor wafers and process excursions can become excessive. New cleaning requirements will arise related to immersion lithography, but they will be tied to the implementation of that lithographic method and should be itemized by the Lithography TWG in the future. Surface preparation overlaps with defect reduction technology in the need for defining appropriate purity levels in chemicals and DI water. To minimize CoO, aggressive purity targets should be adopted only when there is technological justification. In all areas of surface preparation, a balance must be achieved among process and defect performance, cost, and ESH issues. Refer to the *Environment, Safety, and Health* chapter for a comprehensive overview.

THERMAL/THIN FILMS, DOPING, AND ETCHING

Front end processing requires the growth, deposition, etching and doping of high quality, uniform, defect-free films. These films may be insulators, conductors, or semiconductors (for example, silicon). The difficult challenges in front end processing include: (1) the growth or deposition of reliable very thin (with electrical equivalent thickness ≤ 1.0 nm) gate dielectric layers; (2) the development of alternate high dielectric constant layers, including suitable interface layers, for both logic and DRAM capacitor applications; (3) the development of depletion-free, low-resistivity gate electrode materials, (4) the development of reliable processes to enhance the channel mobility in both NMOS and PMOS devices through channel strain, (5) the formation of low resistivity contacts to ultra-shallow junction devices, and (6) the

UMC 1512
UMC v. AICP

development of resist trim and gate etch processes that provide excellent CD control. Other important challenges include the achievement of abrupt channel doping profiles, defect management for minimum post-implant leakage under reduced thermal budget environments, and formation of precise sidewall spacer structures.

An array of "Technology Innovations" are expected to be required to sustain the trend for increased transistor performance for deeply scaled devices, as detailed in the PIDS Chapter. Strained-Si channels have recently been introduced (to boost carrier mobility and drive currents), and work continues to assess the limits of alternative methods to introduce strain. One of the more dramatic announcements this year is the planned deployment of high-κ gate dielectrics (to reduce gate leakage and control short channel effects) and metallic gates (to eliminate doped polysilicon depletion effects that limit practical scaling of gate stack layers) by 2008. Even with the successful introduction of these new materials and structures, the limitations of planar bulk CMOS transistors, particularly the increased sub-threshold leakage currents exhibited at reduced threshold and drive voltages, will drive the introduction of new device structures such as fully depleted silicon (or Ge) on insulator and multi-gate transistors. This rapid introduction of new materials and device structures in the next five to seven years constitutes an unprecedented multiplicity of challenges to develop, and also to integrate these developments into effective, cost-efficient production technologies. During this period of transition, the plethora of choices for the device structure seems likely to lead to some divergence within the industry, some companies choosing to aggressive scale bulk CMOS parameters, while others make the transition to FDSOI and multi-gate structures where the requirements may be less stringent. The thermal and thin film, doping, and etching requirements are given in Tables FEP4a and FEP4b.

Thermal/Thin-Films

The gate dielectric has emerged as one of the most *difficult challenges* for future device scaling. Requirements summarized in Tables FEP4a and FEP4b indicate an equivalent oxide thickness progressing to substantially less than 1 nm. Direct tunneling currents and boron penetration (from the polysilicon layer) preclude the use of silicon oxynitride dielectric layers below about 1 nm thickness. Even in high-performance applications that have high allowable leakage, progress in scaling oxynitrides to 1 nm and below seems to have stalled since the 2003 ITRS, largely because of high leakage currents. Fortunately, the implementation of enhanced-mobility channels has delayed the need for high-κ dielectrics by a couple of years. Nevertheless higher dielectric constant materials are expected as soon as 2008. At the same time depletion-free, metal-gate electrodes are needed. For low-power applications where the allowable gate leakage is very low, higher dielectric constant materials may also be needed as early as 2008, albeit while still using poly-Si gate electrodes. In the meantime, the near-term gate dielectric solution requires the fabrication and use of ultra-thin silicon oxynitride films.

UMC 1512
UMC v. AICP

Table FEP4a Thermal, Thin Film, Doping and Etching Technology Requirements—Near-term Years

Grey cells indicate the requirements projected only for intermediate, or long-term years. Near-term line items are not included.

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*68*	*59*	*52*	*45*	*40*	*36*	*32*	*28*	*25*
MPU Physical Gate Length (nm)	*25*	*23*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
Equivalent physical oxide thickness for bulk MPU/ASIC T_{ox} (nm) for 1E20-doped poly-Si [A, A1, A2]	1								
Equivalent physical oxide thickness for bulk MPU/ASIC T_{ox} (nm) for 1.5E20-doped poly-Si [A, A1, A2]	1.1	0.5							
Equivalent physical oxide thickness for bulk MPU/ASIC T_{ox} (nm) for 3E20-doped poly-Si [A, A1, A2]	1.2	0.71	0.54	0.41					
Equivalent physical oxide thickness for bulk MPU/ASIC T_{ox} (nm) for metal gate [A, A1, A2]		0.9	0.75	0.65	0.55	0.5			
Gate dielectric leakage at 100 °C (A/cm^2) bulk high-performance [B, B1, B2]	8.0E+02	8.7E+02	1.0E+03	1.1E+03	1.3E+03	1.4E+03			
Metal gate work function for bulk MPU/ASIC $\lvert E_{c,v} - \phi_m \rvert$ (eV) [C]		<0.2	<0.2	<0.2	<0.2	<0.2			
Channel doping concentration (cm^{-3}), for bulk design [D]	4.8E+18	3.7E+18	4.1E+18	5.4E+18	6.6E+18	8.4E+18			
Bulk/FDSOI/DG – Long channel electron mobility enhancement factor due to strain for MPU/ASIC [E]	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
Drain extension X_j (nm) for bulk MPU/ASIC [F]	12.5	11	10	9	8	7			
Maximum allowable parasitic series resistance for bulk NMOS MPU/ASIC × width ((Ω–μm) from PIDS [G]	200	200	200	180	180	180			
Maximum drain extension sheet resistance for bulk MPU/ASIC (NMOS) (Ω/sq) [G]	650	740	810	900	1015	1160			
Extension lateral abruptness for bulk MPU/ASIC (nm/decade) [H]	2.5	2.3	2.0	1.8	1.6	1.4			
Contact X_i (nm) for bulk MPU/ASIC [I]	27.5	25.3	22	19.8	17.6	15.4			
Allowable junction leakage for bulk MPU/ASIC (μA/μm)	0.34	0.71	0.7	0.64	0.74	0.68			
Sidewall spacer thickness (nm) for bulk MPU/ASIC [J]	27.5	25.3	22	19.8	17.6	15.4			
Maximum silicon consumption for bulk MPU/ASIC (nm) [K]	13.8	12.7	11	9.9	8.8	7.7			
Silicide thickness for bulk MPU/ASIC (nm) [L]	17	15	13	12	11	9			
Contact silicide sheet R_s for bulk MPU/ASIC (Ω/sq) [M]	9.6	10.5	12.1	13.5	15.1	17.3			
Contact maximum resistivity for bulk MPU/ASIC (Ω-cm^2) [N]	1.2E-07	1.0E-07	9.2E-08	7.0E-08	6.2E-08	5.6E-08			
STI depth bulk (nm) [O]	353	339	335	331	323	316			
Trench width at top (nm) [P]	65	57	50	45	40	35			
Trench sidewall angle (degrees) [Q]	>87.4	>87.6	>87.9	>88.1	>88.2	>88.4			
Trench fill aspect ratio – bulk [R]	6.0	6.5	7.2	7.9	8.6	9.5			
Equivalent physical oxide thickness for FDSOI MPU/ASIC T_{ox} (nm) for metal gate [A, A1, A2]				0.7	0.6	0.55	0.5	0.5	0.5
Gate dielectric leakage at 100°C (A/cm^2) FDSOI high-performance [B, B1, B2]				1.1E+03	1.3E+03	1.4E+03	1.5E+03	1.8E+03	2.0E+03

UMC 1512
UMC v. AICP

Table FEP4a Thermal, Thin Film, Doping and Etching Technology Requirements—Near-term Years

Grey cells indicate the requirements projected only for intermediate, or long-term years. Near-term line items are not included.

Year of Production	2007	2008	2009	2010	2011	2012	2013	2014	2015
Metal gate work function for FDSOI MPU/ASIC $\phi_m - E_i$ (eV) NMOS/PMOS [S]				±0.15	±0.15	±0.15	±0.15	±0.15	±0.15
Saturation velocity enhancement factor MPU/ASIC [T]	1	1.1	1.1	1.1	1*	1*	1*	1*	1*
Si thickness FDSOI (nm) from PIDS [T]				5.5	5.2	4.5	4	3.5	3.2
Maximum allowable parasitic series resistance for FDSOI NMOS MPU/ASIC × width ((Ω–μm) [G]				180	180	180	170	160	160
Maximum drain extension sheet resistance for FDSOI MPU/ASIC (NMOS) (Ω/sq) [G]				730	770	890	1000	1150	1250
Spacer thickness, FDSOI elevated contact [J]				9.9	8.8	7.7	7.2	6.1	5.5
Thickness of FDSOI elevated junction (nm) [U]				18	16	14	13	11	10
Maximum silicon consumption for FDSOI MPU/ASIC (nm) [K]				18	16	14	13	11	10
Silicide thickness for FDSOI MPU/ASIC (nm) [L]				22	19	17	16	13	12
Contact silicide sheet R_s for FDSOI MPU/ASIC (Ω/sq) [M]				7.4	8.3	9.5	10.2	12.1	13.3
Contact maximum resistivity for FDSOI MPU/ASIC (Ω-cm^2) [N]				7.2E-08	6.5E-08	5.8E-08	4.8E-08	4.0E-08	3.5E-08
Trench fill aspect ratio – FDSOI [V]				0.6	0.6	0.6	0.6	0.6	0.6
Equivalent physical oxide thickness for multi-gate MPU/ASIC T_{ox} (nm) for metal gate [A, A1, A2]					0.8	0.7	0.6	0.6	0.6
Gate dielectric leakage at 100°C (nA/μm) muti-gate high-performance [B, B1, B2]					1.25E+03	1.43E+03	1.54E+03	1.82E+03	2.00E+03
Metal gate work function for multi-gate MPU/ASIC [S]					midgap	midgap	midgap	midgap	midgap
Si thickness for multi-gate (nm) from PIDS [U]					9.5	8.5	7.5	6.5	6
Maximum allowable parasitic series resistance for multi-gate NMOS MPU/ASIC × width ((Ω–μm) [G]					180	180	170	160	160
Maximum drain extension sheet resistance for multi-gate MPU/ASIC (NMOS) (Ω/sq) [G]					425	475	535	615	670
Spacer thickness, multi-gate elevated contact [J]					8.8	7.7	7.2	6.1	5.5
Thickness of multi-gate elevated junction (nm) [T]					16	14	13	11	10
Maximum silicon consumption for multi-gate MPU/ASIC (nm) [K]					16	14	13	11	10
Silicide thickness for multi-gate MPU/ASIC (nm) [L]					19	17	16	13	12
Contact silicide sheet R_s for multi-gate MPU/ASIC (Ω/sq) [M]					8.3	9.5	10.2	12.1	13.3
Contact maximum resistivity for multi-gate MPU/ASIC (Ω-cm^2) [N]					6.6E-08	6.1E-08	5.1E-08	4.2E-08	3.8E-08
Physical gate length low operating power (LOP) (nm)	32	28	25	23	20	18	16	14	13
Equivalent physical oxide thickness for bulk low operating power T_{ox} (nm) for 1.5E20-doped poly-Si [A, A1, A2]	1.2	0.8	0.7	0.6	0.5	0.5			

UMC 1512
UMC v. AICP

Table FEP4a Thermal, Thin Film, Doping and Etching Technology Requirements—Near-term Years

Grey cells indicate the requirements projected only for intermediate, or long-term years. Near-term line items are not included.

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
Equivalent physical oxide thickness for bulk low operating power T_{ox} (nm) for metal gate [A, A1, A2]		1.1	1	0.9	0.8	0.8			
Gate dielectric leakage at 100°C for bulk (A/cm^2) LOP [B, B1, B2]	7.8E+01	8.9E+01	1.0E+02	1.1E+02	1.3E+02	1.4E+02			
Metal gate work function for bulk low operating power $\lvert E_{c,v} - \phi_m \rvert$ (eV) [S]		<0.2	<0.2	<0.2	<0.2	<0.2			
Allowable junction leakage for bulk LSTP (pA/μm)	10	10	10	10	16	21			
Equivalent physical oxide thickness for FDSOI low operating power T_{ox} (nm) for metal gate [A, A1, A2]					0.9	0.9	0.8	0.8	0.8
Gate dielectric leakage at 100°C for FDSOI (A/cm^2) LOP [B, B1, B2]					1.3E+02	1.4E+02	1.6E+02	1.8E+02	1.9E+02
Metal gate work function for FDSOI and multi-gate LOP [S]					midgap	midgap	midgap	midgap	midgap
Equivalent physical oxide thickness for multi-gate low operating power T_{ox} (nm) for metal gate [A, A1, A2]					0.9	0.9	0.9	0.8	0.8
Gate dielectric leakage at 100°C for multi-gate (A/cm^2) LOP [B, B1, B2]					1.3E+02	1.4E+02	1.6E+02	1.8E+02	1.9E+02
Physical gate length low standby power (LSTP) (nm)	45	37	32	28	25	23	20	18	16
Equivalent physical oxide thickness for bulk low standby power T_{ox} (nm) for 1.5E20-doped poly-Si [A, A1, A2]	1.9	1.2	1.1	1	0.9	0.8	0.7		
Equivalent physical oxide thickness for bulk low standby power T_{ox} (nm) for metal gate [A, A1, A2]		1.6	1.5	1.4	1.3	1.2	1.1		
Gate dielectric leakage at 100°C for bulk (A/cm^2) LSTP [B, B1, B2]	6.7E-02	8.1E-02	9.4E-02	1.1E-01	1.2E-01	1.3E-01	1.5E-01		
Metal gate work function for bulk LSTP $\lvert E_{c,v} - \phi_m \rvert$ (eV) [S]		<0.2	<0.2	<0.2	<0.2	<0.2	<0.2		
Equivalent physical oxide thickness for FDSOI low standby power T_{ox} (nm) for metal gate [A, A1, A2]						1.3	1.2	1.1	1.1
Gate dielectric leakage at 100°C for FDSOI (A/cm^2) LSTP [B, B1, B2]						1.3E-01	1.5E-01	1.7E-01	1.9E-01
Metal gate work function for FDSOI and multi-gate LSTP $\phi_m - E_i$ (eV) NMOS/PMOS [S]						± 0.1	± 0.1	± 0.1	± 0.1
Equivalent physical oxide thickness for multi-gate low standby power T_{ox} (nm) for metal gate [A, A1, A2]						1.4	1.3	1.2	1.1
Gate dielectric leakage at 100°C for multi-gate (A/cm^2) LSTP [B, B1, B2]						1.3E-01	1.5E-01	1.7E-01	1.9E-01
Thickness control EOT (% 3σ) [W]	<±4	<±4	<±4	<±4	<±4	<±4	<±4	<±4	<±4
Poly-Si or metal gate electrode thickness (approximate) (nm) [X]	50	46	40	36	32	28	26	22	20
Gate etch bias (nm) [Y]	17	15	14	12	11	10	8	8	7
L_{gate} 3σ variation (nm) [Z]	3	2.76	2.4	2.16	1.92	1.68	1.56	1.32	1.2
Total maximum allowable lithography 3σ (nm) [AA]	2.60	2.39	2.08	1.87	1.66	1.45	1.35	1.14	1.04
Total maximum allowable etch 3σ (nm), including photoresist trim and gate etch [AA]	1.5	1.38	1.2	1.08	0.96	0.84	0.78	0.66	0.6
Resist trim maximum allowable 3σ (nm) [AB]	0.87	0.8	0.69	0.62	0.55	0.48	0.45	0.38	0.35

UMC 1512
UMC v. AICP

Table FEP4a Thermal, Thin Film, Doping and Etching Technology Requirements—Near-term Years

Grey cells indicate the requirements projected only for intermediate, or long-term years. Near-term line items are not included.

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
Gate etch maximum allowable 3σ (nm) [AB]	1.22	1.13	0.98	0.88	0.78	0.69	0.64	0.54	0.49
CD bias between dense and isolated lines [AC]	≤15%	≤15%	≤15%	≤15%	≤15%	≤15%	≤15%	≤15%	≤15%
Minimum measurable gate dielectric remaining (post gate etch clean) [AD]	>0	>0	>0	>0	>0	>0	>0	>0	>0
Profile control (side wall angle) [AE]	90	90	90	90	90	90	90	90	90
Allowable threshold voltage variation from charge in dielectric (mV) [AF]	11	9.5	10	10	10	9	9	9	8.5
Allowable interfacial charge in high-κ gate stack (cm^{-2})[AG]		1.54E+11	1.54E+11	1.41E+11	1.62E+11	1.76E+11	1.67E+11	2.00E+11	2.00E+11
Allowable bulk charge in high-κ gate stack (cm^{-3}) [AH]		5.50E+17	5.50E+17	5.42E+17	6.74E+17	8.02E+17	7.58E+17	8.90E+17	8.90E+17
Allowable bulk charge in high-κ gate stack (ppm) [AH]		25.0	25.0	24.7	30.6	36.5	34.4	40.5	40.5
Allowable critical metal impurity level in high-κ dielectric (ppm) [AI]		2.5	2.5	2.5	3.1	3.6	3.4	4.1	4.1
Allowable critical metal impurity level in high-κ dielectric (ppm) [AJ]	2.2	2.5	2.5	2.5	3.1	3.6	3.4	4.1	4.1

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

UMC 1512
UMC v. AICP

Table FEP4b Thermal, Thin Film, Doping and Etching Technology Requirements—Long-term Years

Grey cells indicate the requirements projected only for intermediate, or long-term years. Near-term line items are not included.

Year of Production	2015	2016	2017	2018	2019	2020	2021	2022	Driver
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	25	22	20	18	16	14	13	11	MPU
MPU Physical Gate Length (nm)	10	9	8	7	6.3	5.6	5.0	4.5	MPU
Bulk/FDSOI/DG – Long channel electron mobility enhancement factor for MPU/ASIC [E]	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	MPU/ASIC
Equivalent physical oxide thickness for multi-gate MPU/ASIC T_{ox} (nm) for metal gate [A, A1, A2]	0.6	0.55	0.55	0.55	0.5	0.5	0.5	0.5	MPU/ASIC Multigate
Gate dielectric leakage at 100°C (nA/μm) muti-gate High-performance [B, B1, B2]	2.0E+03	2.2E+03	2.5E+03	2.9E+03	3.3E+03	3.3E+03	4.0E+03	4.4E+03	MPU/ASIC Multigate
Metal gate work function for multi-gate MPU/ASIC [S]	midgap	midgap	midgap	midgap	midgap	midgap	midgap	midgap	MPU/ASIC Multigate
Si thickness for multi-gate (nm) [T]	6.0	5.4	4.5	4.2	3.8	3.5	3.2	3.0	Multigate
Maximum allowable parasitic series resistance for multi-gate NMOS MPU/ASIC × width ((Ω–μm) from PIDS [G]	160	155	150	145	145	145	135	135	MPU/ASIC Multigate
Maximum drain extension sheet resistance for multi-gate MPU/ASIC (NMOS) (Ω/sq) [G]	670	745	890	960	1060	1150	1250	1340	MPU/ASIC Multigate
Spacer thickness, multi-gate elevated contact [J]	5.5	5	4.4	3.9	3.5	3.1	2.8	2.5	MPU/ASIC Multigate
Thickness of multi-gate elevated junction (nm) [U]	10	9	8	7	6.3	5.6	5.0	4.5	MPU/ASIC Multigate
Maximum silicon consumption for multi-gate mpu/asic (nm) [K]	10	9	8	7	6.3	5.6	5.0	4.5	MPU/ASIC Multigate
Silicide thickness for multi-gate MPU/ASIC (nm) [L]	12	11	10	8	7.6	6.7	6.0	5.4	MPU/ASIC Multigate
Contact silicide sheet R_s for multi-gate MPU/ASIC (Ω/sq) [M]	13.3	14.8	16.7	19	21.1	23.8	26.6	29.6	MPU/ASIC Multigate
Contact maximum resistivity for multi-gate MPU/ASIC (Ω-cm^2) [N]	3.8E-08	3.3E-08	2.8E-08	2.4E-08	2.2E-08	1.9E-08	1.6E-08	1.4E-08	MPU/ASIC Multigate
Physical gate length low operating power (LOP) (nm)	13	11	10	9	8	7	6.3	5.6	LOP
Equivalent physical oxide thickness for FDSOI low operating power T_{ox} (nm) for metal gate [A, A1, A2]	0.8	0.7							LOP FDSOI
Gate dielectric leakage at 100 °C for FDSOI (A/cm^2) LOP [B, B1, B2]	1.9E+02	2.3E+02							LOP FDSOI
Metal gate work function for FDSOI and multi-gate LOP [S]	midgap	midgap	midgap	midgap	midgap	midgap	midgap	midgap	LOP
Equivalent physical oxide thickness for multi-gate low operating power T_{ox} (nm) for metal gate[A, A1, A2]	0.8	0.8	0.7	0.7	0.7	0.7	0.6	0.6	LOP Multigate
Gate dielectric leakage at 100°C for multi-gate (A/cm^2) LOP [B, B1, B2]	1.9E+02	2.3E+02	2.5E+02	2.8E+02	3.1E+02	3.6E+02	4.2E+02	4.2E+02	LOP Multigate
Physical gate length low standby power (LSTP) (nm)	16	14	13	11	10	9	8	7	LSTP
Equivalent physical oxide thickness for FDSOI low standby power T_{ox} (nm) for metal gate [A, A1, A2]	1	0.9	0.8						LSTP FDSOI
Gate dielectric leakage at 100°C for FDSOI (A/cm^2) LSTP [B, B1, B2]	1.9E-01	2.1E-01	2.3E-01						LSTP FDSOI
Metal gate work function for FDSOI and multi-gate LSTP | Ei – ϕ_m | (eV)| NMOS/PMOS [S]	±0.1	±0.1	±0.1	±0.1	±0.1	±0.1	±0.1	±0.1	LSTP
Equivalent physical oxide thickness for multi-gate low standby power T_{ox} (nm) for metal gate [A, A1, A2]	1.1	1.1	1.0	1.0	0.9	0.9	0.8	0.8	LSTP Multigate
Gate dielectric leakage at 100°C for multi-gate (A/cm^2) LSTP [B, B1, B2]	1.9E-01	2.1E-01	2.3E-01	2.7E-01	3.0E-01	3.3E-01	3.8E-01	4.3E-01	LSTP Multigate
Thickness control EOT (% 3σ) [W]	<±4	<±4	<±4	<±4	<±4	<±4	<±4	<±4	MPU/ASIC
Poly-Si or Metal Gate electrode thickness (approximate) (nm) [X]	20	18	16	14	12.6	11.2	10.0	9.0	MPU/ASIC
Gate etch bias (nm) [Y]	7	6	5	5	4.7	3.4	3.0	3.5	MPU/ASIC
L_{gate} 3σ variation (nm) [Z]	1.20	1.08	0.96	0.84	0.76	0.67	0.60	0.54	
Total maximum allowable lithography 3σ (nm) [AA]	1.04	0.94	0.83	0.73	0.65	0.58	0.52	0.47	MPU/ASIC
Total maximum allowable etch 3σ (nm), including photoresist trim and gate etch [AA]	0.60	0.54	0.48	0.42	0.38	0.34	0.30	0.27	MPU/ASIC
Resist trim maximum allowable 3σ (nm) [AB]	0.35	0.31	0.28	0.24	0.22	0.19	0.17	0.16	MPU/ASIC
Gate etch maximum allowable 3σ (nm) [AB]	0.49	0.44	0.39	0.34	0.31	0.27	0.24	0.22	MPU/ASIC
CD bias between dense and isolated lines [AC]	≤15%	≤15%	≤15%	≤15%	≤15%	≤15%	≤15%	≤15%	MPU/ASIC

UMC 1512
UMC v. AICP

Table FEP4b Thermal, Thin Film, Doping and Etching Technology Requirements—Long-term Years

Grey cells indicate the requirements projected only for intermediate, or long-term years. Near-term line items are not included.

