INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

2007 EDITION

PROCESS INTEGRATION, DEVICES, AND STRUCTURES

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

TSMC-1016
U.S. Patent 8,796,779

TABLE OF CONTENTS

Process Integration, Devices, and Structures 1
Scope 1
Logic 1
Memory 1
Reliability 1
Difficult Challenges 2
Description of Process Integration, Devices, and Structures Difficult Challenges 3
Logic Technology Requirements and Potential Solutions 5
Logic Technology Requirements 5
Logic Potential Solutions 27
Memory Technology Requirements and Potential Solutions 29
DRAM 29
Non-volatile Memory 35
Reliability Technology Requirements and Potential Solutions 51
Introduction 51
Top Reliability Challenges 52
Reliability Requirements 53
Reliability Potential Solutions 55
Cross TWG Issues 57
Modeling and Simulation 57
Inter-focus ITWG Discussion 57
Emerging Research Devices 57
Front End Processes 57
References 58

LIST OF FIGURES

Figure PIDS1 High-Performance Logic: Scaling of Transistor Intrinsic Speed, 1/τ 7
Figure PIDS2 High-Performance Logic: Jg,limit versus Simulated Gate Leakage Current Density for SiON Gate Dielectric 9
Figure PIDS3 LSTP: Jg,limit versus Simulated Gate Leakage Current Density for SiON Gate Dielectric 9
Figure PIDS4 LOP: Jg,limit versus Simulated Gate Leakage Current Density for SiON Gate Dielectric 10
Figure PIDS5 Logic Potential Solutions 29
Figure PIDS6 Cell FET Devices 34
Figure PIDS7 Storage Node Capacitor 34
Figure PIDS8 DRAM Potential Solutions 35
Figure PIDS9 Non-volatile Memory Potential Solutions 51
Figure PIDS10 Reliability Potential Solutions 56

LIST OF TABLES

Table PIDS1a Process Integration Difficult Challenges—Near-term Years 2
Table PIDS1b Process Integration Difficult Challenges—Long-term Years 3
Table PIDS2a High-performance Logic Technology Requirements—Near-term Years 11
Table PIDS2b High-performance Logic Technology Requirements—Long-term Years 13
Table PIDS3a Low Standby Power Technology Requirements—Near-term Years 17
Table PIDS3b Low Standby Power Technology Requirements—Long-term Years 19
Table PIDS3c Low Operating Power Technology Requirements—Near-term Years 21
Table PIDS3d Low Operating Power Technology Requirements—Long-term Years 23
Table PIDS4a DRAM Technology Requirements—Near-term Years 31
Table PIDS4b DRAM Technology Requirements—Long-term Years 32
Table PIDS5a Non-volatile Memory Technology Requirements—Near-term Years 36
Table PIDS5b Non-volatile Memory Technology Requirements—Long-term Years 40
Table PIDS6 Reliability Difficult Challenges 53
Table PIDS7a Reliability Technology Requirements—Near-term Years 54
Table PIDS7b Reliability Technology Requirements—Long-term Years 55

MASTAR Modeling Program Link

PROCESS INTEGRATION, DEVICES, AND STRUCTURES

SCOPE

The *Process Integration, Devices, and Structures (PIDS)* chapter deals with overall IC process flow integration, with the main IC devices and structures, and with the reliability tradeoffs associated with new options. Physical and electrical requirements and characteristics are emphasized within PIDS, encompassing parameters such as physical dimensions, key device electrical parameters, including device electrical performance and leakage, and reliability criteria. The focus is on nominal targets, although statistical tolerances are discussed as well. Key technical challenges facing the industry in this area are addressed, and some of the best-known potential solutions to these challenges are discussed. The chapter is subdivided into the following major subsections: logic, memory (including both DRAM and non-volatile memory [NVM]), and reliability.

Main aims of the ITRS include identifying key technical requirements and challenges critical to sustaining the historical scaling of CMOS technology per Moore's Law and stimulating the needed research and development to meet the key challenges. The objective of listing and discussing potential solutions in this chapter is to provide the best current guidance about approaches that address the key technical challenges. However, the potential solutions list here is not comprehensive, nor are the solutions in the list necessarily the most optimal ones. Given these limitations, the potential solutions in the ITRS are meant to stimulate and not limit research exploring new and different approaches.