Year of Production	2015	2016	2017	2018	2019	2020	2021	2022	Driver
Minimum measurable gate dielectric remaining (post gate etch clean) [AD]	>0	>0	>0	>0	>0	>0	>0	>0	MPU/ASIC
Profile control (side wall angle- degrees) [AE]	90	90	90	90	90	90	90	90	MPU/ASIC
Allowable threshold voltage variation from charge in dielectric (mV) [AF]	8.5	8	8	8	7.5	7.5	7	7	LSTP
Allowable interfacial charge in high-κ gate stack (cm^{-2})[AG]	1.8E+11	1.9E+11	2.2E+11	1.7E+11	1.8E+11	1.8E+11	1.9E+11	1.9E+11	LSTP
Allowable bulk charge in high-κ gate stack (cm^{-3}) [AH]	9.2E+17	1.1E+18	1.3E+18	8.6E+17	1.0E+18	1.0E+18	1.2E+18	1.2E+18	LSTP
Allowable bulk charge in high-κ gate stack (ppm) [AH]	41.7	48.4	61.3	39.2	45.4	45.4	53.6	53.6	LSTP
Allowable critical metal impurity level in high-κ dielectric (ppm) [AI]	4.2	4.8	6.1	3.9	4.5	4.5	5.4	5.4	LSTP

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

Notes for Tables FEP4a and FEP4b:

[A] This number represents the effective thickness of the dielectric alone, at the maximum operating frequency of the technology, without substrate or electrode effects. This parameter is obtained through an electrical measurement of capacitance corrected for substrate (quantum) and electrode (depletion) effects. In contrast to EOT, the electrical equivalent thickness, or capacitance equivalent, thickness (CET), includes a contribution due to gate (Poly-Si) depletion and a contribution due to the quantum effects which cause the centroid of the inversion charge to be beneath the Si surface. A more detailed discussion of the measurement of EOT is on a separate workbook page of the linked file. Values for EOT were derived from the electrical device requirements (CET) as given in the PIDS chapter. MASTAR and other simulations were used to subtract the substrate dark space and gate depletion for the prescribed channel configuration, doping and voltage at each technology generation.

[A1] EOT values are reported for alternate gate electrode options: Poly-Si whose doping at the dielectric interface is $1 \times 10^{20}/cm^3$ (light doping), $1.5 \times 10^{20}/cm^3$ (the nominal case) and $3 \times 10^{20}/cm^3$ (representing aggressive doping) and metal gate. In approximate terms, poly depletion for 1.5E20 doping was about 0.4 nm, and it was about 0.3 nm for 3e20. Thus, increasing poly-Si doping from 1E20 to 3E20 increases the allowable EOT by 0.2 nm. Similarly, metal gates can use EOTs that are about 0.4 nm thicker than 1.5E20-doped poly-Si. Due to numerous practical difficulties at the high-κ/poly-Si interface, it is envisioned that many companies may want to introduce metal gate at the same time, or maybe even before, high-κ dielectrics are introduced.

[A2] The color-coding of each technology generation considers the ability of known dielectrics to meet gate leakage, uniformity, and reliability requirements. For all three applications (HP, LOP, and LSTP), the gate leakage requirements, in this scenario, can no longer be met by optimized oxynitride (which is taken to have a leakage of 1/30 that of SiO_2); hence high-κ dielectric is needed. Based on early announcements and encouraging results with high-κ dielectrics and poly-Si gates, particularly at 0.7 nm EOT and above, (many of which employed a layered SiON-HfSiON system), were colored yellow. All other high-κ dielectrics, i.e., those thinner than 0.7 nm and those requiring metal gates, are colored red because a manufacturing solution to all known problems is not at hand.

[B] The gate leakage, specified at 100°C, is derived from the target values of transistor sub-threshold leakage at room temperature: 200 nA/μm, 5 nA/μm, and 300 pA/μm for HP, LOP and LSPT, respectively. This device leakage is specified in the PIDS chapter section on Logic—High Performance and Low Power Technology Requirements as the off-state leakage (excluding the junction and the gate leakage components) at room temperature. Although short channel effects cause the actual off-state leakage to exceed the target values, especially in the longer term years, the gate leakage specification is derived from the target values which remain constant over all years of the roadmap.

[B1] The areal gate leakage is modeled as the allowable gate leakage divided by the physical gate length. However, it should be noted that the total gate leakage is the sum of three leakage components: 1) leakage between the source and the gate in the gate-source overlap area, 2) leakage between the channel and the gate over the channel region, and 3) leakage between the gate and the drain in the gate-drain overlap area. The magnitude of each of these three components will depend on the gate, source, and drain biasing conditions. The color coding of leakage values is based on UTQUANT simulations of tunneling current from an inversion channel to the gate for the mid-point EOT. (These simulation results are given in a separate worksheet file online.) It should be emphasized that the tunneling current density will generally be much higher between the junction and gate than between an inversion channel and gate. Thus these simulations represent a best case (lowest leakage) condition, where the gate-to-junction overlap area is minimal. When oxide will meet the leakage specification, the value is coded white. Based on present experience, optimized oxynitride dielectrics have about 30 times less leakage current than oxide; values are coded white when optimized oxynitride can meet the leakage specs. Based on more preliminary data, the leakage current of the Hf-family of high-κ dielectrics is about 10^6 lower than that of oxide; accordingly if this high-κ seems likely to meet the leakage requirement, the values are coded yellow; however, since high-κ dielectrics having an EOT < 0.7 nm are coded red (as discussed in note A2), leakage values for such thin dielectrics are also coded red..

[B2] The unmanaged gate leakage power is the total static chip power that would occur if all the devices on a chip had gate leakage equal to the maximum allowable value. Power management will require the extensive use of power reduction techniques, such as power-down or multiple V_t devices to achieve an acceptable static power level.

UMC 1512
UMC v. AICP

[C] The gate electrode work functions come from the PIDS device design. In bulk devices, the electrode work function and the channel doping jointly control device threshold, which is selected to maximize I_{on}, while meeting the I_{off} specification. In addition, the doping affects both short channel effects and channel mobility and, thus, requires an optimization. The PIDS design shows that work functions 0.1 eV below E_c and 0.1 eV above E_v are best for NMOS and PMOS respectively. The requirement stated in the table is for the work function to be within 0.2 eV of the Silicon band edge. Even though there is some leeway in the choice of the gate work functions, the work function itself needs to be controlled to within about 10 mV 3σ, since that it becomes a component of the device threshold voltage tolerance.

[D] The channel doping for bulk CMOS devices comes from the PIDS device design. The doping, along with the gate dielectric thickness and the junction depth control short channel effects and thus must be co-optimized. The reduced short channel effects associated with higher channel doping must also be traded off for reduced channel mobility and increased tunnel leakage. The values presented in the table reflect a representative co-optimization. Channel doping above $5 \times 10^{18}/cm^3$ was colored yellow because of concerns about excessive band-to-band tunneling leakage in junctions.

[E] Bulk/FDSOI/DG – Long channel Electron Mobility Enhancement Factor due to strain, representing the enhancement in peak electron mobility in NMOS devices.

*[F] X_j at Channel (Extension Junction) are given by the PIDS Bulk device designs (with a range of ±25%) in which they are taken as 0.5*Physical Gate Length. Alternative device designs, employing offset spacers and deeper extension junctions which preserve or even extend the effective channel length, are needed to allow deeper extension junctions. Junction depths for NMOS and PMOS are the same.*

[G] The maximum allowable parasitic series resistance for NMOS devices comes from the PIDS device design to meet the performance requirements. The allowable resistance for PMOS is taken to be 2.2 times the NMOS values. The components of this series resistance include: the spreading resistance (affected by the abruptness of the extension junction), the accumulation resistance, the sheet resistance of the drain extension region, and the contact resistance. The maximum drain extension sheet resistance is modeled by averaging the sheet resistances of a Gaussian and a box doping profile where the active surface doping concentration was taken as $2 \times 10^{20}/cm^3$. (See the worksheet labeled DopingModels in the linked file). The drain extension sheet resistance value must be optimized together with the contact resistance and junction lateral abruptness (which effects spreading resistance), in order to meet the overall parasitic resistance requirements. This is a relatively crude model and the resultant sheet resistance values should only be used as a guide.

*[H] The extension junction lateral abruptness requirement, based on Short Channel effect,[8] is consistent with a 3 decade fall off of doping over the lateral extent of the junction, which is taken to be 60% of the vertical junction depth. Since the extension junction depth is taken as $0.5*L_{gate}$, the abruptness = 0.1 * Physical Gate Length (nm). Note discussion of the integration choices in the supplemental material online at http://public.itrs.net.*

*[I] Contact Junction Depth = 1.1*Physical Gate Length (with a range of ±33%) for Bulk devices. Junction depths for NMOS and PMOS are the same.*

[J] Spacer thickness (width) is taken as the same as the Contact Junction Depth, namely $1.1 \times L_{gate}$, for bulk devices.. Validity established using response surface methodology in "Response Surface Based Optimization of 0.1 μm PMOSFETs with Ultra-Thin Oxide Dielectrics".[9] For FDSOI and Multi-gate devices, the spacer width was taken to be half that value, i.e., $0.55 \times L_{gate}$. (See the worksheet labeled Doping Models in the linked file of the electronic version of this chapter, online at http://public.itrs.net).

[K] Silicon consumption is based on half the contact junction depth, for bulk devices. For advanced fully depleted and multi-gate devices, having elevated contacts, the silicide thickness is such that the silicide/silicon interface is coplanar with the channel/gate dielectric interface. The silicon consumption is equal to the added silicon thickness.

[L] Silicide thickness is based on the silicon consumption, which is taken to be 1/2 of the Contact X_j midpoint to avoid consumption-induced increase in contact leakage for bulk devices. Less than half of the junction can be consumed.[10] For fully-depleted and multi-gate devices, having elevated contact structures, the silicide thickness is that thickness yielded by consumption of the contact silicon added above the plane of the gate dielectric/channel interface. For cobalt and titanium di-silicide layers this silicide thickness is nominally equal to the silicon consumed. For nickel mono-silicide the silicide thickness is equal to 2.22/1.84× of the silicon consumed. In the table we have assumed NiSi implementation. Silicide thicknesses less than 15 nm were coded as red, along with Si consumption and silicide sheet resistance. (See the worksheet labeled Doping Models in the linked file of the electronic version of this chapter, online at http://public.itrs.net).

[M] Contact silicide sheet resistance: assumes 16 μΩ-cm silicide resistivity for NiSi.

[N] The Si/Silicide maximum interfacial contact resistivity values were calculated from the PIDS total allowed MOSFET Source/Drain resistance allocated to the extension junction resistance contact resistivity. The extension junction resistance was taken to be the extension junction sheet resistance (given as a line item in the table) times the length of the extension junction, which was taken to be the length of the spacer plus the lateral excursion of the extension junction less the lateral excursion of the contacting junction. The lateral excursion of a junction was assumed to be 0.6 times the junction depth. The calculation further assumes that the transistor contact length is taken to be twice the MPU half pitch, where length is in the direction of current flow. These values should be appropriately modified if different transistor contact lengths are assumed. (See the worksheet on Doping Models in the linked file of the electronic version of the chapter online at http://public.itrs.net). Note that this contact resistivity is the maximum allowable. The contact resistivity was colored red below 5×10^{-8} Ohm-cm^2 and white above 1×10^{-7} Ohm-cm^2. The values of contact resistivity, drain extension sheet resistance, and drain extension lateral abruptness must be co-optimized in order to meet the overall parasitic resistance requirements.

[O] Assumes that the trench depth for bulk is proportional to the contact junction depth plus depletion width into the well. The constant of proportionality was determined by setting the 2003 value equal to 400 nm.

[P] Assumes a minimum trench width equal to the MPU half-pitch.

[Q] Assumes that the trench width is reduced by no more than half of the top dimension.

[R] Assumes a mask thickness equal to half of the MPU half-pitch adds to the trench depth in the substrate

[S] In fully-depleted and multi-gate devices, the gate work function is the prime determinant of device threshold; accordingly values near midgap are more appropriate. The scenario depicted in the table is one which seeks to maintain the same work function over time for a given device type and to minimize the number of different work functions needed for different applications. Dual work function gates are best served with work functions that are ±0.15 eV from midgap for NMOS and PMOS respectively (-/+0.1 eV for LSTP applications). Several applications, including some low cost ones, can be satisfied with a single midgap work function for both NMOS and PMOS. As with gate electrodes for bulk devices, work function control of 10 mV, 3σ is required.

UMC 1512
UMC v. AICP

[T] Si thicknesses for FDSOI and multi-gate devices was based on PIDS device optimization to control short channel effects. Although some company-to-company differences in the final optimized nominal thickness is expected, the tolerance on the final thickness is ±10%. The colorization of the FDSOI thickness is based on thinning the material specified in the Starting Materials tables (in Tables FEP2a and FEP2b), which are controlled to ±5%, to the final thicknesses required by PIDS devices, which require a ±10% tolerance, assuming that the thinning process introduces no additional variation in thickness. Silicon thickness for all multi-gate requirements was colored red, where control of the thickness, sidewall angle, and channel mobility have not been demonstrated.

[U] The thickness of the elevated junctions in FDSOI and in Multi-gate was taken as equal to the Physical Gate Length. In this model, the entire thickness of the elevated junction is consumed to form silicide. By adjusting this thickness tradeoffs can be made between silicide sheet resistance and lateral parasitic junction-to-gate capacitance.

[V] Based on a trench depth equal to the FDSOI thickness

[W] The 4% 3 σ tolerance on EOT comes from Modeling of Manufacturing Sensitivity and of Statistically Based Process Control.[11] Requirements for 0.18 micron NMOS device. Color coding is taken the same as the dielectric—see footnote A2.

[X] Gate thickness is taken as two times the physical gate length. Thicker gates reduce gate series resistance, but at the expense of increased topography and aspect ratio. Thicknesses < 40 nm are colored yellow, and thicknesses <25 nm are colored red.

[Y] Bias is defined as the difference between the printed gate length and the final post-etch gate length.

[Z] The total gate length 3σ variation encompasses all random process variation including point-to-point on a wafer, wafer-to-wafer, and lot-to-lot variations. It excludes systematic variations such as lithography proximity effects, and etch variations such as CD bias between densely spaced and isolated lines. This total variability is taken to be less than or equal to 12% of the final feature size. A conventional MOS structure is the basis for these calculations. MOS transistor structures that vary in any way from the conventional structure (e.g. Vertical MOS transistors) will have different technical challenges and will not fall within these calculations. The data is computed taking into account lithographic errors during resist patterning and combined etch errors due to both resist trim and gate etch.

[AA] The allowable lithography variance σ^2_L is limited to 3/4 of the total variance, σ^2_T of the combined lithography and etch processes. It is further assumed that the lithographic and etch processes are statistically independent and therefore that the total variance is the sum of the etch and lithography variances. This implies among other things that the printed features in the resist have vertical wall profiles and be sufficiently thick to withstand the etch process without loss of dimensional fidelity. Refer to the Etch supplemental file in the electronic version of this chapter online at http://public.itrs.net.

[AB] It is assumed that the resist trim and gate etch processes are statistically independent and therefore that the respective variances, σ^2, of the two processes are additive. 1/3 of the combined trim-etch variance is allocated to the trim process, with the remaining 2/3 allocated to the etch process.

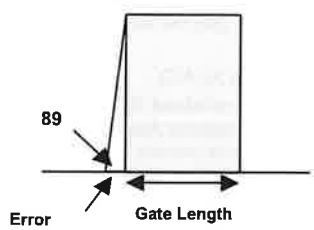
[AC] 15% dense-iso CD budget is a combination of measurements from Etch, Lithography and Metrology.

[AD] It is important that some dielectric remains after the gate etch clean step. Between technology generations the dielectric thickness decreases and there is an onset of using high-κ materials (2008) to replace the gate dielectric. Both advances represent challenges to ensure there is an amount of remaining dielectric and the ability to measure the remaining material.

[AE] Profile can be a major contributor to etch errors (see inset). Accurate measurement of vertical profiles remains difficult. Long term, the effect of edge roughness on device performance needs to be addressed and methodology of the measurement determined.

Gate error produced @ 89 degrees = 3.5 nm

Gate Length:	65nm	53nm	45nm	37nm	32nm	30nm	25nm
% error =	5.4	6.6	7.8	9.4	10.9	11.7	14



[AF] The total allowable threshold voltage variation (ATVV) is defined by PIDS as being 3% of the power supply voltage (V_{dd}). The amount partitioned to charges in the high k dielectric, either at an interface or in the bulk, was taken as 1/3 of the total ATVV. The values given here are for LSTP applications, which impose the most stringent requirements on allowable charge because the EOTs are large.

[AG] Assumes that all of the charge is at the Si-gate dielectric interface, i.e., there is no bulk charge and no charge at an SiO_2/high-κ interface

[AHI] Assumes: i) a single (high-κ) dielectric with uniformly-distributed charge, and ii) a relative dielectric constant of 4 times that of SiO_2. Conversion of the bulk concentrations to units of ppm in the dielectric assume the metal atom density in the high-κ dielectric is the same as that of Si in SiO_2, namely $2.2 \times 10^{22}/cm^3$.

[AI] Assumes that 90% of the charge (and traps) in the high-κ are due to intrinsic bonding defects and that 10% can be due to metallic impurities. The critical metals are expected to be: a) transition metals with low or mid-gap d-states, including Ti, Sc, Nd, V, Ta, Nb, b) transition metals having more d electrons than the high-κ metal, c) Cu, Ag, Ag, and d) radioactive isotopes of high-κ metals.

Intermediate and long-term solutions require the identification of materials with a higher dielectric constant (>10 suggested for intermediate term and >20 for long term) with other electrical characteristics (such as stability and interface-state densities) and reliability approaching that of high quality gate SiO_2. A progression from Hf-based dielectrics to Group III and rare earth (RE) oxides to ternary oxides might be required. A major problem with a material

UMC 1512
UMC v. AICP

other than SiO_2 is the anticipation that a very thin SiO_2 or SiON layer may still be required at the channel interface to preserve interface-state characteristics and channel mobility. This interface layer would increase the equivalent oxide thickness and severely degrade any benefits that accrue from the use of the high-κ dielectric; epitaxial dielectrics may eliminate the interfacial layer, but there is considerable, unresolved concern about high interfacial charge levels and degraded channel mobility in those systems

The presence of an intermediate layer of O-Si-O bonding to bridge between the silicon substrate and high-κ metal ions is expected to limit the scaling of equivalent oxide thickness to nominally 0.4 nm. It is also anticipated that an appropriate material may be required between the high-κ material and the gate electrode to minimize mutual interaction, to inhibit the growth of additional dielectric layers during subsequent processing, and to control/tune the effective gate electrode work function. Improved thickness control and uniformity will also be essential to achieve V_t control for 300 mm and larger wafers. Sensitivity to post-gate, process-induced damage associated with ion implant and plasma etching is expected to increase, especially as it relates to leakage current associated with the gate dielectric perimeter.

Another challenge is the realization of dielectric properties that meet both the gate leakage specification and the reliability requirements. To achieve these needs, the high κ dielectric must have a band gap of 4–5 eV with a barrier height of >1 eV to limit thermionic emission and direct tunneling. In addition, the candidate dielectric material must have negligible trap densities to suppress Frenkle-Poole tunneling and to provide sufficient V_t stability against PBTI and NBTI mechanisms. Finally, the material must have excellent diffusion barrier properties to prevent contamination of the transistor channel by gate electrode material or gate electrode dopant.

The gate electrode also represents a major challenge for future scaling, where work function, resistivity, and compatibility with CMOS technology are key parameters for the new candidate gate electrode materials. Further scaling of polysilicon gates is limited by: a) dopant depletion even when enhanced dopant activation is achieved, b) dopant penetration of the gate dielectric, and c) an apparent incompatibility of polysilicon on high k dielectrics due to poor control of the workfunction in both NMOS and PMOS devices.

Intermediate and longer-term solutions involving metal gates are much more complex and are actively being researched. For one thing, the optimal gate electrode work function differs between different device type and between applications. In bulk NMOS and PMOS devices, band-edge work functions provide the best trade-off between drive current and short-channel-effect control. Yet, fully-depleted SOI and multi-gate devices are better optimized with dual work function gates whose Fermi levels are a couple hundred meV above and below mid gap. Low-cost, low-power applications, may be able to advantageously employ a single (mid-gap) gate work function. Hence, tunable work function systems are of high importance. Lacking a tunable metal gate system, two entirely independent gate stacks (dielectric plus metal) may be needed. Fully- or totally-silicided (FUSI, TOSI) gate electrodes have received attention in an attempt to define the range of work function tuning that is possible in these systems. They appear to be potential solutions for applications requiring workfunctions a few hundred meV around mid gap, but not for applications requiring band-edge workfunctions. Ultimately cladded gate electrodes may be required, where an interface layer is used to achieve the desired gate work function and the second layer is used to prevent oxygen diffusion into or out of the high-κ dielectric, lower the overall gate sheet resistance, and/or provide a more-easily-etched layer. Another very difficult challenge in device scaling is channel mobility enhancement, making mechanical stress a first-order consideration in the choice of front end materials and processes. Potential solutions are complex, in part because electron and hole mobilities are enhanced in different ways by stress, so that NMOS and PMOS devices need to be stressed differently. Conventional processes (trench isolation, gate electrodes, silicides) introduce local stress, which must be accounted for. In addition, global stress can be introduced using alternating layers of Si and SiGe; furthermore, strained Si (or Ge) layers can be used on SOI. Finally stressed layers can be deposited on top of devices or into the substrate (SiGe recessed junctions). Orienting PMOS devices along <100> directions, rather than the traditional <110> direction, can also be employed to enhance hole mobility. The challenge is to integrate multiple sources of local and global stress in such a way that the mobility enhancement from each source is additive, that both NMOS and PMOS devices are enhanced, and that the critical shear stress limit of the substrate is not exceeded (locally).

In order to maintain high device drive currents, technology improvements are required to increase channel mobility of traditional bulk CMOS devices, as well as partially depleted-, and fully depleted SOI devices. The use of strained channel layers, such as strained Si on relaxed Si-Ge for NMOS and strained Si on strained Si-Ge for PMOS will help in achieving this objective but will require considerable process optimization. These enhanced mobility, e.g., strained, channel devices may be needed in conjunction with oxynitride gate dielectrics, before the introduction of high-κ materials. Alternate devices such as non-standard, double gate devices anticipated in the longer-term would also benefit from strained silicon channels.

UMC 1512
UMC v. AICP

The incorporation of mobility-enhanced channels, alternate interfacial layers, high-κ dielectrics, and new gate electrode materials into CMOS configurations pose-significant integration challenges. The limited thermal stability of many of the candidate material systems is incompatible with junction annealing cycles typically used after gate formation. The use of these new materials may require that either junction annealing temperatures be dramatically lowered, or alternate processes be used that reverse the sequence of gate stack and junction formation. Examples of these include the "replacement gate" or gate-last processes. These schemes increase manufacturing complexity, CoO, and may impact device performance and reliability. Consequently strenuous efforts are in place to retain the conventional CMOS process architecture.

Sidewall spacers are currently used to achieve isolation between the gate and source/drain regions, as well as to facilitate the fabrication of self-aligned, source/drain-engineered dopant structures. In addition, offset spacers, formed before implant of the extension junction, may be required to minimize the overlap capacitance and allow slightly deeper junctions. The robustness of the sidewall spacer limits the gate and source/drain contacting structure and processes that can be used to form these contacts. Sidewall spacers have traditionally been formed from deposited oxides, thermal oxidation of polysilicon, deposited nitrides, and various combinations thereof. Traditional sidewall processes will continue to be used at least until the time (2010 estimated) when elevated or raised source/drain structures are required, at which time process compatibility with the side-wall spacer will become critical. Fully-depleted SOI devices will require thin, robust sidewalls having gate dielectric-like reliability and stability. In addition, they must be optimized to minimize parasitic capacitance and series resistance. Below a physical gate length of about 20 nm, even the best state-of the-art thermal oxides are susceptible to defect formation when subjected to selective epitaxial silicon or silicide processes anticipated for elevated contact structures. Nitrides or oxynitrides may offer a better alternative than oxide; however, additional research is needed to find and qualify an acceptable sidewall spacer, compatible with the high-κ gate dielectric.

Thermal and deposited thin films are also very important for filling shallow isolation trenches as well as for pre-metal dielectrics. Trends for decreasing trench width, and higher aspect ratio gaps, suggest that top and bottom corner profile control, and controlled uniform filling of dense/isolated structures, are the key requirements for this application. In the fabrication of shallow trench isolation structures, the top corner of the active region is generally exposed by HF etching of pad and sacrificial oxides prior to the growth or deposition of the gate dielectric. The gate conforms to this corner, forming a region of higher electric field and potentially high defectivity. This region can be thought of as a transistor in parallel with the bulk transistor, with both a lower threshold voltage and saturation current. This leads to a 'hump' in the I_d/V_g characteristics and higher subthreshold leakage. Accordingly, the top corner of the STI trench is rounded, usually by oxidation prior to the deposition of the isolation oxide. Increasing the radius of curvature of this corner increases the V_t of the parasitic transistor and decreases the magnitude of this 'hump'. However, unless new processes are used, device scaling will lead to a decreased radius of curvature.

The magnitude of the parasitic drain current also depends on the degree of recession of the field oxide adjacent to the active edge, since that will in part determine the cross section of the edge transistor. Therefore, as the radius of curvature is scaled down with the isolation width, hopefully so is the recession of the field oxide, resulting in at least partial mitigation of the degradation associated with the decrease of the radius of curvature. The recession of this oxide depends on the "hardness" of the deposited isolation oxide to CMP processing and to HF dipping, as well as to the thickness of the pad and sacrificial oxides, all of which are process design choices that are optimized at each year.

The key thermal/doping integration issues are maintaining shallow junction profiles, junction abruptness, obtaining high dopant activation, ensuring thermal compatibility of materials, and controlling the impact of these issues on device electrical performance. A potential solutions roadmap for Thermal/Thin Films is given in Figure FEP4. The technology changes associated with incorporation of strained substrates, high-κ dielectrics, metal electrodes, strain layers, and non-bulk CMOS are sufficiently major that two years of process qualification and pre-production will likely be needed before they are ready for full production. For example, extraordinary amounts of reliability data will be needed before totally new gate stack materials would be released for sale to customers. This is in contrast to previous, less radical changes, which only required a year for qualification.

UMC 1512
UMC v. AICP

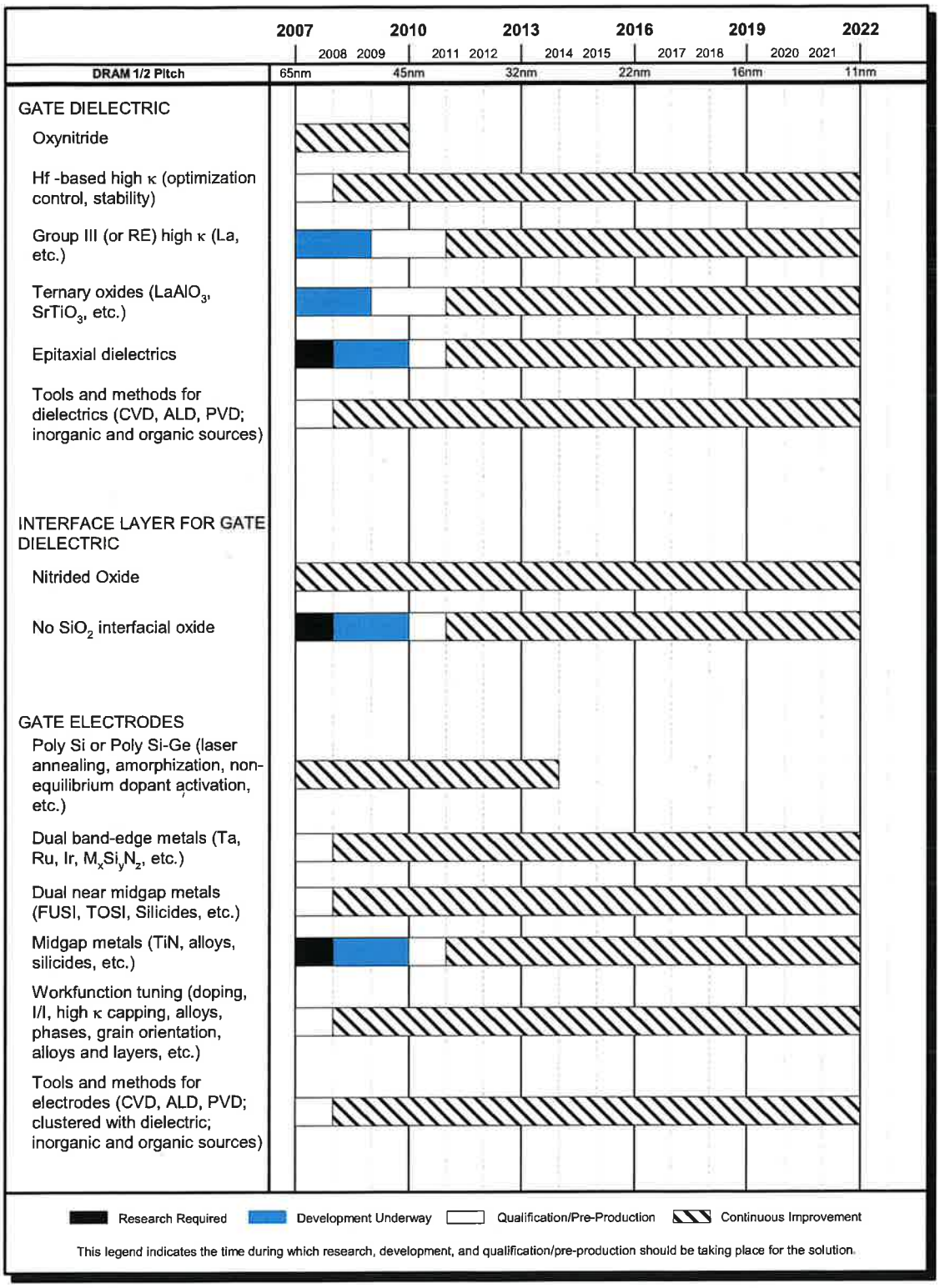


Figure FEP4 Thermal/Thin Films Potential Solutions

UMC 1512
UMC v. AICP

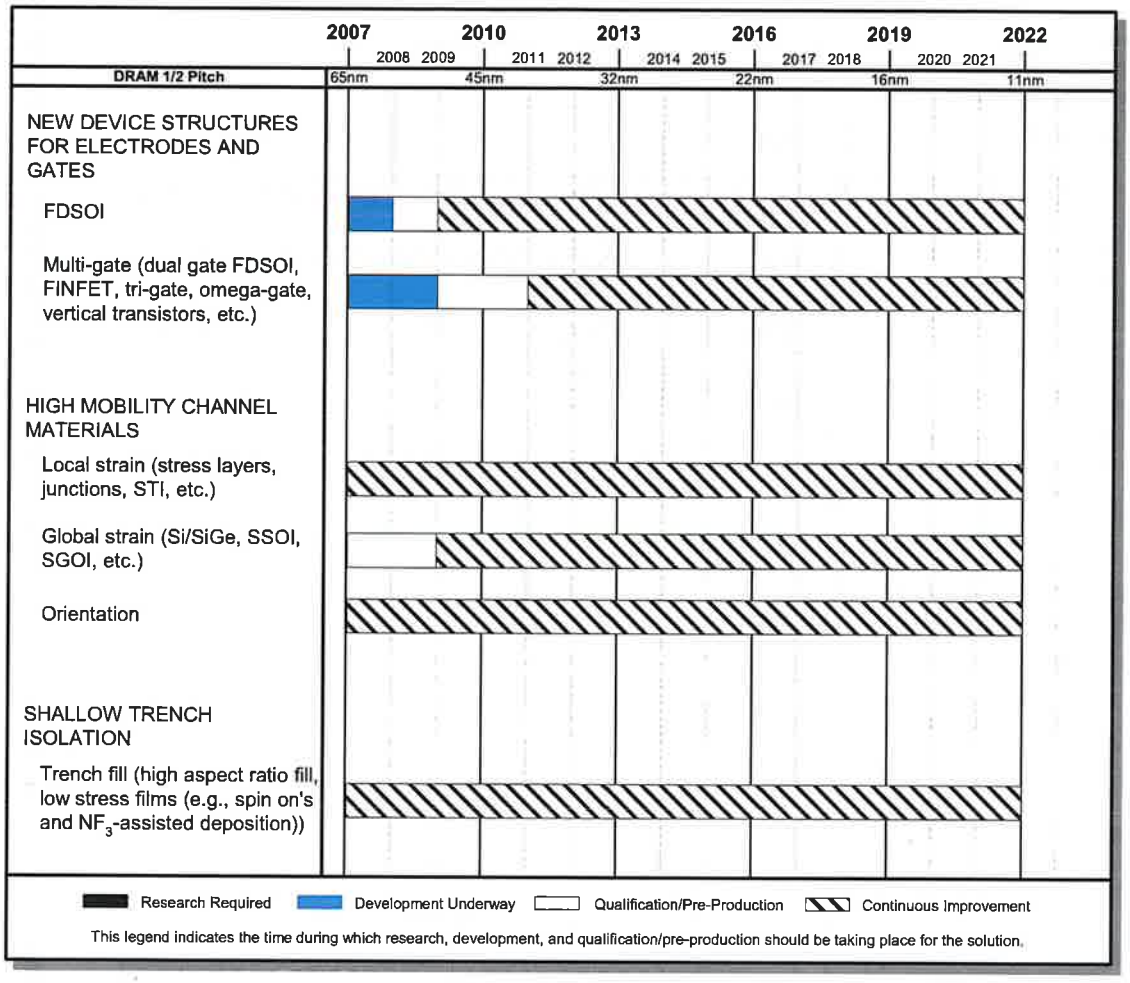


Figure FEP4 Thermal/Thin Films Potential Solutions (continued)

DOPING TECHNOLOGY

The traditional scaling of bulk CMOS devices is becoming increasingly difficult with the consequence that the introduction of numerous new materials and device structures is anticipated within the next few years. The transition to non-classical CMOS devices is expected to be staggered among different companies so that different device architectures may be present at any given time. This is discussed in detail in the PIDS chapter where the following device scenario may be inferred for the high-performance transistor:

Years 2007 through 2012—bulk silicon MOSFETS with the following enhancements:

- Optimized oxynitride gate dielectric
- High-κ gate dielectric and metal gate electrode stacks starting in 2008
- Elevated contacts

Years 2010 through 2015—Fully depleted SOI single gate planar devices with elevated contacts
Years 2011 through 2020—Fully-depleted, dual—or multi-gate devices, e.g. FINFET.

Difficult Challenges—In the very short term, through 2009, the difficult challenges for doping of CMOS transistors are 1) introducing high-κ/metal gate technologies to manufacturing; 2) achieving doping profiles in the source/drain extension regions to attain progressively shallower junction depths needed for control of short-channel effects (~10 nm), while concomitantly optimizing the sheet resistance (~500 Ohms/sq)-junction depth product, doping abruptness at the

UMC 1512
UMC v. AICP

extension-channel junction, and extension-gate overlap; 3) achieving controlled doping profiles in the channel region to set the threshold voltage while concomitantly minimizing short channel effect and maximizing carrier mobility; 4) the formation of, and making low-resistance contact to, shallow highly doped source/drain regions;[1] and 5) the challenges of using msec anneal such as Flash or laser anneal to fabricate high quality interface between high-κ and silicon channel and meet the reliability targets

Also in the near term, but beyond 2009, the grand challenge is more directly stated as "Transistor Structure," where extensions of planar bulk devices will become increasingly difficult to control short channel effects—even with very aggressively scaled junctions and high-κ/metal gate stacks. To alleviate the need for such aggressive scaling, planar, bulk CMOS will likely start to be replaced with non-classical CMOS, i.e., FDSOI and double- and multi-gate devices, which are likely to be implemented on vertical pillars. These non-classical devices present a new set of challenges, including the need for ultra-thin SOI starting material and need for elevated contacting junctions.

Series resistance, particularly of contacts, takes on increased significance since it seriously threatens the further scaling of devices. Since W/L of devices remains relatively constant with channel length scaling, the device resistance remains relatively constant. Yet contact hole sizes scale as the square of the lithographic dimension, causing contact resistances to rapidly increase for smaller feature sizes. Non-equilibrium doping levels at the metal/semiconductor seem to be needed by 2010 when an interfacial contact resistivity of 5×10^{-8} Ω-cm^2 will be needed to meet device performance objectives. Dual metal contacts will be needed by 2013, and more research into suitable materials is urgently needed.

Source and Drain Extensions—For planar bulk CMOS, the management of short channel effects is expected to have a significant impact on processes used for doping drain extensions, channels, halos, and channel edges. Drain extension doping levels are expected to increase, driven by the need to reduce junction depth while concomitantly minimizing parasitic resistance. The implant energy and dose requirements as well as the resulting peak active dopant concentration in the supplemental material are derived from the need to achieve an extension series resistance equal to 15% of the PIDS total series resistance, assuming dopant activation with negligible diffusion (i.e., flash or non-melt laser annealing or solid phase epitaxial regrowth).

In a bulk planar MOSFET the as-implanted (vertical) junction depth with its proportional lateral straggle strongly influences subsequent lateral diffusion and encroachment of the channel. Short channel behavior is therefore strongly linked to the vertical junction depth, and the drain extension resistance is strongly linked to doping concentration and lateral abruptness. The conventional assumption has always been that a more abrupt (box like) lateral junction is better for short channel behavior, essentially since there is less encroachment of the extension doping into the channel, and hence less counter-doping for a more abrupt junction. However, it has recently been shown that due to charge sharing very abrupt junctions also degrade threshold voltage roll-off, and DIBL increases monotonically as junctions become increasingly abrupt (i.e., have steeper doping gradients). Consequently there exists a minimum abruptness of finite value for optimum device performance.

Theoretically an accumulation resistance of the source extension can be defined that strongly depends on lateral abruptness with the smallest accumulation resistance achieved for the most abrupt lateral junction. However, the accumulation resistance cannot be simply viewed as a resistive component in the current path of the device, since any change in its value will change the whole behavior of the device, most notably its short channel behavior. Any change of abruptness has to be followed by a new optimization of the device. Authors who have done so have found no relevant improvement in device performance from abrupt profiles.

The effort to model the requirements for sheet resistance, junction depth, junction abruptness, and series resistance has led to an appreciation of the complexity of the interdependence between these parameters with each other and their combined influence on the overall transistor design. Therefore, the process that collectively optimizes junction depth, doping concentration, and lateral abruptness essentially requires the design of the complete transistor characteristics for each year. This is a task beyond the scope of this roadmap. To that end therefore, all three requirements in the technical requirements tables have been indicated as "guidance" rather than well-defined requirements. In general terms however, for P-channel devices, sensitivity simulations indicate that above a critical abruptness value there is only a marginal reduction in parasitic resistance. Improving the abruptness beyond some critical value therefore gives only minor improvements. On the other hand, for n-channel devices, a more abrupt source extension junction leads to a higher source injection velocity and higher resulting drive current. Therefore, for NMOS devices, higher abruptness values continue to be desirable.