LOGIC

A major portion of semiconductor device production is devoted to digital logic. In this section, both high-performance and low-power logic (which is typically for mobile applications) are included, and detailed technology requirements and potential solutions are considered for both types. Key considerations are performance, power, and density requirements and goals. One key theme is continued scaling of the MOSFETs for leading-edge logic technology in order to maintain historical trends of improved device performance. This scaling is driving the industry toward a number of major technological innovations, including material and process changes such as high-κ gate dielectric, metal gate electrodes, etc., and in the long term, new structures such as ultra-thin body, multiple-gate MOSFETs (such as FinFETs). These innovations are expected to be introduced at a rapid pace, and hence understanding, modeling, and implementing them into manufacturing in a timely manner is expected to be a major issue for the industry

MEMORY

Logic and memory together form the predominant majority of semiconductor device production. The types of memory considered in this chapter are DRAM and non-volatile memory (NVM). The emphasis is on commodity, stand-alone chips, since those chips tend to drive the memory technology. However, embedded memory chips are expected to follow the same trends as the commodity memory chips, usually with some time lag. For both DRAM and NVM, detailed technology requirements and potential solutions are considered

The NVM discussion in this chapter is limited to devices that can be written and read many times; hence read-only memory (ROM) and one-time-programmable (OTP) memory are excluded. The current mainstream types of NVM currently is Flash, both NAND and NOR. There are serious issues with scaling that are dealt with at some length in the chapter. Other, non-charge-storage types of NVM are also considered, include ferroelectric RAM (FeRAM), magnetic RAM (MRAM), and phase change RAM. For DRAM type memory, the key issue is dealing with increasing scaling difficulties, especially with ensuring very low levels of leakage.

RELIABILITY

Reliability is a critical aspect of process integration. Emerging technology generations require the introduction of new materials and processes at a rate that exceeds current capabilities for gathering information and generating the required database and models on new failure regimes and defects. Because process integration must then be performed without the benefit of extended learning, it will be difficult to maintain current reliability levels. Uncertainties in reliability can also lead to unnecessary performance, cost, and time-to-market penalties. These issues place difficult challenges on testing and wafer level reliability (WLR). Packaging interface reliability is particularly vulnerable to reliability problems because of new materials and processes, form factors, tighter lead and bond spacing, severe environments, adhesion, and customer manufacturing capability issues.

DIFFICULT CHALLENGES

Table PIDS1a Process Integration Difficult Challenges—Near-term Years

Difficult Challenges ≥ 22 nm	*Summary of Issues*
1. Scaling of MOSFETs to the 22 nm technology generation	Scaling planar bulk CMOS will face significant challenges due to the high channel doping required, band-to-band tunneling across the junction and gate-induced drain leakage (GIDL), random doping variations, and difficulty in adequately controlling short channel effects. Also, keeping parasitics, such as series source/drain resistance with very shallow extensions and fringing capacitance, within tolerable limits will be significant issues. Implementation into manufacturing of new structures such as ultra-thin body fully depleted silicon-on-insulator (SOI) and multiple-gate (e.g., FinFET) MOSFETs is expected at some point. This implementation will be challenging, with numerous new and difficult issues. A particularly challenging issue is the control of the thickness and its variability for these ultra-thin MOSFETs, as well as control of parasitic series source/drain resistance for very thin regions.
2. With scaling, difficulties in inducing adequate strain for enhanced mobility.	With scaling, it is critically important to maintain (or even increase) the current significantly enhanced CMOS channel mobility attained by applying strain to the channel. However, the strain due to current process-induced strain techniques tends to decrease with scaling.
3. Timely assurance for the reliability of multiple and rapid material, process, and structural changes	Multiple major changes are projected over the next seven years, such as.: Material: high-κ gate dielectric, metal gate electrodes, lead-free solder Process: elevated S/D (selective epi) and advanced annealing and doping techniques Structure: ultra-thin body (UTB) fully depleted (FD) SOI, multiple-gate MOSFETs, multi-chip package modules It will be an important challenge to ensure the reliability of all these new materials, processes, and structures in a timely manner.
4. Scaling of DRAM and SRAM to the 22 nm technology generation	DRAM main issues with scaling—adequate storage capacitance for devices with reduced feature size, including difficulties in implementing high-κ storage dielectrics; access device design; holding the overall leakage to acceptably low levels; and deploying low sheet resistance materials for bit and word lines to ensure desired speed for scaled DRAMs. SRAM—Difficulties with maintaining adequate noise margin and controlling key instabilities and soft error rate with scaling. Also, difficult lithography and etch issues with scaling.
5. Scaling high-density non-volatile memory to the 22 nm technology generation	Flash—Non-scalability of tunnel dielectric and interpoly dielectric. Dielectric material properties and dimensional control are key issues. FeRAM—Continued scaling of stack capacitor is quite challenging. Eventually, continued scaling in 1T1C configuration. Sensitivity to IC processing temperatures and conditions. MRAM—Magnetic material properties and dimensional control. Sensitivity to IC processing temperatures and conditions