---

[1] *Already selectively-deposited,* in situ-*doped junctions have started to be used to provide uniaxial stress to enhance channel mobility and at the same time to replace ion implantation and thermal annealing. The co-optimization of channel stress, junction doping, and contact materials adds to the challenge.*

UMC 1512
UMC v. AICP

The realization of ultra-shallow source and drain extension junction depths, that are vertically and laterally abrupt, requires not only the development of new and enhanced methods for implanting the doping species, but requires as well the development of thermal activation processes that have an extremely small thermal budget. This is required to truncate the enhanced diffusion that accompanies the activation of the implanted dopant species. The current methods under investigation are identified in the potential solutions Figure FEP5. These methods may introduce significant cost adders to the CMOS process flow. Therefore, one should carefully evaluate the incremental benefits in lateral and vertical abruptness that these processes deliver versus the costs incurred. Sub-nanometer spatially resolved 2D metrology is needed to monitor the location and shape of both vertical and lateral dopant profiles in the extension region.

For non-bulk, fully-depleted ultra-thin-body (FD-UTB) MOSFETs, envisioned in year 2010 and beyond, doping processes will require modification for optimized device drive current and threshold voltage stability. The critical extension junction depth is determined by the thickness of the active silicon layer; thus it becomes somewhat less challenging to make from an implant and anneal perspective. The vertical junction depth in particular looses its meaning since it is now constrained by geometry, the thickness of the Si layer. However, this does not imply that any implant energy is suitable for the extension of an UTB device, since the lateral junction is still linked to the (virtual) vertical one. To derive reasonable values for junction depth, doping concentration and lateral abruptness essentially requires the design of the complete transistor characteristics for each generation—a task beyond the scope of this roadmap. Contacts to these ultra-thin extension junctions becomes much more difficult than in bulk devices, and elevated junctions are required, at least as sacrificial layers for contact silicidation. It remains to be seen how effective such elevated junctions will be in imparting sufficient strain to adequately enhance channel mobility.

FD-UTB devices do not require channel doping to manage the short channel effect, and therefore may be implemented using intrinsic, undoped silicon channels. However, the precise control of doping around the gate edge to optimize gate/drain overlap (or underlap) and the management of parasitic resistance remain important technology challenges.

Vertical channel transistors, such as the FINFET, provide the additional challenge of doping closely spaced arrays of potentially high-aspect-ratio pillars. Such structures seem likely to require isotropic doping processes to form extension junctions.

Contacts and Series Resistance—Scaling of contact area, source/drain junction depth, and contact silicide thickness will lead to increases in parasitic resistance effects unless new materials and processes are developed for producing the self-aligned silicide contact and shunt. The fundamental contact-scaling problem arises from the lateral scaling of the contact area in two dimensions. As a consequence, the contact resistivity associated with the interface between the silicide and the doped contact silicon ultimately becomes the dominant component of the overall source/drain parasitic resistance. The control of this issue requires that: a) dopant concentration at the interface is maximized, b) a lower-barrier-height junction material such as silicon/germanium is used as the contact junction and/or c) low-barrier-height, dual metal (silicides) be used to contact n+ and p+ junctions. An alternative, yet to be practically demonstrated, is to form Schottky barriers that serve as junctions and contacts. The use of selectively deposited silicon-germanium materials in the contact and the controlled profiling of the contact dopant provide potential solutions to these problems. However, the CMOS integration of these processes that mandate different dopant species for the p-channel and n-channel devices makes this a significant challenge. These integration challenges are made more difficult by the fact that the transistor gates are also doped and silicided together with the contact regions.

In bulk devices, several interdependent scaling issues arise that require mutual optimization between contact junction depth, silicide thickness, and silicon/silicide contact resistivity. The contact junction depth, despite the strategic placement of halo implants must still scale with gate length, as shown in Table FEP4. Because of this, progressively less of the contact depth remains available for silicide formation. To avoid high contact resistance and high contact leakage, no more than half the contact depth can be consumed in the formation of the silicide. Therefore with scaled contacts, the silicide must become progressively thinner to accommodate the progressively more shallow contact junction. This practice cannot be continued beyond a certain point because the silicide will tend to become discontinuous and therefore not adequately shunt the contact. Self-aligned NiSi contacts alleviate this problem since they form slightly thicker (more stable) films for a given amount of junction consumption and they are formed at lower temperatures where agglomeration is not as severe compared to the predecessor, $CoSi_2$. Even in bulk devices, ultimately selective deposition of silicon or germanium epitaxial layers in the contact region is required thereby making more silicon available for the silicide formation process. However, as previously discussed, selective epitaxial deposition places increased demands for perfection and robustness on the sidewall spacer.

Another challenge is posed by the introduction of high-κ dielectric/metal gate electrode stacks, also anticipated in the near term. The limited thermal budget of the candidate high-κ materials, will significantly impact the contact formation and shunting strategies.

UMC 1512
UMC v. AICP

Planar single-gate and vertical multi-gate, fully depleted CMOS transistor structures present a new array of challenges for formation of contacts, e.g., to thin, vertical multi-channel arrays linked to heavily-doped contact bus structures. Mastering the intricacies of formation of reliable contacts to these 3D structures will require an additional set of rapid innovations in contact technology. Here, the management of the series resistance of the contact structure remains a major challenge. For the planar single gate devices, the introduction of elevated contacts cannot be avoided without incurring major resistance penalties. Similarly, the research literature contains many references to the strategic use of selective epitaxial shunting of the contact regions of double gate devices in order to realize the required reduction in parasitic resistance. On the other hand, elevated junctions increase the junction-to-diffusion parasitic capacitance, so that both the resistance and the capacitance must be considered. The whole issue of CMOS integration and its associated dual-doping requirements and how doping is accomplished on these epitaxially enhanced contacts remain a major development issue.

Channel—The maintenance of acceptable off-state leakage with continually decreasing channel lengths will require channel-doping levels for planar CMOS transistors to increase in order to control short channel effects for extremely small devices. Increasingly precise control of both vertical and lateral channel doping profiles is required to deal with short channel effects, introducing new challenges for doping tools, process and metrology. The circuit speed advantages of increased drive current for high-performance logic applications has and will continue to drive the use of strained-Si channels materials to increase carrier mobility and to counter the trend towards lower carrier mobility driven by increased channel doping for control of short channel effects. Junction leakage, whether due to band-to-band tunneling, carrier recombination, or contact tunneling and thermionic emission, continues to be an issue, particularly for bulk devices. Part of the leakage concern arises because of direct tunneling as channel levels increase and part because low-thermal budget annealing processes may not remove all crystal damage and dopant diffusion is insufficient to move the junction depletion region beyond regions of un-annealed damage. Leakage is sensitive to junction and channel doping, junction abruptness, and damage removal.

Channel designs for fully depleted CMOS, either in planar SOI or vertical, multi-gate devices, favor the use of intrinsic, undoped silicon. This approach avoids the carrier mobility degradation associated with channel doping but requires that threshold voltage be exclusively controlled by the work function of the gate electrode. These devices usually require dual work function gates that might be achieved by a single metal whose work functions are "tuned" by changes in composition, e.g. through doping.

Multiple issues must be dealt with before the successful introduction of multigate, fully-depleted 3D transistors to CMOS production. These include optimization of the doping transition from highly-doped contact regions to intrinsic channels; reduction of subthreshold leakage at high-field channel edges in multi-gate designs, and producing the solutions to a large number of process integration issues arising from a fully 3D transistor design. These challenges, added to the anticipated shifts to high-κ gate dielectrics and dual-work function, metal-gate materials, constitute a revolutionary change in transistor technology in the coming four to seven years.

UMC 1512
UMC v. AICP

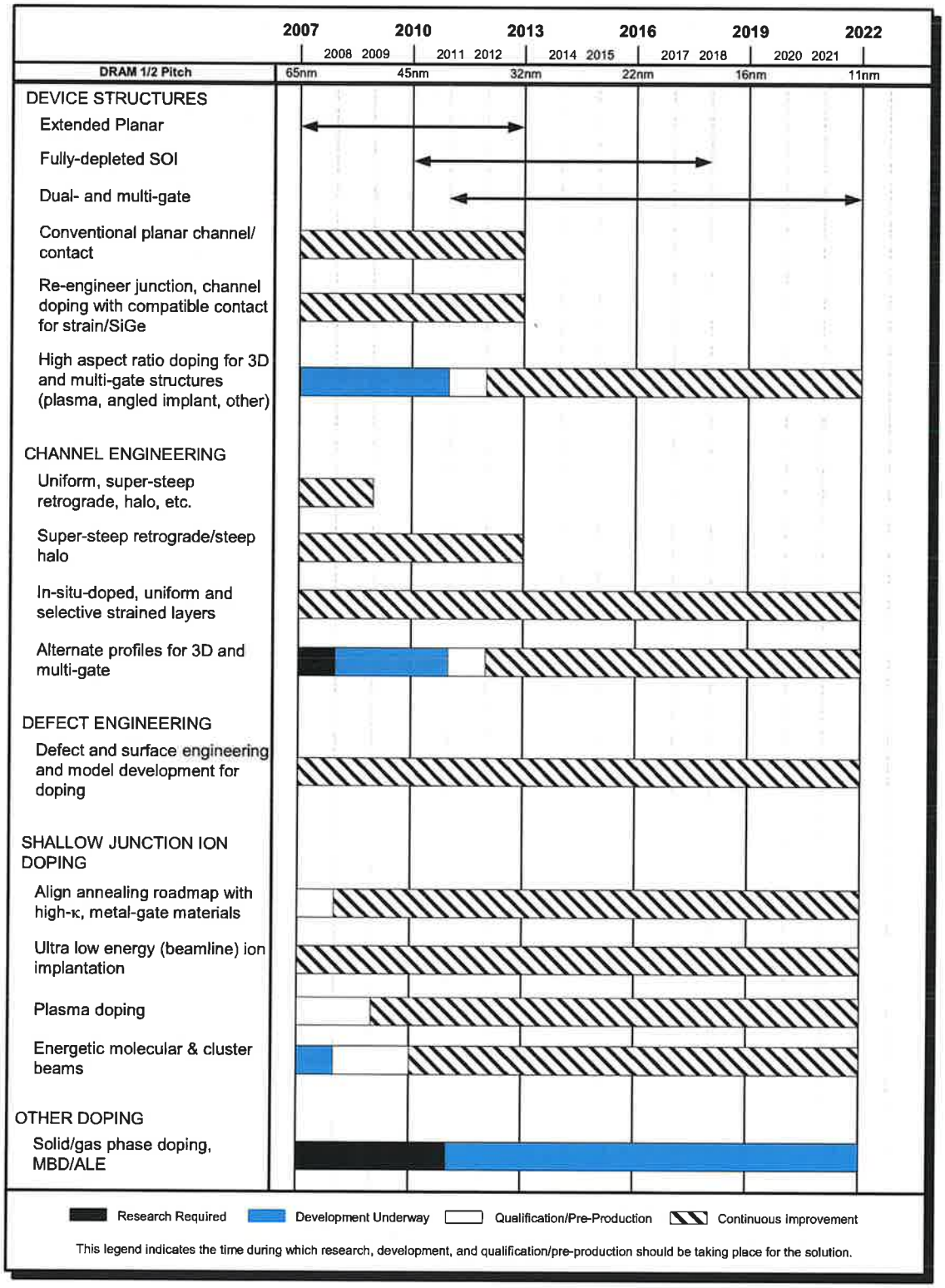


Figure FEP5 Doping Potential Solutions

UMC 1512
UMC v. AICP

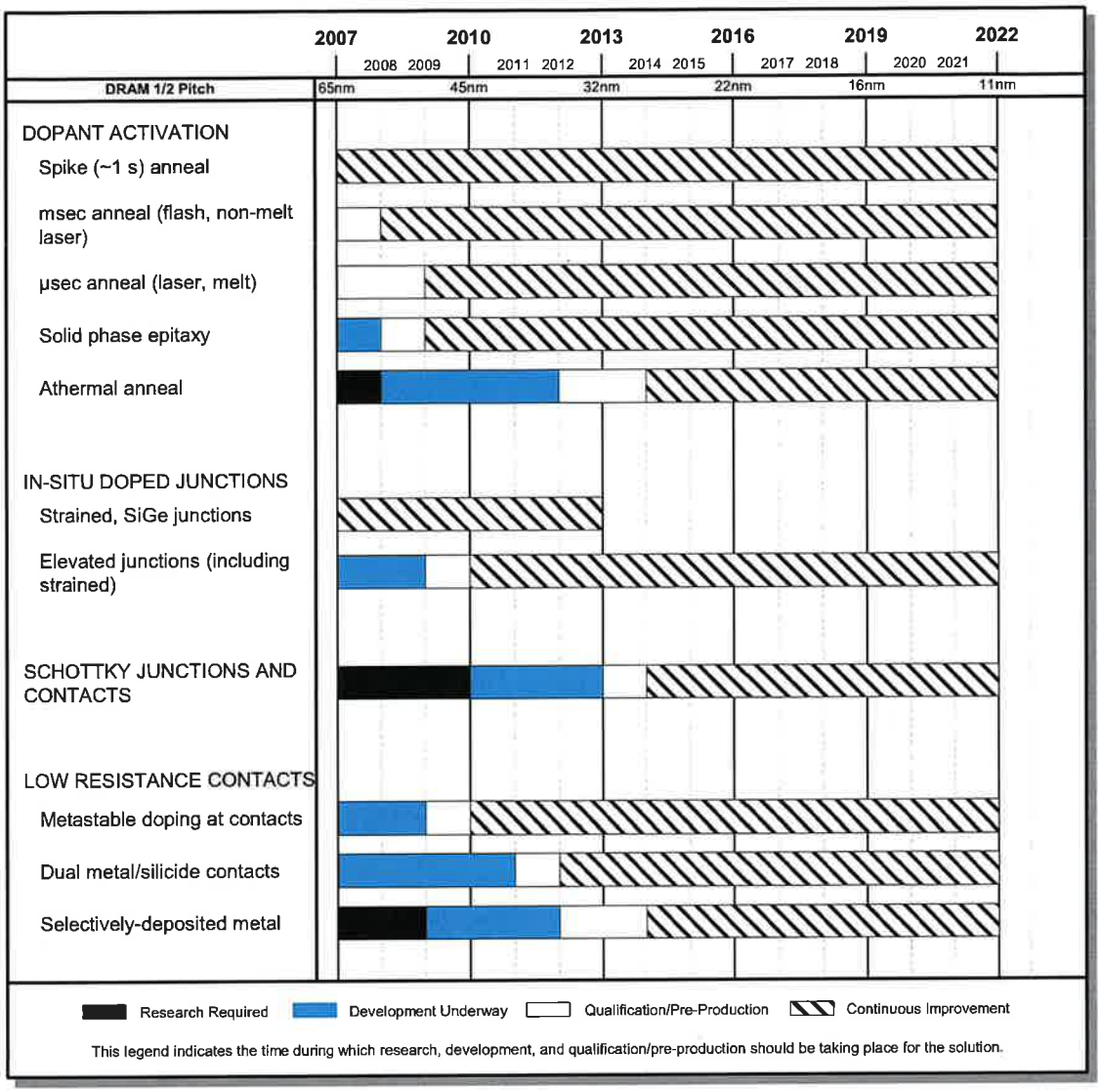


Figure FEP5 Doping Potential Solutions (continued)

Front End Etch Processing

The persistent need in scaling device sizes is to control the CDs, and reduce the median and narrow the distribution, at least in absolute terms. Now that high-κ dielectrics and metal gates are in production, etch processes with sufficient selectivity and damage control for use with these materials have been identified. Scaling requires honing and optimizing these processes to allow for CD reduction. The allowable variation of CD in the ITRS (12% of the physical gate width, 3 sigma) is shared among lithography, resist trimming, and etching. Refer to Table FEP4. The etch portion of the variation includes hardmask etch, doped polysilicon etch, and metal etch, and is assumed to be a global variation including PMOS and NMOS metal gates, and across wafer and wafer-to-wafer variation.

Etch nonuniformity and etch bias determine the variation of the CD attributable to etch. These metrics are strong functions of chamber design, which includes the details of gas distribution, the spatial variation of plasma density and temperature, and the bias voltage. Compensation can be used to improve etch rate uniformity, but its use can result in a narrow process window, and nonuniformity in parameters other than rate, which include etch profiles and electrical damage. Uniformity all the way to the wafer edge is a particularly difficult issue because of the step (or gap) at the edge,

UMC 1512
UMC v. AICP

and its effect on gas flow and distortion of field lines. Edge effects can therefore have an important influence on die yield. CD 3-sigma control at the sub-nm level needs to be achieved with good selectivity and control of damage to the gate dielectric, vertical hard mask and polysilicon profiles, and no undercut or footing of the metal layer used to establish the work function. Integrated metrology designed to feed information forward and back have become standard techniques for achieving sub nm CD 3-sigma control in the trimming process.

Many plasma sources have been developed for improving etch performance. (See Figure FEP6, the Etch Potential Solutions chart.) The required characteristics include controllable and uniform sidewall angle, tightly controlled CDs and etch rate uniformity, and no electrical damage. Historically, high density plasma (HDP), magnetically enhanced reactive ion etching (MERIE), and electron cyclotron resonance (ECR) have been used for these purposes. Pulsed plasma has been added to silicon plasma etching to reduce charge accumulation in the substrate. Gas clustered ion beams have been used for the same purpose, but have not become mainstream. Atomic layer control of etch has been established by cycled polymer deposition and etching steps, modulating the bias power.[12] However this method is inherently slow, and will be relegated only to overetch steps, to achieve the highest selectivity. In this method the surface is passivated by a polymerizing low bias power plasma, and the polymer is removed, along with a "monolayer" of the underlying silicon by a step with higher bias power. Some of these advanced methods may be needed to volatilize products of metals used for metal gate CMOS to suppress chamber contamination, and the frequency of required chamber cleaning, to control mean time between failure (MTBF) and mean time to repair (MTTR).

Trimming of resist has been used for most of this decade to perform the final sizing of the gate. The amount of trim in Table FEP4a and FEP4b is about 40% of the gate length printed in the photoresist. This process is therefore critical to maintaining control of the physical gate length. Trimming also allows compensation for within-wafer and dense/isolated line width variation in the etch step to enable the meeting of overall profile and CD requirements. Scaling of the gate length generally means scaling of the resist thickness, and trimming resist to achieve gate length reduction reduces its height by half the amount trimmed from the length. Sufficient resist must be retained after this operation to allow for the gate hard mask etch. 193 nm photoresists will be used for the foreseeable future due to the adoption of immersion lithography. These resists are prone to line edge roughness (LER) and poor etch resistance.

As linewidth shrinks, the presence of LER is becoming more important to CD control. The LER is at best staying constant as the linewidth shrinks, which makes it a major scaling concern. There is also some evidence that LER contributes to gate leakage. Both lithography and etch can contribute to this. The choice of gate material, photoresist type and etch chemistry contribute to its degree. It is thought that resist trim in an isotropic plasma etch can reduce the amplitude of the roughness.[13] Current methods of quantification need to be standardized to allow the industry to address the problem.

Defect density is a constant concern, especially at the gate level. The multistep gate etch processes required for metal gate structures will generally have to be done in the same chamber with a potential for contamination of the polysilicon etch with metal etch byproducts. Process and equipment development will address chamber cleanliness, and the potential for defects from these etch processes, which may deposit some forms of less volatile metal halides from the metal gate and high-κ layers on the chamber parts. *In situ* interferometry will become more important in allowing preemptive endpoint to terminate the aggressive main etch before clearing each layer, so as to improve selectivity and defectivity.

As non-planar transistors become necessary, etch becomes much more challenging. FinFET configurations bring new constraints to selectivity, anisotropy, and damage control. The fins are actually the smallest features in the whole process, and involve etching using a sidewall defined mask, a feature as small as 0.7 times the gate length.[14] Profile control must be very tight to make very parallel fin surfaces without defects. The gate etch provides many new challenges such as cleaning stringers from the bottom of fins; etching thick potentially planarized poly;, and stopping on very thin oxynitrides or high-κ material, and preserving photoresist. Spacer etch will present unique problems. The spacer must be removed from the face of the fin, which can be many hundreds of angstroms high, without removing it from the polysilicon sidewalls where it defines the very critical length of the extension implant. This may require more selective processes, and improved anisotropy without photoresist present.

Sidewall spacer width is determined by the isotropy of the spacer etch. The silicon loss in the source/drain area is controlled by the selectivity of this etch and the amount of overetch required to clear the spacer film. Control of this important parameter will be improved by integrated metrology.

Shallow trench isolation (STI) also has some challenging integration issues in the 32 nm regime and beyond. Many device manufacturers are using etch processes rather than oxidation to round the top corner of the STI trench to alleviate the transistor double-hump effect. Etching to round the corners also has the advantage of less encroachment into the active area.[15] For this application, integration challenges are top and bottom corner rounding radius control and STI wall slope control. Etching to round the corners also mitigates difficulties in filling high aspect ratio isolation trenches.

UMC 1512
UMC v. AICP

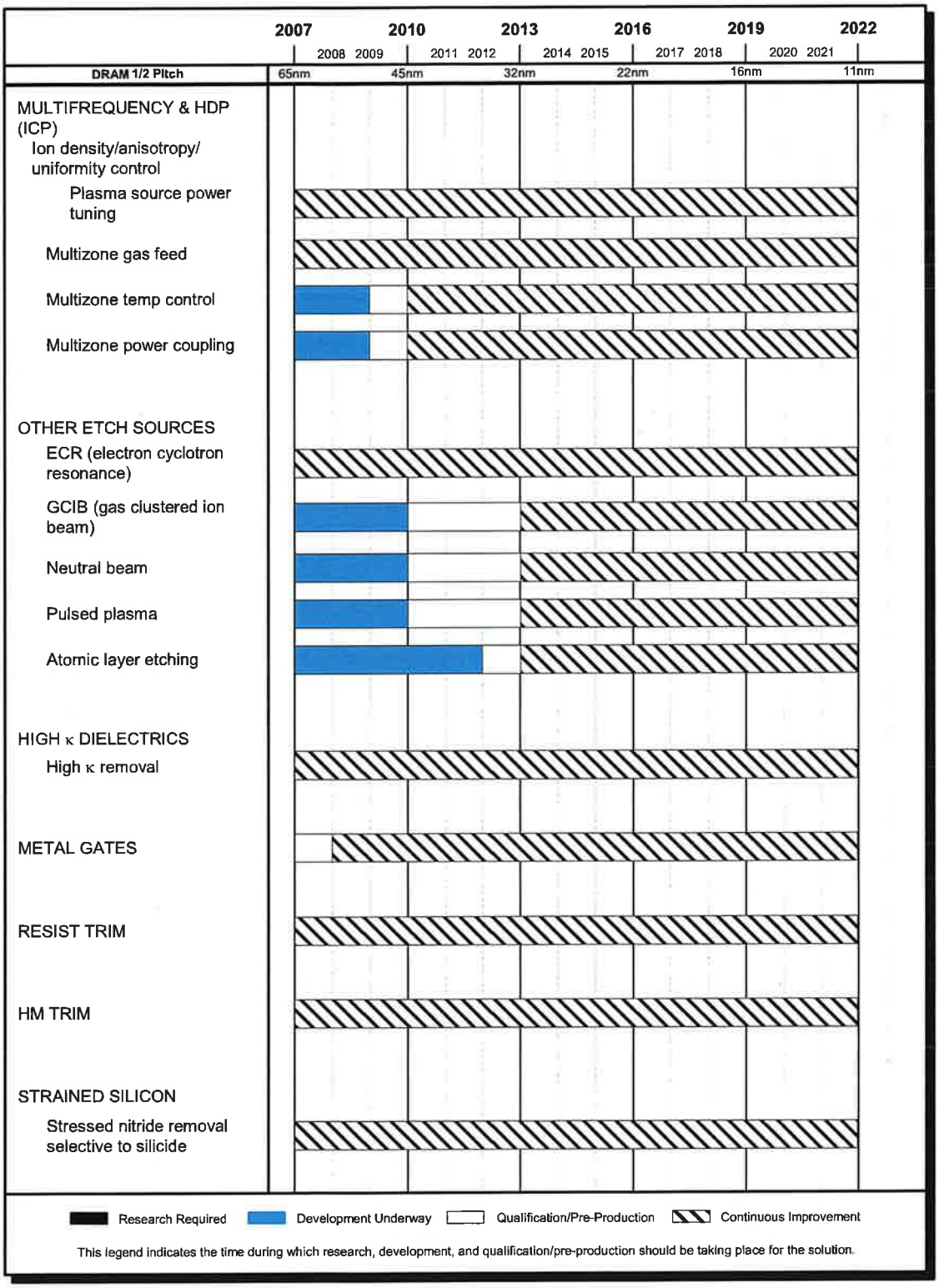


Figure FEP6 Front End Etch Processing Potential Solutions

UMC 1512
UMC v. AICP

DRAM Stacked Capacitor

DRAM capacitor technology faces new challenges of introducing new storage capacitor dielectric and electrode materials. Table FEP5 summarizes the technology requirements for the DRAM stacked capacitor. The DRAM cell size is being scaled down and an area of 6F2 (F is feature size) was realized at the 65 nm technology generation. Each of the target values in Table FEP5 is based on the assumption that a cell capacitance retains at least 25fF/cell to assure stable circuit function and sufficient soft-error immunity.

For the technology generation larger than 90 nm Metal 1 half pitch, nitride/oxide dielectrics films with 3D polysilicon capacitor structures have been used to keep the cell capacitance sufficiently high for sensing and noise immunity. However, it was difficult to keep a sufficiently high cell capacitance value using these materials and structure at the 80 nm technology generation and beyond. Thus, metal electrode and high-κ dielectrics (MIM) have been introduced.

In the metal electrode, besides common requirements of capacitor electrode such as good step coverage, good adhesion, and high work function, low deposition temperature for minimizing thermal degradation of capacitor characteristics and high film density for mechanical hardness in cylindrical structure are also needed. For those reasons, atomic layer deposition (ALD) methods or ALD-like CVD methods for electrode deposition were introduced. TiN which was adapted as metal electrode at 80 nm will be used through the 45 nm technology generation for the MIM capacitor. However, beyond the 45 nm generation, where equivalent oxide thickness (T_{eq}) of below 0.5 nm is needed, new metal electrode materials such as Ru, RuO_2, Pt, IrO_2, $SrRuO_3$ will be needed, for their high work function and to provide a template for preferred microstructure.[16]

The major obstacle in scaling of the DRAM stack capacitor is scaling of t_{eq} for capacitor dielectrics. HfO_2, ZrO_2, and Ta_2O_5 high-κ dielectrics have been already implemented in mass production in 80 nm DRAM devices. However, these dielectrics have a dielectric constant below 50. Development of new dielectric materials such as TiO_2, STO, and BST, which have a dielectric constant higher than 80, will be required to satisfy the t_{eq} requirement beyond the 45 nm technology generation. One of the difficult challenges is scaling of the physical dielectric thickness, T_{phy} while maintaining dielectric constant and leakage current of dielectrics. In general, as the physical thickness of high-κ materials such as BST decreases, the dielectric constant decreases and the leakage current increases. This means that a limit on the reduction of T_{phy} exists and materials with higher dielectric constant should be used to maintain higher T_{phy} while decreasing T_{eq}. Future high-κ dielectrics require not only development of process equipment with high throughput but also development of precursors appropriate for ALD or CVD.

In order to relax the burden of severe T_{eq} scaling of dielectrics, continued process development to increase the capacitor area is also required. Improvement of oxide etching capability for forming the high aspect ratio storage node and the process scheme for obtaining mechanical stability without bridging between storage nodes in the 3D capacitor are major challenges along with T_{eq} scaling of capacitor dielectrics.

The process technology requirements for system-on-a-chip (SOC) with embedded DRAM exhibit many variations depending on the ratio of logic area and memory area. Cell capacitance requirements for embedded DRAM may be smaller than those for stand-alone DRAM. One of the serious problems for SOC is contact via formation. In general, the stacked capacitor DRAM process requires relatively deep contact vias as compared with those in the logic process. Therefore, the contact via diameter for DRAM has to be enlarged to minimize aspect ratio. For this reason it will be difficult to achieve the same metal line pitch for the logic section of SOC using standalone DRAM design rules. In the logic-based SOC, cell size expansion is needed to reduce the capacitor height and to decrease the contact via aspect ratio. On the other hand, in the memory-based SOC, the metal line pitch has to be adjusted so that the DRAM contact via size may be kept large enough. Therefore, some additional break-through in SOC is required to solve this contact via density issue.

UMC 1512
UMC v. AICP

Table FEP5a DRAM Stacked Capacitor Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ Pitch (nm) [A]	*65*	*57*	*50*	*45*	*40*	*36*	*32*	*28*	*25*
Cell size factor a [B]	6	6	6	6	6	6	6	6	6
Cell size (μm^2) [C]	0.025 =0.13x0.195	0.019 =0.11x0.17	0.015 =0.10x0.15	0.012 =0.090x0.14	0.0096 =0.080x0.12	0.0077 =0.071x0.11	0.0061 =0.064x0.96	0.0048 =0.057x0.085	0.0048 =0.051x0.076
Storage node size (μm^2) [D]	0.00845 =0.065x0.13	0.0064 =0.057x0.11	0.0051 =0.051x0.10	0.0041 =0.045x0.090	0.0032 =0.040x0.080	0.0026 =0.036x0.071	0.002 =0.032x0.064	0.0016 =0.032x0.064	0.0013 =0.025x0.051
Capacitor structure	Cylinder /Pedestal MIM	Cylinder /Pedestal MIM	Cylinder /Pedestal MIM	Cylinder /Pedestal MIM	Cylinder /Pedestal MIM	Cylinder /Pedestal MIM	Pedestal MIM	Pedestal MIM	Pedestal MIM
t_{eq} at 25fF (nm) [G]	1.15	0.9	0.8	0.6	0.5	0.4	0.3	0.3	0.3
Dielectric constant	40	43	49	65	78	98	130	130	98
SN height (μm)	1.4	1.3	1.9	1.6	1.5	1.3	1.1	1.3	1.4
Cylinder factor [E]	1.5	1.5	1	1	1	1	1	1	1
Roughness factor	1	1	1	1	1	1	1	1	1
Total capacitor area (μm^2)	0.83	0.65	0.58	0.43	0.36	0.29	0.22	0.22	0.22
Structural coefficient [F]	33.3	34.3	38.6	36.2	37.7	37.6	35.6	45.2	45.2
t_{phy} at 25fF (nm) [H]	11.8	10.0	10.0	10.0	10.0	10.0	10.0	10.0	7.5
A/R of SN [I]	21.6	22.0	38.3	35.4	37.4	36.9	35.0	45.8	57.6
A/R of SN (OUT) for cell plate deposition [I]	33.8	33.8	63.8	63.8	74.7	83.0	93.4	160.4	144.0
HAC diameter (μm) [J]	0.08	0.07	0.06	0.05	0.05	0.04	0.04	0.03	0.03
Total interlevel insulator and metal thickness except SN (μm) [K]	0.78	0.75	0.73	0.7	0.68	0.66	0.63	0.61	0.59
HAC depth (μm) [L]	2.2	2.0	2.6	2.3	2.2	2.0	1.8	1.9	2.0
HAC A/R	28.0	29.3	44.1	42.5	43.5	49.7	43.8	63.1	67.7
$V_{capacitor}$ (Volts)	1.3	1.2	1.1	1.1	1.1	1.1	1.1	1	0.9
Retention time (ms) [M]	64	64	64	64	64	64	64	64	64
Leak current (fA/cell) [N]	0.76	0.70	0.64	0.64	0.64	0.64	0.64	0.59	0.53
Leak current density (nA/cm^2)	91.5	107.9	111.3	148.4	178.0	222.6	296.7	269.8	242.8
Deposition temperature (degree C)	~500	~500	~500	~500	~500	~500	~500	~500	~500
Film anneal temperature (degree C)	~750	~750	<750	<750	~650	~650	~650	<650	<650
Word line R_s (Ohm/sq.)	2	2	2	2	2	2	2	2	2

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

UMC 1512
UMC v. AICP

Table FEP5b DRAM Stacked Capacitor Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
DRAM ½ Pitch (nm) [A]	*22*	*20*	*18*	*16*	*14*	*13*	*11*
Cell size factor a [B]	**6**	**6**	**6**	**6**	**6**	**6**	**6**
Cell size (μm^2) [C]	0.003 =0.045x0.068	0.0024 =0.040x0.060	0.0019 =0.036x0.054	0.0015 =0.032x0.048	0.012 =0.028x0.043	0.010 =0.026x0.039	0.007 =0.022x0.033
Storage node size (μm^2) [D]	0.001 =0.023x0.045	0.0008 =0.020x0.040	0.00064 =0.018x0.036	0.00051 =0.016x0.032	0.0004 =0.014x0.028	0.0003 =0.013x0.026	0.0002 =0.011x0.022
Capacitor structure	Pedestal MIM	Pedestal MIM	Pedestal MIM	Pedestal MIM	Pedestal MIM	Pedestal MIM	Pedestal MIM
t_{eq} at 25fF (nm) [G]	0.3	0.3	0.3	0.25	0.2	0.15	0.1
Dielectric constant	91	78	78	70	80	91	98
SN height (µm)	**1.6**	**1.8**	**2.0**	**1.9**	**1.7**	**1.4**	**1.1**
Cylinder factor [E]	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Roughness factor	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Total capacitor area (μm^2)	**0.22**	**0.22**	**0.22**	**0.18**	**0.14**	**0.11**	**0.07**
Structural coefficient [F]	**72.4**	**90.5**	**114.3**	**120.7**	**120.7**	**120.7**	**120.7**
t_{phy} at 25fF (nm) [H]	7.0	6.0	6.0	4.5	4.1	3.5	2.5
A/R of SN [I]	74.5	90.2	111.4	117.5	122.8	106.8	99.4
A/R of SN (OUT) for cell plate deposition [I]	204.8	225.4	334.2	267.6	296.7	231.3	182.2
HAC diameter (µm) [J]	**0.03**	**0.02**	**0.02**	**0.02**	**0.02**	**0.02**	**0.01**
Total interlevel insulator and metal thickness except SN (µm) [K]	**0.57**	**0.55**	**0.53**	**0.51**	**0.49**	**0.47**	**0.45**
HAC depth (µm) [L]	**2.2**	**2.4**	**2.5**	**2.4**	**2.2**	**1.9**	**1.5**
HAC A/R	**73.6**	**117.7**	**126.8**	**119.5**	**110.5**	**92.9**	**154.3**
$V_{capacitor}$ (Volts)	**0.7**	**0.6**	**0.6**	**0.6**	**0.6**	**0.5**	**0.5**
Retention time (ms) [M]	**64**	**64**	**64**	**64**	**64**	**64**	**64**
Leak current (fA/cell) [N]	**0.41**	**0.35**	**0.35**	**0.35**	**0.35**	**0.29**	**0.29**
Leak current density (nA/cm^2)	**188.8**	**161.9**	**161.9**	**194.2**	**242.8**	**269.8**	**404.7**
Deposition temperature (degree C)	**~500**	**~500**	**~500**	**~500**	**~500**	**~500**	**~500**
Film anneal temperature (degree C)	**<650**	**<650**	**<650**	**<650**	**<650**	**<650**	**<650**
Word line R_s (Ohm/sq.)	**2**	**2**	**2**	**2**	**2**	**2**	**2**

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

Notes for Tables FEP5a and FEP5b:

[A] 2005 Overall Roadmap Technology Characteristics, Table 1a and b

[B] a = (cell size)/F^2 (F : minimum feature size)

[C] Cell size = aF^2 (cell shorter side = 2F)*

[D] SN size = (a/2 -1)F^2 (SN shorter side = F)*

[E] Cylinder structure increase the capacitor area by a factor of 1.5

[F] SC = (total capacitor area) / (cell size)

*[G] teq = 3.9*E0*(total capacitor area)/25fF*

[H] $t_{phy} = t_{eq}*E_r/3.9$ If polysilicon is used as a bottom electrode. $t_{phy} = (t_{eq}-1)*E_r/3.9$

[I] A/R of SN = (SN height)/F

[J] A/R of SN (OUT) = (SN height) / (F - 2 t_{phy})*

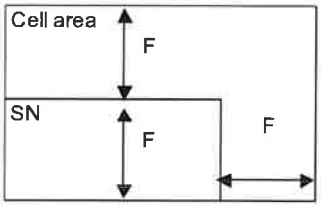
*[K] HAC diameter = 1.2*F*

[L] The thickness is assumed to be 1.05 µm at 180 nm. (10% reduction by each technology generation)

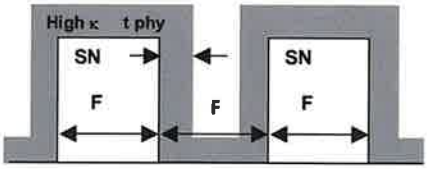
[M] HAC depth = SN height + total interlevel insulator and metal thickness

[N] DRAM retention time (PIDS)

*[O] (Sense Limit*C*V_{dd}/2)/(Retention Time * MARGIN) (Sense limit=30% leak, MARGIN=100)*



Notes[C] & [D] Cell area and Projected SN area



Note [I] A/R of SN (OUT)

UMC 1512
UMC v. AICP

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
DRAM M1 ½ pitch (nm)	65			45			32			22			16			11
Top Electrode	TiN				Ru, RuO_2, Pt, IrO_2, SrRuO...											
Capacitor Dielectric Material	HfO_2, ZrO_2, Ta_2O_5				TiO_2, STO, BST...											
Bottom Electrode	TiN				Ru, RuO_2, Pt, IrO_2, SrRuO...											

Figure FEP7 DRAM Stacked Capacitor Potential Solutions

DRAM Trench Capacitor

Tables FEP6a and FEP6b summarize the technological requirements for trench DRAM capacitors. The target values for cell capacitance that are needed to ensure sufficient signal and retention margin are indicated in the table. Due to the decrease in bit-line capacitance by process measures and cell structure, the cell capacitance can be reduced compared previous ITRS roadmaps. Starting in 2009 the cell capacitance is set at 25fF, which is consistent with the stacked capacitor roadmap. The cell size is assumed to stay at $8F^2$.

For technology generations down to and including the 75 nm Metal 1 (bitline) half-pitch generation, a conventional nitride/oxide dielectric has been used as capacitor dielectric in trench cells. Additional surface enhancement techniques have been implemented starting with the 90 nm generation. The trench profile is widened below a certain depth (bottle-shaped trenches) and the trench side walls are roughened by hemispherical grain deposition (HSG) to increase the capacitor surface area.

High-κ materials such as HfSiO will be introduced at the 48–58 nm generation. ALD will be utilized to deposit new materials into high aspect ratio trenches. Metal top electrodes are also anticipated at the 58 nm generation and will allow the progression from silicon-insulator-silicon (SIS) to metal-insulator-silicon (MIS) capacitors. Conductive metal nitrides and carbon are the most attractive material candidates for the top electrode. Ultimately, metal-insulator-metal (MIM) capacitors are required.

Trench technology allows the fabrication of rigid and mechanically extremely stable high aspect ratio capacitor structures. As a result of ground-rule shrinking, the trench aspect ratio (trench depth divided by trench top width after etch) will increase up to values of ~ 80:1 for the 48 nm generation. Even higher aspect ratios are anticipated at smaller ground rules. The introduction of a 3D cell transistor is expected for the 58 nm generation to resolve the cell transistor scaling issues.

New integration schemes to be introduced beyond the 40nm generation will reduce the temperature budget for the cell capacitor. Thus a more aggressive shrink of the equivalent oxide thickness (T_{eq}) than shown in the 2005 roadmap will be enabled. As a consequence the trench aspect ratio can be kept at less than 90 nm down to 32 nm.

Beyond 25 nm, t_{eq} will need to be reduced to less than 0.7 nm. At this time, it is not clear yet which capacitor high-κ materials will satisfy the leakage current and reliability conditions. Therefore in addition to material and process development, new memory cell concepts will be required at 25 nm and beyond.

For embedded applications, the trench technology with its capacitor buried in the substrate enables a planar transition between the DRAM cell array and the logic circuit. The trench DRAM concept also avoids deep, high aspect ratio contact holes.