Table PIDS1b Process Integration Difficult Challenges—Long-term Years

Difficult Challenges<22 nm	*Summary of Issues*
6. Implementation of advanced, non-classical CMOS with enhanced drive current and acceptable control of short channel effects for highly scaled MOSFETs	Advanced non-classical CMOS (e.g., multiple-gate MOSFETs) with ultra-thin, lightly doped body will be needed to scale MOSFETs to 10 nm gate length and below effectively. Control of parasitic resistance and capacitance will be critical. To attain adequate drive current for the highly scaled MOSFETs, quasi-ballistic operation with enhanced thermal velocity and injection at the source end appears to be needed. Eventually, nanowires, carbon nanotubes, or other high transport channel materials (e.g., germanium or III-V thin channels on silicon) may be needed.
7. Dealing with fluctuations and statistical process variations in sub-11 nm gate length MOSFETs	Fundamental issues of statistical fluctuations for sub-10 nm gate length MOSFETs are not completely understood, including the impact of quantum effects, line edge roughness, and width variation.
8. Identifying, selecting, and implementing new memory structures	Dense, fast, low operating voltage non-volatile memory will become highly desirable Increasing difficulty is expected in scaling DRAMs, especially scaling down the dielectric equivalent oxide thickness and attaining the very low leakage currents and power dissipation that will be required. All of the existing forms of nonvolatile memory face limitations based on material properties. Success will hinge on finding and developing alternative materials and/or development of alternative emerging technologies. See Emerging Research Devices section for more detail.
9. Identifying, selecting, and implementing novel interconnect schemes	Eventually, it is projected that the performance of copper/low-κ interconnect will become inadequate to meet the speed and power dissipation goals of highly scaled ICs. Solutions (optical, microwave/RF, etc.) are currently unclear. For detail, refer to ITRS Interconnect chapter.
10. Eventually, identification, selection, and implementation of advanced, non-CMOS devices and architectures for advanced information processing	Will drive major changes in process, materials, device physics, design, etc. Performance, power dissipation, etc., of non-CMOS devices need to extend well beyond CMOS limits. Non-CMOS devices need to integrate physically or functionally into a CMOS platform. Such integration may be difficult. See Emerging Research Devices sections for more discussion and detail.

DESCRIPTION OF PROCESS INTEGRATION, DEVICES, AND STRUCTURES DIFFICULT CHALLENGES

[1] Scaling of MOSFETs to the 22 nm technology generation—With scaling of planar bulk MOSFETs, the channel doping will need to be increased to undesirably high levels in order to gain adequate control of short-channel effects and to set the threshold voltage properly. As a result of the high channel doping, the mobility of holes and electrons will be reduced and the junction leakage due to band-to-band tunneling and gate-induced drain leakage will increase. Furthermore, due to the small total number of dopants in the channel of extremely small MOSFETs, the percent stochastic (random) variation in the number and location of the dopants will increase sharply, and this will sharply increase the statistical variability of the threshold voltage. Another challenge for highly scaled MOSFETs is reducing the parasitic series source/drain resistance (R_{sd}) to tolerable values with very shallow source and drain junction depth.

Due to the challenges with scaling planar bulk MOSFETs, advanced devices such as ultra-thin body fully depleted SOI MOSFETs and multiple-gate, particularly double-gate (DG) MOSFETs (e.g., FinFETs) are expected to be eventually implemented. Since such devices will typically have lightly doped channels and the threshold voltage will be controlled by the metal gate electrode's work function, the challenges associated with high channel doping and stochastic dopant variation in planar bulk MOSFETs will be avoided, but numerous new challenges are expected. Amongst the most critical of such challenges will be controlling the body thickness and its variability for these ultra-thin structures, and setting the metal gate electrode work function to its desired value. As with the planar bulk MOSFET, it will be highly challenging to reduce the parasitic series source/drain resistance (R_{sd}) to tolerable values, but here the ultra-thin body is an added difficulty.

With scaling, a common issue for both planar bulk and advanced MOSFETs is expected to be increased line edge roughness as a percentage of the gate length.

For high-performance logic, in the face of increased chip complexity and increasing transistor leakage current with scaling, chip static power dissipation is expected to become particularly difficult to control while at the same time meeting aggressive targets for performance scaling. Innovations in circuit design and architecture for performance management, as well as utilization of multiple transistors on chip, are needed to design chips with both the desired

performance and power dissipation. The multiple transistors have different threshold voltages (V_t), with the low V_t, high leakage devices used mainly in the critical paths, and higher V_t, lower leakage devices used in the rest of the chip. For low-power logic, control of static power dissipation with scaling is the critical goal. To meet this goal, the transistor leakage current is projected to be much lower than for high-performance logic, and circuit and architectural innovations as well as multiple transistors on the chip will be needed, similarly to high-performance logic.