UMC 1512
UMC v. AICP

Table FEP6a DRAM Trench Capacitor Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ pitch "F" (nm)	65	57	50	45	40	36	32	28	25
Cell size factor "a" [A]	8	8	8	8	8	8	8	8	8
Cell size (μm^2) [B]	0.045	0.028	0.018	0.016	0.0128	0.0104	0.0082	0.0063	0.0050
Trench structure	bottled	bottled	bottled	bottled	bottled	bottled	bottled	bottled	bottled
Trench bottle circumference (nm) [C]	549	483	399	374	333	300	266	233	208
Trench etch depth (µm) [D]	6.8	6.0	5.8	5.6	5.0	4.5	4.0	3.7	3.4
Bottled trench depth (µm) [E]	6.0	5.3	5.1	4.9	4.3	3.8	3.3	3.1	2.8
Storage node size (μm^2) [F]	3.3	2.6	2.0	1.8	1.4	1.1	0.9	0.7	0.6
Trench surface area enhancement factor (HSG) [G]	1.2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Cell capacitance (fF) [H]	35.0	30.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
teq at Cs (nm) [I]	3.9	3.5	2.8	2.5	2.0	1.6	1.2	1.0	0.8
Trench top opening (nm) [J]	98	81	70	63	56	50	45	39	35
Trench etch aspect ratio [K]	70	74	83	89	89	89	89	94	97
Capacitor structure	Cup SIS	Cup MIS	Cup MIS	Cup MIM	Cup MIM	Cup MIM	Cup MIM	Cup MIM	Cup MIM

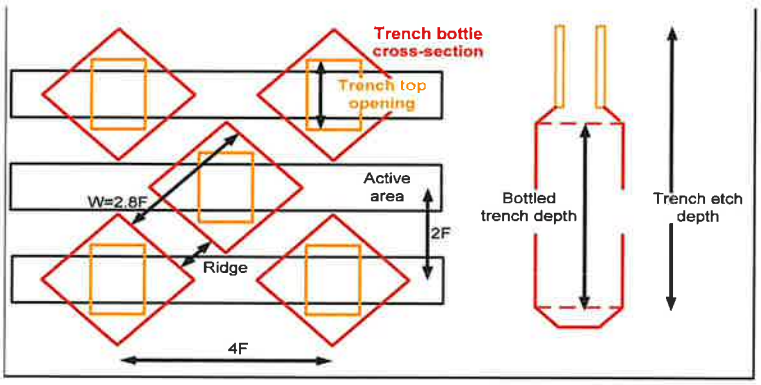
Table FEP6b DRAM Trench Capacitor Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
DRAM ½ pitch "F" (nm)	22	20	18	16	14	13	11
Cell size factor "a" [A]	8	8	8	8	8	8	8
Cell size (μm^2) [B]	0.0039	0.0032	0.0026	0.0020	0.0016	0.0014	0.0010
Trench structure	bottled	bottled	bottled	bottled	bottled	bottled	bottled
Trench bottle circumference (nm) [C]	183	166	150	133	116	108	92
Trench etch depth (µm) [D]	3.0	2.8	2.6	2.4	2.3	2.2	2.1
Bottled trench depth (µm) [E]	2.5	2.3	2.1	1.9	1.8	1.7	1.6
Storage node size (μm^2) [F]	0.5	0.4	0.3	0.3	0.2	0.2	0.1
Trench surface area enhancement factor (HSG) [G]	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Cell capacitance (fF) [H]	25.0	25.0	25.0	25.0	25.0	25.0	25.0
teq at Cs (nm) [I]	0.6	0.5	0.4	0.3	0.3	0.3	0.2
Trench top opening (nm) [J]	31	28	25	22	20	18	15
Trench etch aspect ratio [K]	97	100	103	107	117	121	136
Capacitor structure	Cup MIM	Cup MIM	Cup MIM	Cup MIM	Cup MIM	Cup MIM	Cup MIM

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

Notes for Tables FEP6a and FEP6b:

[A] a=(cell size)/F2 (F equals DRAM half pitch)
*[B] Cell size = a*F2*
[C] Trench bottle circumference = 4(2.83 - ridge)*F*
[D] Trench etch depth (um)
[E] Depth of the bottled part of the trench capacitor
[F] Storage node size = (trench bottle circumference](bottled trench depth)*
[G] Storage node surface enhancement by hemispherical Si grains (HSG)
[H] Cell capacitance = Cs
*[I] teq = 3.9*E0*(storage node surface area)/Cs*



UMC 1512
UMC v. AICP

[J] Trench top opening after etch
[K] Trench etch aspect ratio = (trench etch depth)/(trench top opening)

NON-VOLATILE MEMORY (FLASH)

Tables FEP7a and FEP7b summarize the main technology requirements for NOR and NAND flash memories. The most important issues are related to the cell area reduction (see the non-volatile memory technology requirements, in the *PIDS chapter*) and to the consequent scaling down of the thickness of the two key active dielectrics of the memory cell, namely the tunnel oxide and the interpoly dielectric, in a way that guarantees the charge retention and endurance requirements for the memory cell. For NAND Flash the best definition of the minimum feature size is the half-pitch of the memory cell when viewing a cross section parallel to the bit line, that is also the half pitch of the poly 2-word line. Refer to Figure FEP8.

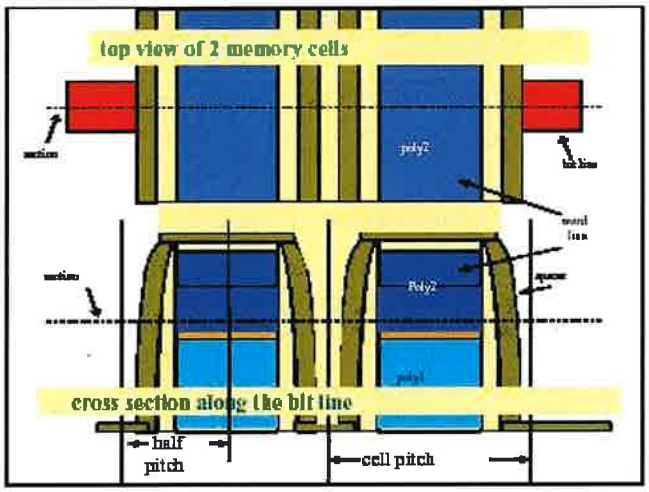


Figure FEP8 Minimum Feature Size of NAND Flash Memory

For NOR Flash memories the definition of the minimum feature size is difficult and can vary among the different Flash manufacturers. Referring to Figure FEP9, the following are definitions of the minimum feature size specific for NOR Flash memories, as follows:

- The half pitch when viewing a cross section parallel to the poly 2 word line
- The poly 1 to poly 1 distance along the word line
- The minimum contact size

UMC 1512
UMC v. AICP

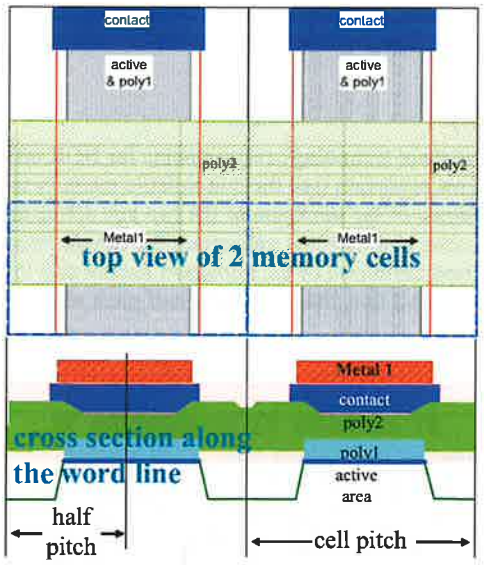


Figure FEP9 Minimum Feature Size of NOR Flash Memory

The tunnel oxide thickness must be reduced for the programming/erasing performances while scaling the interpoly dielectric thickness reduction is necessary to keep the capacitance coupling ratio, α_g, at an almost constant value in order to achieve acceptable ratios between the control and floating gate voltages. The coupling ratio is typically improved by reducing the interpoly dielectric thickness and increasing the tunnel oxide thickness and the floating/control gate coupling area. Scaling the tunnel oxide thickness is one of the key challenges for Flash memories, since this dielectric must simultaneously guarantee good charge retention properties that are better with a higher thickness, and high write/erasing performances that are better with a lower thickness.

The impact of the floating/control gate coupling area on the α_g factor becomes a critical issue starting from 45-40 nm technology for both NOR and NAND flash devices, when the spacing between two adjacent floating gates (poly 1) becomes too small to allow the control gate (poly 2) to overlap the vertical poly 1 sidewalls, as is done in the present architecture. The lack of electrical coupling between poly 1 and poly 2 along the vertical sidewalls of the poly 1 results in a strong degradation of the α_g value and could require a strong reduction of the interpoly dielectric thickness as a compensation. This situation is illustrated in Figure FEP10.

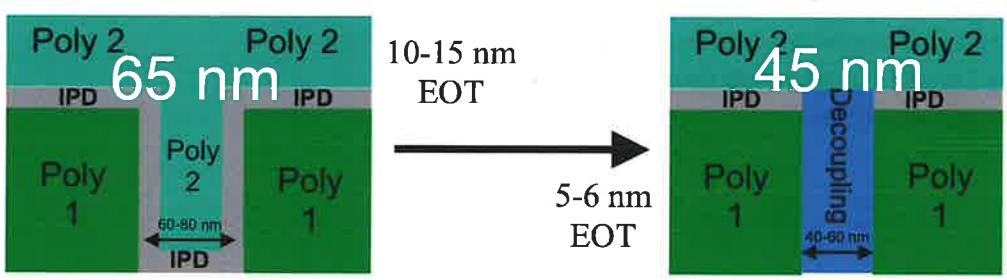


Figure FEP10 Flash Memory Interpoly Dielectric Thickness Scaling at 45 nm

The present interpoly dielectric technology is based on oxy-nitride stacked layers and will probably not be feasible for aggressive reduction of equivalent oxide thickness (EOT), due to unacceptable charge retention properties. Thus, the introduction of high-κ materials at this step will be necessary. Alternatively, new floating gate designs to maintain a high coupling area with the control gate or storage materials different from poly-silicon are potential solutions. From this point of view, the 45–40 nm technology generation will be a transition one with both classical and new solutions depending on the architecture schemes chosen for the memory cells.

UMC 1512
UMC v. AICP

Another challenge for the Flash memory scaling is the formation of shallow trench isolation (STI). The continuous scaling of the dimensions along the X axis (along the word line in Figure FEP9) and the necessity to maintain the depth of the STI trench cause an increase of the trench aspect ratio that needs to be filled with the STI oxide. (Figures FEP11 and FEP12). In response to these challenges an overview of the Flash memory Potential Solutions is shown in Figure FEP13.

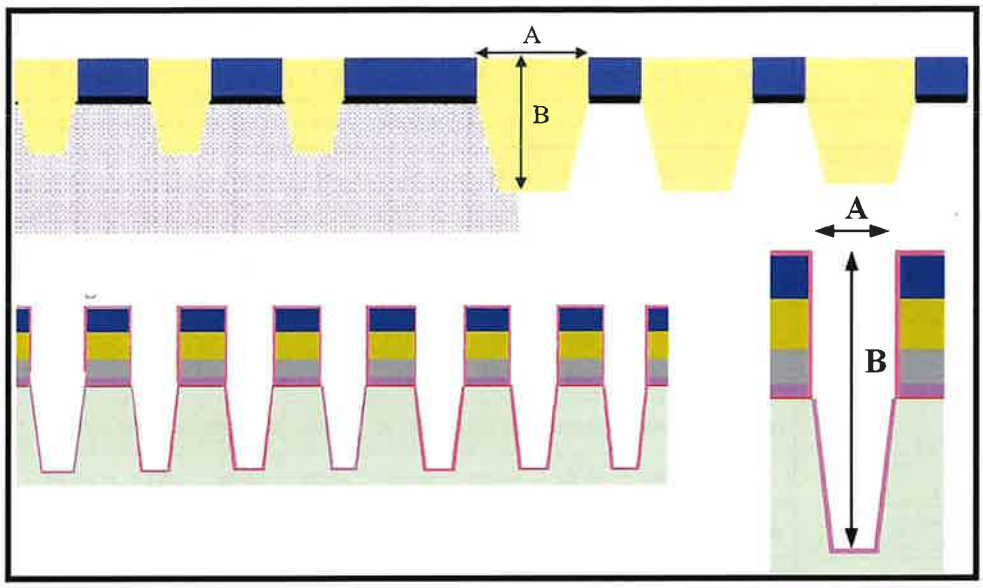


Figure FEP11 Schematics of STI Isolation Trenches

The aspect ratio is defined as the B/A ratio, including both the depth of the trench inside silicon and the height of the stack deposited on the silicon surface. The factor A is based on the minimum feature size F while the factor B depends on the type of isolation scheme utilized.

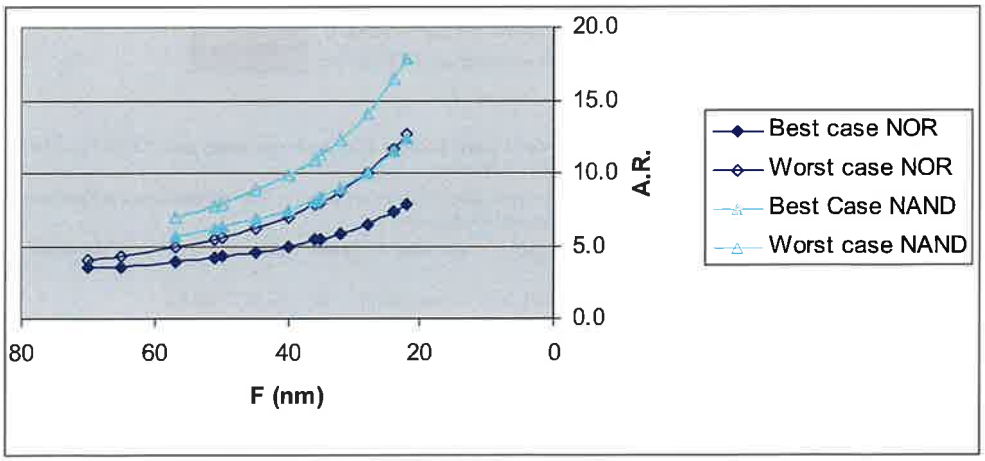


Figure FEP12 Evolution of the STI Aspect Ratio for Flash Memories with the Minimum Feature Size

Filling high aspect ratio trenches with the isolation oxide is a major challenge to be faced. Additional challenges for isolation include the overall thermal budget in the STI formation, especially in the case of self-aligned STI schemes, and the co-existence of different STI trench geometries on different parts of the chip (memory array versus I/O circuitry).

UMC 1512
UMC v. AICP

Table FEP7 FLASH Non-volatile Memory Technology Requirements

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*	*2016*	*2017*
NAND Flash poly ½ Pitch (nm) [A]	51	45	40	36	32	28	25	22	20	19	18
NOR Flash –F (nm) [A]	65	57	50	45	40	35	32	28	25	22	20
Flash NOR tunnel oxide thickness (EOT-nm) [B]	8.5-9.5	8.5-9.5	8.5-9.5	8-9	8-9	8-9	7-8	7–8	7–8	7–8	7–8
Flash NOR tunnel dielectric material [C]	Oxide	Oxide	Oxide	Oxide	Oxide	Oxide	Ox / High-κ	Ox / High-κ	Ox / High-κ	Ox / High-κ	Ox / High-κ
Flash NAND tunnel oxide thickness (EOT-nm) [B]	7-8	7-8	6–7	6–7	6–7	6–7	6–7	6–7	6–7	6–7	6–7
Flash program/erase window minimum DVT SLC/MLC (V) [D]	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4
Flash erase/program time degradation t_{max}/t_0 at constant V [E]	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2
Flash NOR interpoly dielectric thickness (EOT-nm) [F]	13–15	13–15	13–15	◆ 6–13	◆ 6–13	◆ 6–13	4–6	4–6	4–6	3–5	3–5
Flash NAND interpoly dielectric thickness (EOT-nm) [F]	10–13	10–13	◆ 5–12	◆ 5–12	4–6	4–6	4–6	3–5	3–5	3–5	3–5
Flash NAND Interpoly Dielectric Material [G]	ONO	ONO	ONO/ High-κ	ONO/ High-κ	High-κ	High-κ	High-κ	High-κ	High-κ	High-κ	High-κ
Flash interpoly dielectric thickness control EOT (% 3s) [H]	<±5	<±5	<±5	<±5	<±5	<±5	<±5	<±5	<±5	<±5	<±5
Flash interpoly dielectric T_{max} of formation t >5'/<5' (°C) [I]	750/900	750/900	750/900	650/800	650/800	650/800	600/700	600/700	600/700	600/700	600/700
Flash interpoly dielectric conformality on floating gate EOT_{min}/EOT_{max} [J]	>0.98	>0.98	>0.98	>0.98	>0.98	>0.98	>0.98	>0.98	>0.98	>0.98	>0.98
Tunnel / Interpoly max leakage current (A) at 2 V for 10 years data retention [K]	5 E-25	5 E-25	5 E-25	2.50E-25	2.50E-25	2.50E-25	1.30E-25	1.30E-25	1.30E-25	6.00E-26	6.00E-26
Flash NAND STI Filling Aspect Ratio(min-max) [L]	6.3-7.9	6.8-8.8	7.5-9.9	8.1-10.9	9-12.3	10-14.1	11.5-16.4	12.4-17.9	12.9-18.8	13.5-19.8	14.2-21.0
Flash NAND STI Filling Technology [M]	HDP/CVD	HDP/SOD	HDP/SOD	HDP/SOD	SOD	SOD	SOD	SOD	SOD	SOD	SOD
Flash NOR STI Filling Aspect Ratio(min-max) [L]	3.6-4.3	3.9-4.9	4.3-5.6	4.6-6.2	5.0-7.0	5.4-7.8	5.8-8.7	6.5-10.0	7.3-11.6	7.8-12.7	8.1-13.4

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

Notes for Table FEP7:

[A] In the past Flash devices tended to lag behind the current CMOS technology, but that delay no longer exists and NAND Flash are now leading the scaling rush. This entry provides the F value for designs in the indicated time period.

[B] Tunnel oxides must be thick enough to assure retention but thin enough to allow ease of erase/write. This difficult trade-off problem hinders scaling. Tunnel oxides less than 7 nm seem to pose fundamental problems for retention reliability.

[C] A combination of silicon dioxide and high-κ materials will be required to continue the further thickness reduction of the tunnel dielectric and to guarantee the charge retention properties of the memory device.

[D] Between minimum and maximum values of the program/erase distributions for single/multilevel cells (SLC/MLC).

[E] Time degradation after maximum specification number of write/erase cycles considering no erasing/program voltage correction.

[F] Interpoly dielectric must be thick enough to assure retention but thin enough to assure an almost constant coupling ratio. Charge retention when the dielectric is scaled downward is the major issue.

[G] High-κ interpoly will help reducing the interpoly EOT and maintain constant coupling ratio without loosing retention.

[H] Thickness control to assure correct coupling ratio and minimum thickness for charge retention.

[I] For long (>5 min) and short (<5 min) thermal processes to avoid tunnel oxide and device degradation.

[J] Uniform step coverage is important to assure charge retention, especially when the floating gate sidewall is electrically coupled with the control gate to enhance the coupling ratio.

[K] Maximum leakage current through the tunnel and interpoly dielectrics to assure ten years data retention. It is calculated considering a floating gate voltage of -2 V when the cell is programmed and a total capacitance that is a half every technology generation. In case of 20 years data retention the leakage current target value is 50% than the reported value.

UMC 1512
UMC v. AICP

[L] Ratio between the height and the length of the trench to be filled. Minimum value is typically for NOR without drain silicidation, The maximum value is typically for NAND. Height of the trench includes both the height of the stack outside silicon and the depth of the silicon trench. Actual considered vertical dimensions are based on literature data[17, 18] and the future trend is based on some overall reduction for NOR, due to operating voltage scaling, and no-silicon trench reduction for NAND, due to hard scaling of the operating voltage. Length of the trench is the minimum feature size given for the technology generation, or higher.

[M] SOD (Spin-On-Dielectric) technology will be mandatory to fill very high aspect ratio trenches

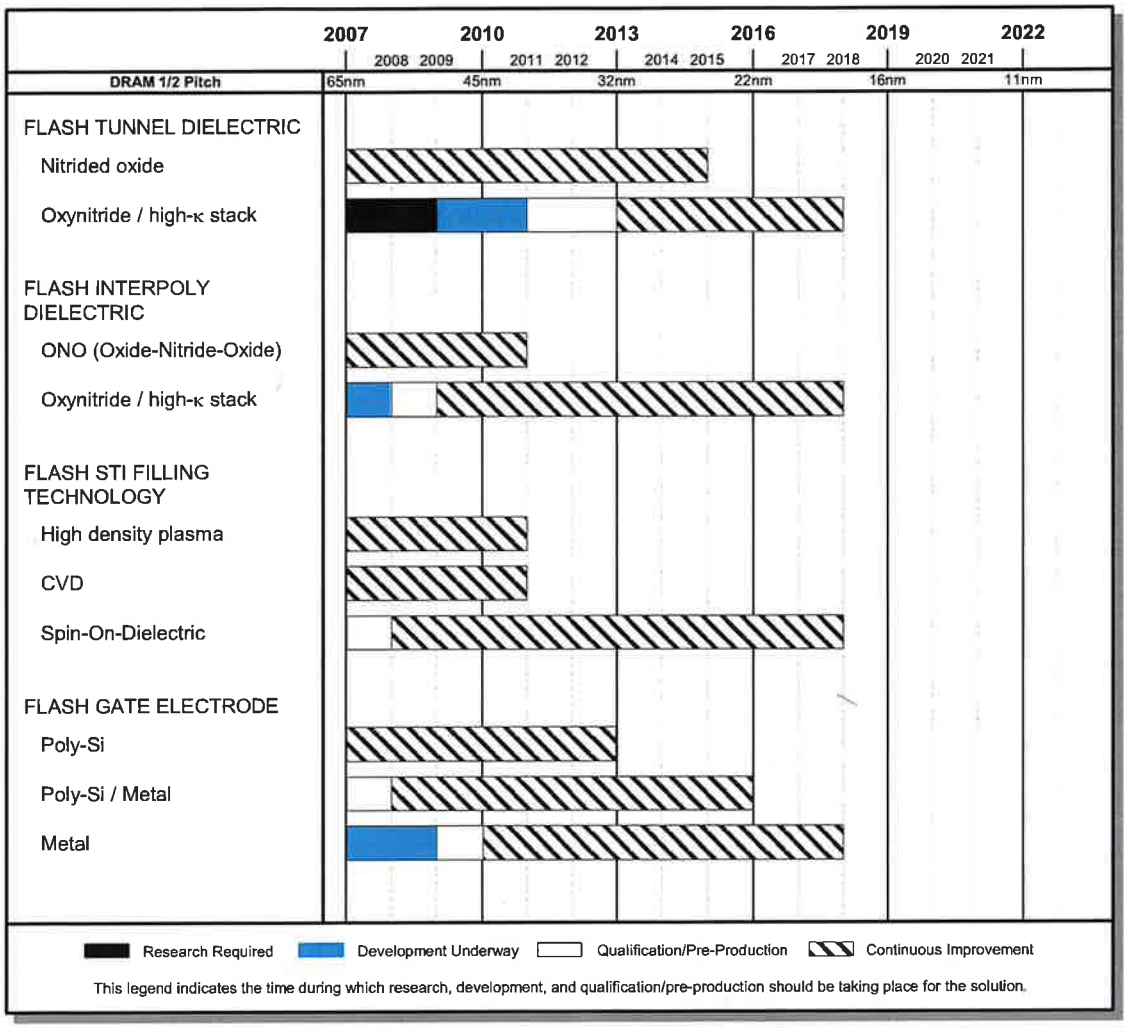


Figure FEP13 Flash Non-Volatile Memory Floating Gate Potential Solutions

CHARGE TRAPPING FLASH MEMORY TECHNOLOGY

Challenges for continued scaling of both NOR and NAND devices indicate a future need for an alternative to traditional floating gate technology. Generally, there are more issues in scaling the NOR cell than the NAND cell, however, both suffer from cross-talk effect, which mostly impacts multilevel devices. Moreover, both NOR and NAND Flash memories will face a hard limit in scaling the poly1 to poly1 pitch along the word line, due to the ONO dielectric thickness on the floating gate side wall, as discussed above. Charge trapping memory technology, based on storing charge inside a silicon nitride or a high-k dielectric, can solve these two issues.

UMC 1512
UMC v. AICP

The major advantages of the charge trapping approach are:

- Suppression of floating-gate to floating-gate disturbance
- High scalability (feasibility for FinFET structure)
- Simple integration for embedded memory applications
- Immunity to localized defects in tunnel oxide (ex: SILC effects) reducing challenges to tunnel oxide scaling
- Tight erase distributions

The most common recognized challenges of the charge trapping approach are:

- Low gate to nitride coupling ratio
- High-κ blocking layer between the trapping material and the gate is mandatory
- A high work function gate, i.e. a metal gate is mandatory in order to erase the memory cell with reasonable bias
- Optimization of charge redistribution retention after cycling related to localized charge storage inside a dielectric.

As a consequence of these challenges, some major technology issues must be faced before the charge trapping approach moves into manufacturing, including triggering of the charge trapping material and integration of the high-κ blocking layer and the metal gate.

PHASE CHANGE MEMORY

Phase Change Memory (PCM) technology is based on the basic properties of the chalcogenide alloy[2], so the integration of the material into a standard CMOS process represents a challenging matter: not for the single cell concept, already proven to be very strong, but for the manufacturability of very high density non-volatile memory, where the technology can be considered robust only if demonstrated over many billions of cells. Considering the electronic and transport properties of the chalcogenide alloy, either in the crystalline or in the amorphous state, in order to form a functional compact cell array, a PCM cell is be formed by a variable resistor (heater and chalcogenide – called data-storage) in series with a selector device (transistor). Refer to Figures FEP13, FEP14 and FEP15 for illustrations of the basic concepts involved in inducing the phase change and it's desired properties. Hence, the basic PCM cell has a 1T/1R structure. The type of transistor and of data-storage varies respectively as a function of the application and of the process architecture strategy. For high-density memory, a more compact cell layout is achieved via the vertical integration of a pnp bipolar transistor,[19, 20] while for embedded memory the transistor is a n-channel MOS, where a larger cell size is balanced by a minimum process cost overhead with respect to the reference CMOS.

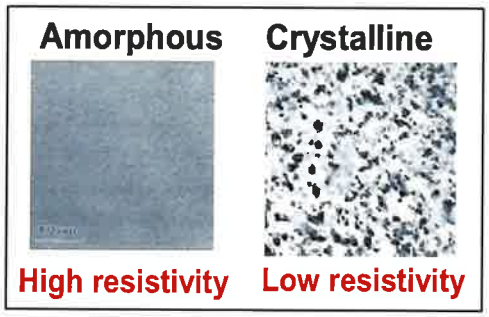


Figure FEP14 Amorphous / Poly-crystal Phases of a Chalcogenide Alloy, usually $Ge_2Sb_2Te_5$ (GST)

[2] *Chalcogenides are alloys based on the VI group elements that have the interesting characteristic to be stable at room temperature both in the amorphous and in the crystalline phase. In particular, the most promising are the GeSbTe alloys which follow a pseudobinary composition (between GeTe and Sb_2Te_3), often referred as GST.*

UMC 1512
UMC v. AICP

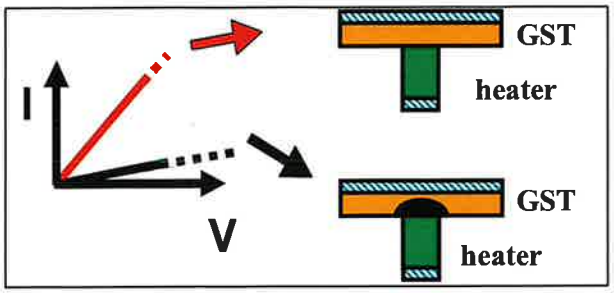


Figure FEP15 Resistance Change of GST

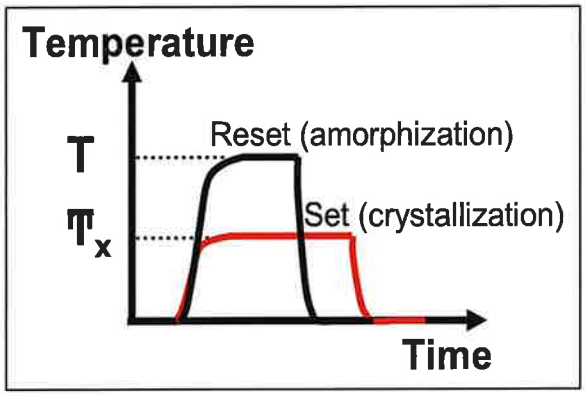


Figure FEP16 Set/Reset Thermal Cycles to Change the Crystal Phase of the GST Material and to Write/Erase the PCM

The integration of the data-storage occurs between the front-end and the back-end of the CMOS process. The "simple" variable resistor, i.e., the heater and chalcogenide system, may be obtained in different ways and the choice is a function of the understanding of the process complexity, current performances, thermal properties and scaling perspective.[21] A possible reported approach is to use a sub-lithographic contact heater with a planar chalcogenide[22] or a modified version with a recession in the contact and chalcogenide confinement, which should improve the thermal properties and hence reduce the reset current.[23]

A completely different approach relies on the definition of the contact area between the heater and the chalcogenide by the intersection of a thin vertical semi-metallic heater and a trench, called a "µtrench",[20] in which the chalcogenide is deposited. Since the µtrench can be defined by sub-litho techniques and the heater thickness by film deposition, the cell performance can be optimized by tuning the resulting contact area still maintaining a good dimensional control.

Despite the high potential of the PCM concept and the good integration results so far achieved,[24, 25] some practical challenges must still be addressed. In particular, large efforts are being dedicated to the integration of a compact PCM cell structure with the chalcogenide compound, to achieve a full compatibility with an advanced CMOS technology and to reduce the programming current without degrading the appealing features of the PCM technology. The easier integration of PCM cell is achieved with the pillar-like structure, but the resulting programming currents are quite large, thus posing additional constraints on the selecting device and on the overall power consumption. Several approaches so far proposed to reduce the programming currents of PCM devices rely on the confinement of the chalcogenide compounds in trenches,[20, 24] contact recessions[26] or contact holes.[27] The main idea is to force the maximum current crowding directly in the active region of the cell, where the chalcogenide material switches between the two phases. In fact, it has been estimated[21] that the programming currents can be reduced up to the 50% by employing fully confined structures with the chalcogenide material that fills a contact hole. The superior capabilities of confined structures has been demonstrated by the µTrench PCM cell architecture that achieved a programming current of 450 µA at 180 nm[20] and 350 µA at 90 nm.[24] A

UMC 1512
UMC v. AICP

programming current of 260 μA in a 50 nm contact has been demonstrated for a fully confined structure obtained with a CVD-deposited chalcogenide.[27] These results point out that the fabrication of efficient PCM cell architectures will be required to fill confined structures with very high aspect ratio. A continuous improvement of the chalcogenide material minimum conformality is thus expected to support the cell architecture evolutions.

To support reliable large array products, PCM technologies must be able to retain data over the product's lifetime with very low defect rates. Data retention is limited by resistance loss of the amorphous phase of the material, a process that is controlled by the kinetics of crystallization. Prior work on data retention of reset cells shows that data retention of GST is much longer than ten years at 85°C, and therefore adequate for typical non-volatile memory applications.[28, 29] This value is satisfactory to address consumer applications, but it is not matching industrial requirements for high temperature operation (e.g., automotive applications). This drives a need for improvement of the maximum storage temperature. This improvement will mainly come from the development of different chalcogenide compounds and compositions.

One of the most attracting features of PCM technology is the expected superior endurance when program and erase repetitive operation are performed. Several publications reports endurance capabilities that range from 10^7 up to 10^{12} cycles. Such impressive results depend on the intrinsic endurance of the chalcogenide compounds as well as on the overall stability of the PCM cell surrounding material. Among them, the heater electrode is the part of the cell that undergoes to the heavier stressing conditions, with temperatures much higher than the 600°C and current densities that can exceed 1 A/μm^2. The most important electrical property of the heater material that must be preserved during cycling is the electrical resistivity, which is requested to remain stable according to the endurance specifications. The maximum resistivity variation is thus intended to provide a guideline on the main electrical property of the heating element in PCM cells, providing the requirement that guaranties to be able, with the same current, to switch the PCM cells among the two logic states for the required number of P/E cycles.

The requirement on the heater resistance stability is intimately related to the maximum reset current density specification. It has been reported that, under the simple assumption of isotropic scaling, the expected reset current density will increase linearly with the scaling factor,[21] and a more aggressive trend could be expected according to the forecasted roadmap. A detrimental effect of this increase could be more aggressive stressing conditions for the heater material and for the chalcogenide-heater interface, that should be faced with a slower growth of the required current density. It follows that a better heating efficiency will be required to downscale the PCM devices that could be achieved through an increase of the heater resistivity, still preserving the stability requirements. It is thus expected that the scaling roadmap of PCM technology will face the need to provide novel material for the heating electrode capable to reach a trade off among the endurance requirements and the performance specifications.

UMC 1512
UMC v. AICP

Table FEP8a Phase Change Memory (PCM) Technology Requirements—Near-term Years

Grey cells indicate the requirements projected for years before it reaches volume production.

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
PCM ½ Pitch (nm) (contacted)	65	57	50	45	40	35	32	28	25
Phase change material min. conformality (%) [A]	30	60	70	70	90	90	90	90	90
PCRAM phase change material minimum operating temperature (°C) [B]	85	100	100	100	125	125	125	125	125
Heater max resistivity change during reset cycle and after 1E12 cycles (%)	5	5	5	2	2	2	1	1	1
Maximum Reset Current Density (A/μm²)	0.3-0.8	0.3-0.8	0.3-0.8	0.4-1.0	0.4-1.0	0.4-1.0	0.5-1.3	0.5-1.3	0.5-1.3

Table FEP8b Phase Change Memory (PCM) Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
PCM ½ Pitch (nm) (contacted)	22	20	18	16	14	12	10
Phase change material min. conformality (%) [A]	90	90	90	90	90	90	90
PCRAM phase change material minimum operating temperature (°C) [B]	125	125	125	125	125	125	125
Heater max resistivity change during reset cycle and after 1E12 cycles (%)	1	1	1	1	1	1	1
Maximum Reset Current Density (A/μm²)	0.5-1.3	0.5-1.3	0.5-1.3	0.5-1.3	0.5-1.3	0.5-1.3	0,5-1.3

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known

Notes for Tables FEP8a and FEP8b:

[A] Conformality requirements of the phase change material are given for μtrench PCM to optimize the operating current and the formation technology [20, 24, 26, 27, 30, 31, 32, 33]

[B] Expected operating temperature for a PCM that guarantees a minimum data retention capability of 10 year [22, 32, 34]

[C] Expected maximum resistance variation of the heater to match the PCM W/E (writing/erasing) cycling endurance specifications reported in the PIDS Table "Non-volatile Memory Technology Requirements" [21, 22, 35]

[D] Expected maximum current density required to reset the PCM cell (from crystalline to amorphous state). It is expected that current density could affect the PCM cell reliability in term of endurance. The reported value is thus intended to meet the endurance specification reported in the PIDS Table "Non-volatile Memory Technology Requirements" [21, 35]

Ferroelectric Random Access Memory (FeRAM)

FeRAM (also abbreviated FRAM) was a new addition to the 2001 ITRS, and was the result of collaboration between the FEP and PIDS technology working groups. The critical requirement tables, Tables FEP9a and FEP9b were revised in 2007 based on the results of a survey of the FeRAM manufacturers.

Historically speaking, FeRAM devices had been proposed much earlier than semiconductor memory devices.[36] At present however, memory capacity is limited to ~1/1000 of that of commodity DRAM, due to difficulties associated with capacitor fabrication, integration and reliability. These difficulties together with the lack of a "killer application" have constrained commercial production. FeRAMs depend substantially on the continued development of materials such as ferroelectric films which make forecasts presented here somewhat speculative. Nevertheless, the roadmap covers the years 2007 to 2022 in order to provide a strategic overview of the technology directions and the challenges that must be overcome.

Mass Production Based Table Entries

Since the FeRAM table was introduced in 2001, the specifications that have been included in the tables from 2001–2006 reflected FeRAM technology that was presented at conferences because obtaining accurate information on devices in the marketplace was not readily available. As a result, a large gap existed between what FeRAM semiconductor manufacturers presented at conferences and what was commercially available on the market. To eliminate this gap, the specifications for the 2007 table were defined from judgment comparing three different measures. The first metric is the manufacturer's specification available on their homepage. The second metric is derived from surveys. The third metric

UMC 1512
UMC v. AICP

uses a previous precedent established for DRAMs, where the level of technology in the roadmap is based upon the two leading manufacturers which have achieved a production volume of at least 10,000 chips per month.

MIXED SIGNAL DEVICES AND FEATURE SIZE

As noted above, FeRAM technology has lagged behind leading edge memories such as Flash and DRAM. Although this gap exists, manufacturers have developed devices which use advanced CMOS such as 0.13um technology but use 0.18 μm technology for the Metal 1 half-pitch width for FeRAM as shown in Table FEP9a. Combining advanced CMOS technology with relaxed designed rules for FeRAM, is expected to increase the number of applications for FeRAM. Table FEP9a shows a feature size of 0.13 μm for 2007 commercial product using the same criteria as DRAM. Feature size scaling is forecasted to occur at approximately 0.7× every three years, which is at a slower pace than other established memories.

CELL SIZE

Currently, the most efficient cell structure is the one transistor-one capacitor (1T-1C) cell. It is replacing the 2T-2C cell that is less efficient but offers greater operating range stability. However, in the market both cell structures will likely be available for some time depending upon the device application. As far as the capacitor structure is concerned, the change from the planar capacitor type to a stack configuration has resulted in a cell size reduction. The timing of a shift from a stacked structure to a 3D structure will depend upon the ferroelectric material used and it is expected to occur approximately in 2016. The different capacitor configurations are shown in the drawing accompanying Tables FEP9a and FEP9b. The cell structure and capacitor configuration changes are forecasted to reduce the cell area factor to 16 in 2013–2015 after which the cell area factor will continue with further scaling.

FERROELECTRIC MATERIALS ALTERNATIVES

There are several ferroelectric materials under evaluation at the present time,[37] but at present there is no clear, single material choice. The two current materials are PZT, or $Pb(Zr,Ti)O_3$ and SBT, or $SrBi_2Ta_2O_9$. SBT has superior fatigue-free characteristics with a Pt bottom electrode and is more suitable for low voltage operation because of its smaller coercive field (E_c). (Fatigue is defined as a loss of polarization or charge that develops after bipolar cycling of the memory capacitor.) PZT has a larger switching charge per unit area, Q_{SW}, which is important since it is allows for further scaling without shifting to a 3D cell. Both materials may suffer damage due to process integration during the device fabrication which has hampered device development.

The most important issues with PZT and SBT films are suppression of film deterioration that is attributed to hydrogen diffusion[38] and oxygen loss; the achievement of stable data read/write characteristics; and data retention during integration. Process improvements are also required for embedding FeRAM. It is important to avoid high temperature annealing or hydrogen incorporation into ferroelectric films after the oxygen anneal used to crystallize the films. For example, AlOx and TiN are often used as hydrogen barrier layers. Also, conductive oxides such as IrO_2 or $SrRuO_3$ (SRO) are often used as capacitor electrode materials for PZT since their use improves ferroelectric capacitor reliability.

Physical vapor deposition (PVD) and chemical solution deposition (CSD) including Sol-Gel methods are currently the most commonly used methods for ferroelectric film deposition. However, continued scaling dictates the need to shift to methods with better step coverage such as MOCVD as noted in Figure FEP17. A previously reported MOCVD study has shown that a (111) oriented PZT film is very effective at increasing the switching charge.[39] Etching of capacitor electrodes remains a challenge with RIE, because the capacitor electrodes do not react to form volatile etch by-products. Therefore, sputter etching is widely used. This limits critical dimension (CD) control and makes scaling more difficult. High temperature etching technology for improving sidewall slope of the capacitor was developed to overcome this difficulty.[39]

PZT and SBT are often doped to improve their electrical properties. For instance, PZT may be doped with lanthanum, and SBT with niobium. Doping is used to achieve the following film enhancements: leakage current suppression, improved endurance or imprint characteristics, suppression of post process film degradation, and others. Besides PZT and SBT, one of the promising new materials is BLT or $(Bi,La)_4Ti_3O_{12}$,[40] of which characteristics are in-between the foregoing two.[36] In addition, $BiFeO_3$ (BFO) has gained much attention as a new candidate material. BFO has a giant ferroelectric polarization of $150\mu C/cm^2$ or more.[41] Although BFO exhibits a large polarization, it also requires a higher switching voltage, which means that the film needs to be thinner or possibility doped to accommodate low voltage operation. Since the ferroelectric properties of each material have improved in recent years due improvement in process technology, it essential for the process to be optimized for the integrated ferroelectric capacitor in order to obtain good ferroelectric properties.

UMC 1512
UMC v. AICP

Estimated Minimum Switching Charge

The estimated minimum switching charge has been derived as follows: The sense amplifier for FeRAM is assumed to be basically the same as that of DRAM. Therefore, the bitline signal voltage was calculated using DRAM data from the 1999 ITRS. These data provide that the capacitance Cs remain constant at 25fF/cell independent of the dimension, and the bitline capacitance is 320fF at 0.18 µm. Based on this data with the further assumption that bitline capacitance is proportional to $F^{2/3}$, where F is the feature size[42] that allows for the calculation of $\Delta V_{bitline}$. The $\Delta V_{bitline}$ is about 140 mV, and the assumption is that this is needed for the sense amplifier circuit independent of technology generation. Multiplying $\Delta V_{bitline}$ (140 mV) with $C_{bitline}$ then gives the minimum switching charge.

Dividing the minimum switching charge value derived above by the ferroelectric film switching charge per unit area, Q_{SW}, (assumed to be 30 $\mu C/cm^2$) then yields the desired capacitor area. If this area is larger than the projected capacitor size, then a 3D capacitor should be adopted. Based on this, a 3D capacitor will be needed by year 2016.

The FeRAM forecast of Tables FEP9a and FEP9b is based on these assumptions and calculations. "Red brick walls" begin to appear in 2016. The first priority to break through these walls is the development of highly reliable ferroelectric materials that exhibit negligible process induced degradation.

Endurance

An endurance of 10^{15} read/write cycles is required to replace other RAMs such as SRAM and DRAM. In order to confirm such endurance values, standardized testing within a practical time period is needed based upon accelerated testing methods with an underlying physical model is needed. There are several models in the literature on degradation of ferroelectric capacitors due to endurance testing but so far there are few reports on degradation on integrated capacitors.

Recently, FeRAMs are being used for IC cards and the personal authentication, etc. utilizing the feature of fast program speed and high endurance instead of EEPROM or Flash memory. Security applications show strong potential to be a growing market for FeRAM.

An encouraging fact is that the storage capacity of commodity Flash memory has dramatically increased and is currently almost equal to or even greater than that of DRAM. This increase has occurred because of market demand for large capacity, non-volatile memory. FeRAM may also satisfy this market demand and therefore can be combined with Flash to speed up applications such as solid state disk drives and security log-ins for portable devices. Global efforts by researchers to address fundamental issues in ferroelectric materials and reliability as well further market development for FeRAM application are needed for further growth.

UMC 1512
UMC v. AICP

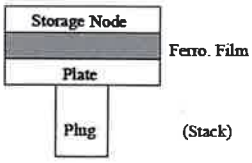
Table FEP9a FeRAM Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
FeRAM technology – F (nm)[A]	180	180	180	150	150	150	130	130	130
FeRAM cell size – area factor a in multiples of F2 [B]	22	22	22	20	20	20	16	16	16
FeRAM cell size (μm^2) [C]	0.713	0.713	0.713	0.450	0.450	0.450	0.270	0.270	0.270
FeRAM cell structure [D]	2T2C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C
FeRAM capacitor structure [E]	stack	stack	stack	stack	stack	stack	stack	stack	stack
FeRAM capacitor footprint (μm^2) [F]	0.330	0.330	0.330	0.199	0.199	0.199	0.106	0.106	0.106
FeRAM capacitor active area (μm^2) [G]	0.330	0.330	0.330	0.199	0.199	0.199	0.106	0.106	0.106
FeRAM cap active area/footprint ratio	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ferro capacitor voltage (V) [I]	1.50	1.50	1.50	1.20	1.20	1.20	1.20	1.20	1.20
FeRAM minimum switching charge density ($\mu C/cm^2$) [J]	13.5	13.5	13.5	19.9	19.9	19.9	34.0	34.0	34.0
FeRAM endurance (read/write cycles) [K]	1.0E+14	1E+14	1E+14	1E+14	1E+14	1E+14	1E+15	1E+15	1E+15
FeRAM nonvolatile data retention (years) [L]	10 Years	10 Years	10 Years	10 Years	10 Years	10 Years	10 Years	10 Years	10 Years

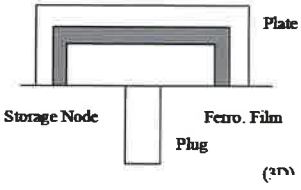
Table FEP9b FeRAM Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
FeRAM technology – F (nm) [A]	90	90	90	65	65	65	65
FeRAM cell size – area factor a in multiples of F2 [B]	14	14	14	12	12	12	12
FeRAM cell size (μm^2) [C]	0.113	0.113	0.113	0.051	0.051	0.051	0.051
FeRAM cell structure [D]	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C
FeRAM capacitor structure [E]	3D	3D	3D	3D	3D	3D	3D
FeRAM capacitor footprint (μm^2) [F]	0.041	0.041	0.041	0.016	0.016	0.016	0.016
FeRAM capacitor active area (μm^2) [G]	0.100	0.100	0.100	0.069	0.069	0.069	0.069
FeRAM cap active area/footprint ratio	2.46	2.46	2.46	4.25	4.25	4.25	4.25
Ferro capacitor voltage (V) [I]	1.00	1.00	1.00	0.70	0.70	0.70	0.70
FeRAM minimum switching charge density ($\mu C/cm^2$) [J]	30	30	30	30	30	30	30
FeRAM endurance (read/write cycles) [K]	>1.0E16	>1.0E16	>1.0E16	>1.0E16	>1.0E16	>1.0E16	>1.0E16
FeRAM nonvolatile data retention (years) [L]	10 Years	10 Years	10 Years	10 Years	10 Years	10 Years	10 Years

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known
Manufacturable solutions are NOT known



Notes for Tables FEP 9a and 9b:



UMC 1512
UMC v. AICP

[A] Feature size "F" is defined as the critical dimension of cells currently in mass production.
[B] Cell size = a × F2.
[C] {(cell size)1/2 - (capacitor space)}2 is assumed, where capacitor space = 1.5 × F.
[D] Besides these cell structures, other configurations are being investigated; ex. Chain-FeRAM.
[E] See illustrations of capacitor structures stack and 3D to the right.
[F] 3D is assumed to be a pedestal structure.
[G] Greater than 1 for 3D capacitors, otherwise 1.
[I] Vop=operational voltage. Low voltage operation is a key issue. Matsushita's 0.18 μm FeRAM with SBT in 2003 operates at 1.1V.
[J] First, the minimum switching charge per cell is estimated by multiplying ΔVbitline × Cbitline, assuming ΔV bitline=140 mV and bitline capacitance = F2/3 which is the same as in DRAM. Then the value is divided by the capacitor area for stacked structures to estimate the minimum switching charge density. The value used for 3D structures is 30.
[K] 100 MHz×10 years=3E+16. Endurance of 1E+15 is required to compete with SRAM and DRAM.
[L] Depends on particular application. 85°C is the temperature specified for IC cards.

INTER-FOCUS ITWG DISCUSSION

Front End Processing shares numerous issues and dependencies with other Focus ITWGs. Chief among those are issues surrounding gate EOT and leakage requirements with the PIDS and to some extent the Design ITWGs. Other issues with these ITWGs revolve around junction depth and sheet resistance requirements as well as requirements driven by alternate device structures. Resolution of these issues is generally attained through compromise and trade-offs. A joint effort between FEP, PIDS, Design, and Lithography is focused on gate length CD tolerance as well as gate length scaling. Despite the relaxation in overall CD tolerance from 10% to 12% starting with the 2005 ITRS, control still remains an issue. This joint effort will also assess the current project rate of gate length scaling and make adjustments in future editions of the ITRS. Also, it is expected that there will be more discussions with PIDS regarding the optimal trade-off of device parameters in non-traditional MOSFETS such as FinFET and other multi-gate transistors. Other interactions include those with the Yield Enhancement ITWG to validate different statistical defect models. A very important interaction continues to be with the Interconnect ITWG, where members of the FEP Surface Preparation team provide technical support in the development of interconnect surface preparation and cleaning technical requirements and potential solutions.

IMPACT OF FUTURE EMERGING RESEARCH DEVICES

Significant challenges must be overcome to continue shrinking integrated circuit technologies and, in the long term, more radical devices may need to be integrated with CMOS to continue increasing performance. Emerging research devices include both memory and logic devices and while these are still in research, challenging issues must be overcome to integrate these with CMOS. Many of the ERD memory and some of the logic devices are based on conventional charge state technology and could use processing modules that are currently on the FEP roadmap. These are briefly highlighted. Many of the longer term ERD would use new device materials and introduce new process modules and integration complexity, and these devices are very speculative and intercept timing has not been defined. This is discussed in detail in the ERD chapter.

Emerging Research Memory Devices

- Engineered tunnel barrier
- Ferroelectric FET
- Fuse/Antifuse
- Ionic memory
- Electronic effects
- Nanomechanical
- Macromolecular
- Molecular

Emerging Research Logic Devices

- Ferromagnetic (including magnetic QCA)
- FET extension–1D structures

UMC 1512
UMC v. AICP

- FET extension–channel replacement
- Resonant tunneling
- Molecular (including electric QCA)
- Single electron transistor (SET)
- Spin transistor

Of these devices, the nanofloating gate, SET, and RTDs could use many existing processes, but would probably need an engineered dielectric. The 1D structures (nanotubes, nanowires, etc.) will need new processes to control diameter, location, orientation, and new doping processes. The polymer and molecular devices would require low temperature processing and reliable contacts that are compatible with CMOS integration. The other devices will introduce more radical materials that will require significant work to make them compatible with CMOS processing.

The 1D structures require catalyst and CVD processes optimized to control diameter, structure, location, and orientation. They will also need new processes to selectively dope these 1D structures and new contact materials and processes to form the low resistance contacts. Nanowires would require extremely tight control of dopant ion implant dose and energy, and new high-κ gate dielectric processes may be needed to passivate the multi orientation surfaces of Si, SiGe or Ge. Carbon nanotubes require new doping processes that do not currently exist, and new gate dielectrics and gate electrodes may be needed to control threshold voltages.

Insulator resistance phase change memory and Ferroelectric FET memory would introduce a radical new material that would require a new deposition capability and new etches and cleans. These materials are often complex metal oxides that must be deposited at high temperature, and contact formation and integration may be challenging.

Traditionally some RTDs are fabricated with III-V semiconductors, and this would introduce complex new processes and materials into the FEP for integration with CMOS. Recent work has demonstrated devices made of SiGe that would require integration, but many challenges must be overcome with these materials, particularly to achieve peak/valley I/V ratios > 5. Further, the best use of Si and SiGe based RTDs is their integration into a CMOS gate, which brings another set of complex materials and integration issues.

Spin transistors will require integration of radical new materials with CMOS and this will require new deposition capabilities and introduce much process complexity. These devices are very speculative at this time, but some include GaMnAs, GeMn, as well as spin injection from ferromagnetic materials into the semiconductor that has dramatic contamination challenges.

The level of process complexity for Emerging Research Devices will continue to increase as new materials are used and then integrated on the CMOS platform. This will require development of new deposition, etch, and clean processes and new barrier layer and contact technologies.

Cross-cut Discussion

FEP Metrology Cross-Cut Issues

FEP Metrology continues to face the challenges associated with rapid introduction of new materials, processes, and structures. In the area of starting materials and surface preparation, detection of small particles (<30 nm) continues to be a challenge. The multiple layer interfaces found in new substrates such as strained silicon on insulator (sSOI) add complexity to measurement of substrate property uniformity. As high-κ materials enter manufacturing, film metrology for both the dielectric and metal gate must become more capable to support process control of the complicated layer structures and the elemental composition of each layer. In addition to new gate stack materials, other new materials and structures are expected to be introduced in future generations. Sidewall measurements, feature shape, and dimension control remain a challenge for new structures such as FinFETs. Electrical measurement capability needs to evolve with the introduction of the new high-κ stacks.

FEP Modeling and Simulation Cross-Cut Issues

The FEP challenges surround the introduction of new materials and of non-classical CMOS. This raises various requirements on Modeling and Simulation. Especially, in the coming era of material-limited device scaling, material issues need to be addressed in most modeling areas. This includes among others strained materials, so the importance of modeling of stress and strain is further growing. New device architectures request especially large progress in numerical device simulation, together with improvements of the simulation of the process steps used to fabricate these devices, e.g.,

UMC 1512
UMC v. AICP

the formation of shallow junctions. Both shrinking device dimensions and the non-planar architectures, especially SOI devices, increase the impact of interfaces because the volumes in-between are decreasing. These effects must be appropriately included in the physical process and device models. Process variations are getting increasingly important as devices further scale—a premier example is the redistribution of variance allowance between lithography and etching in the 2005 roadmap—and simulation can and must contribute to assessing the impact of such variants on the final device and chip. High-κ dielectrics are required to be introduced by 2008, so modeling must be able to appropriately describe them as soon as possible. The formation of ultra-shallow, abrupt, highly activated drain extensions continues to be a major challenge, and support from modeling is required both to improve the physical understanding for the processes used (for example, kinetics of dopants and point defects during annealing) and to subsequently optimize them by numerical simulation. This knowledge is also needed for defect engineering, which aims at achieving shallower junctions by the exploitation of the interaction between dopant atoms and defects. Furthermore, the reduction of critical dimensions (CD) and the control of their variations including LWR and LER are generally a key issue. It is highly desirable to use simulation to identify the most important parameters among the many influencing CD in order to minimize experimental effort.

FEP Environment, Safety, and Health Cross-Cut Issues

Refer to the *Environment, Safety, and Health* chapter for comprehensive information and link to a new chemical screening tool (Chemical Restrictions Table).

Conclusion

The FEP chapter of the 2007 ITRS has attempted to clearly identify the challenges and potential solutions for continued evolution of the integrated circuit beyond traditional scaling. During the next several years front-end processes will require the continued introduction of new materials such as high-κ dielectrics and highly-engineered metal films for applications as diverse as MOSFET gate stacks, DRAM storage capacitors, and Flash-memory storage devices. In addition to these new materials, new device structures, such as FinFET, will be introduced in order to meet performance requirements. Market growth for alternative memories will also require the development and optimization of a broad class of ferroelectric, magnetic, and phase-change thin film materials. Underlying these device changes are rapidly evolving requirements for substrates, such as SOI, and the need for 450 mm diameter substrate within the next five to seven years.

The transition from extended bulk CMOS to non-classical device structures is not expected to take place at the same time for all applications and all chip manufacturers. Instead, a scenario is envisioned where a greater diversity of technologies are competitively used at the same point in time—some manufacturers choosing to make the transition to non-classical devices earlier, while others emphasize extensions of bulk technology. To support this probable scenario the FEP team has provided metrics for parallel paths showing what is required to extend classical CMOS and what can be gained by making a transition to other device structures such as fully-depleted SOI and multi-gate.

The team also notes the acceleration of Flash-memory applications as a new driver of process technologies, such as critical dimension etching. The rapid expansion of the market for Flash memories will bring more focus on the material and process challenges for these devices.

Pathways around potential barriers are found by close cooperation between different ITRS technology working groups. This was demonstrated over the past two years by continuous discussion of the physical gate length variation challenge between the FEP, PIDS, Lithography, and Design groups. This collaboration resulted in changes across several chapters of the ITRS including a shift between printed dimensions and etch bias, a redistribution of allowable variation, and a recognition that devices can be economically manufactured with slightly higher variation than previously prescribed. Continued cross-TWG collaborations such as this will be crucial to finding pathways around future barriers.

UMC 1512
UMC v. AICP

REFERENCES

[1] W. Maly, H.T. Heineken, and F. Agricola, "A Simple New Yield Model," Semiconductor International, No. 7, 1994, 148–154.

[2] Y. Omura, S. Nakashima, K. Izumi, and T. Ishii, "0.1mm-Gate, Ultrathin-Film CMOS Devices Using SIMOX Substrate with 80-nm-Thick Buried Oxide Layer," IEDM Tech. Digest, 1991, 675–678.

[3] W. P. Maszara, R. Dockerty, C.F.H. Gondran and P.K. Vasudev. "SOI Materials for Mainstream CMOS Technology," in: "Silicon-on-Insulator Technology and Devices VIII," S. Cristoloveanu, P.L.F. Hemment, K. Izumi and S. Wilson, eds., PV 97-23, The Electrochemical Society Proceeding Series, Pennington, NJ, 1997, 15–26.

[4] H. Aga, M. Nakano and K. Mitani. "Study of HF Defects in Thin Bonded SOI Dependent on Original Wafers," Extended Abstracts of the 1998 International Conference on Solid State Devices and Materials, Hiroshima, Japan (1998), 304–305.

FRONT END PROCESSES SURFACE PREPARATION

[5] P.W. Mertens, "Advanced Cleaning Technology," invited tutorial, UCPSS 2000, (2000), Ostende, Belgium, 31–48.

[6] P.W. Mertens, et. al., "Recent Advances in Wafer Cleaning Technology," Semicon Europa Front End Technology Conference, Munich, (April 24, 2001).

[7] P. W. Mertens, T. Bearda, M. Houssa, L. M. Loewenstein, I. Teerlinck, R. Vos, I. Cornelissen, S. De Gendt, K. Kenis, M. Meuris and M. M. Heyns, "Advanced Cleaning for the Growth of Ultrathin Gate Oxide", 11th International Conference on Insulating Films on Semiconductors, Erlangen, Germany (June 16-19, 1999), in Microelectronic Engineering 48, 1999, 199.

THERMAL/ THIN FILMS/ DOPING

[8] Y. Taur, "25 nm CMOS Design Considerations," IEDM 1998, Technical Digest, IEEE, December 1998, 789–792.

[9] A. Srivastava and C.M.Osburn, "Response Surface Based Optimization of 0.1 μm PMOSFETs with Ultra-Thin Oxide Dielectrics," SPIE Proc., Vol. 3506, 1998, 253.

[10] C.M. Osburn, J.Y. Tsai and J. Sun, "Metal Silicides: Active Elements of ULSI Contacts," J. Electronic Mater., Vol. 25(11), 1996, 1725.

[11] P. Zeitzoff and A. Tasch, "Modeling of Manufacturing Sensitivity and of Statistically Based Process Control Requirements for 0.18 micron NMOS device," Characterization and Metrology for ULSI Technology: 1998 International Conference, D.G. Seiler, et al. eds., 73.

FRONT END PROCESSES ETCHING

[12] A. Agarwal and M. Kushner, "Plasma Atomic Layer Etching," 33rd IEEE International Conference on Plasma Science, p. 469, (2006).

[13] T. Wallow, A. Acheta, Y. Ma, A. Pawloski, S. Bell, B. Ward, C. Tabery, B. La Fontaine, R.-H. Kim, S. McGowan and H. J. Levinson, "Line-edge roughness in 193-nm resists: lithographic aspects and etch transfer", Advances in Resist Materials and Processing Technology XXIV. Edited by Lin, Qinghuang. Proceedings of the SPIE, Volume 6519, pp. 651919 (2007).

[14] N. Lindert, L. Chang, Y.-K. Choi, E. H. Anderson, W.-C. Lee, T.-J. King, J. Bokor, and C.-M. Hu, "Sub-60 nm Quasi-Planar FinFETs fabricated using a simplified process", IEEE Electron Device Letters, vol. 22, No. 10, pp. 487-489, (2001).

[15] R. Singh, P. J. Kelley, L. B. Fritzinger, C. C. Lee, and J. S. Molloy, "Shallow trench isolation method providing rounded top trench corners", US Patent 6,174,786, January 16, 2001.

DRAM STACKED CAPACITIOR

[16] N. Fukushima, et al., "Epitaxial (Ba, Sr) TiO_3 Capacitors with Extremely High Dielectric Constant for DRAM Applications," IEDM 1997, Technical Digest, Dec. 1997, pages 257-260.

NON-VOLATILE MEMORY

[17] Hong-Ho Park et al., "8 Gb MLC (Multi-Level Cell) NAND Flash Memory using 63 nm Process Technology", IEDM Tech. Dig. 2004.

[18] G. Servalli et al., "A 65nm NOR Flash Technology with 0.042 μm^2 Cell Size for High Performance Multilevel Applications", IEDM Tech. Dig. 2005.

PHASE CHANGE MEMORY

[19] S.Lai and T.Lowrey, "OUM – A 180nm NVM cell element technology for stand alone and embedded applications", IEDM Tech. Dig., 2001.

[20] F. Pellizzer et al., "Novel utrench Phase-Change Memory Cell for Embedded and Stand-Alone Non-Volatile Memory Applications," Symp. on VLSI Tech., pp. 18-19, 2004.

UMC 1512
UMC v. AICP

[21] A. Pirovano et al., "Scaling analysis of phase-change memory technology", IEDM Tech. Dig., p.699-702, 2003.

[22] S.J.Ahn et al., "Highly manufacturable high density Phase Change Memeory of 64Mb and beyond", 2004 IEDM, 37.3, pp. 911-914.

[23] S.J.Ahn et al., "Highly reliable 50nm contact cell technology for 256Mb PRAM", Dig. of 2005 Symp. on VLSI Tech., p.98, 2005.

[24] F. Pellizzer *et al.*, "A 90nm Phase-Change Memory Technology for Stand-Alone Non-Volatile Memory Applications", *Symp. on VLSI Tech.*, pp. 122-123, 2006.

[25] S. Kang *et al.*, "A 0.1μm 1.8V 256Mb 66MHz Synchronous Burst PRAM", *ISSCC Tech. Dig.*, pp. 140-141, 2006.

[26] Y. N. Hwang et al., "Writing Current Reduction for High-density Phase-change RAM",*IEDM Tech. Dig.*, 2003.

[27] J. I. Lee et al., "Highly Scalable Phase Change Memory with CVD GeSbTe for sub 50nm Generation", *Symp. on VLSI Tech.*, pp. 102-103, 2007.

[28] K. Kim and S.-J. Ahn, "Reliability investigations for manufacturable high density PRAM," in *Proc. IRPS*, pp. 157-162, 2005.

[29] A. Pirovano et al., "Reliability study of Phase-Change Nonvolatile Memories," *IEEE TDMR* 4 (3) 2004.

[30] Y.H.Ha et al., "An edge Contact Type Cell for Phase Change RAM Featuring Very Low Power Consumption", VLSI Symp. 2003.

[31] B. J. Choi et al., "Cyclic PECVD of $Ge_2Sb_2Te_5$ Films Using Metallorganic Sources", *J. of the Electrochemical Society,* 154, 2007.

[32] R.-Y. Kim et al., "Structural properties of $Ge_2Sb_2Te_5$ thin films by metal organic chemical vapor deposition for phase change memory applications", *Appl. Phys. Lett.*, 89, 102107, 2006.

[33] J. Lee et al., "GeSbTe deposition for the PRAM application", *Applied Surface Science*, 2006.

[34] A. L. Lacaita et al., "Electrothermal and Phase Change Dynamics in Chalcogenide-Based Materials", 2004 IEDM, 37.3, pp. 911-914.

[35] S. Lai, "Current Status of Phase Change Memory and Its Future", 2003 IEDM, pp. 255-258.

FeRAM

[36] J. L. Moll and Y. Tarui, IEEE Trans. Electron Devices, ED10, 338, 1963.

[37] D. J. Wouters, 28, International Conference on Solid State Devices and Materials, 2003.

[38] J. S. Cross, Y. Horii, N. Mizuta, S. Watanabe and T. Eshita, Jpn. J. Appl. Phys. 41 (2002) 698.

[39] Y. Horii, Y. Hikosaka, A. Itoh, K. Matsuura, M. Kurasawa, G. Komuro, K. Maruyama, T. Eshita and S. Kashiwagi, 539, IEDM, 2002.

[40] B. H. Park, B. S. Kang, S.D. Bu, T. W. Noh, J. Lee, and W. Jo, 682, Nature, 1999.

[41] K. Y. Yun, D. Ricinschi, T. Kanashima, M. Noda and M. Okuyama, Jpn. J. Appl. Phys. 43(2004)L647.

[42] A. Nitayama, Y. Kohyama, and K. Hieda, 355, IEDM, 1998.

UMC 1512
UMC v. AICP