[2] With scaling, difficulties in inducing adequate strain for enhanced mobility— Currently, enhanced channel mobility due to applied strain to the channel is a major contributor to meeting the MOSFET performance requirements. With scaling, it is critically important to maintain (or even increase) the significantly enhanced CMOS channel mobility to continue to meet the performance requirements. However, the strain due to current process-induced strain techniques tends to decrease with scaling, and solutions to maintain the strain in scaled structures are needed. (For more detail, see Logic Potential Solutions section.)

[3] Timely assurance for the reliability of multiple and rapid material, process, and structural changes—In order to successfully scale ICs to meet performance, leakage current, and other requirements, it is expected that numerous major process and material innovations, such as high-κ gate dielectric, metal gate electrodes, elevated source/drain, advanced annealing and doping techniques, new low-κ materials, lead-free solders, multi-chip packages, etc., will need to be implemented in well under a decade. Also, it is projected that new MOSFET structures, starting with ultra-thin body SOI MOSFETs and moving on to ultra-thin body, multiple-gate MOSFETs, will need to be implemented. Understanding and modeling the reliability issues for all these innovations so that their reliability can be ensured in a timely manner is expected to be particularly difficult.

[4] Scaling of DRAM and SRAM to the 22 nm technology generation—For DRAM, a key issue is implementation of high-κ dielectric materials and eventually MIM structures in order to get adequate storage capacitance per cell even as the cell size is shrinking. Also important is controlling the total leakage current, including the dielectric leakage, the storage junction leakage, and the access transistor source/drain subthreshold leakage, in order to preserve adequate retention time. The requirement of low leakage currents causes problems in obtaining the desired access transistor performance. Finally, deploying of low sheet resistance materials for word and bit lines to ensure acceptable speed for scaled DRAMs is critically important.

For SRAM, difficulties with scaling are expected, particularly in maintaining both acceptable noise margins and controlling instability, especially hot electron instability and negative bias temperature instability (NBTI). Also, there are difficult lithography and etch issues with scaling, and difficult issues with keeping the leakage current within tolerable targets for highly scaled SRAMs. Solving these SRAM challenges is critical to system performance, since SRAM is typically used for fast, on-chip memory.

[5] Scaling high-density non-volatile memory (NVM) to the 22 nm technology generation—Inherent in the nature of available nonvolatile semiconductor memory are two challenges. The first is that the memory element structure for each NVM technology differs from the underlying CMOS technology in some way, and accommodating those differences while attempting to scale the memory cell poses some difficult issues. These issues vary depending on which NVM technology is being considered, and specific issues are listed for each NVM type in the table. The second challenge is that the normal operating process used to set and to reset the state of the memory cell generally stresses the materials, and degradation of cell characteristics can be expected. Degradation is usually associated with a defect related mechanism rather than with an intrinsic device characteristic. Endurance and retention requirements provide the user with guidance as to the probable capability of the device and define a "safe" range of use. For both parameters it is a continuous challenge to be able to realistically determine this long-term behavior. Failure causes are difficult to identify and real-time testing is not feasible.

[6] Implementation of advanced, non-classical CMOS with enhanced drive current and acceptable control of short channel effects for highly scaled MOSFETs—For the long-term years, when the transistor gate length is projected to become 10 nm and below, ultra-thin body, multiple-gate MOSFETs with lightly doped channels are expected to be utilized to effectively scale the device, and particularly, to control short-channel effects for such highly scaled devices. The other material and process solutions mentioned above, such as high-κ gate dielectric, metal gate electrodes, strained silicon channels, elevated source/drain, etc., are expected to be incorporated along with the non-classical CMOS structures. For 10 nm gate length and below, body thicknesses well below 10 nm are projected, and the impact of quantum and surface scattering effects on such thin devices are not well understood. Finally, for these advanced, highly scaled MOSFETs, quasi-ballistic operation with enhanced thermal carrier velocity and injection at the source end appears to be necessary. Eventually, high transport channel materials, such as germanium or III-V channels on silicon, or carbon nanotubes or nanowires, may be utilized.

[7] Dealing with fluctuations and statistical process variations in sub-11 nm gate length MOSFETs—For such devices, the impact of statistical variations is not well understood, including the impact of quantum effects, line edge roughness, and variation in the ultra-thin body width.

[8] Identifying, selecting, and implementing new memory structures—In the long term, increasing difficulty is expected in scaling both DRAMs and NVMs, as discussed for each of these memory types in the table. The need for high density, fast, and new non-volatile memory structures is expected to increase, particularly to reduce power dissipation. Implementing such advanced, non-volatile structures will be a major challenge.

[9] Identifying, selecting, and implementing novel interconnect schemes—The resistivity of copper increases somewhat with scaling for widths under ~100 nm, and at κ ~1–1.5, the limits of low-κ dielectric will be reached. At that point, further interconnect performance improvements will require novel architectural and/or materials solutions

[10] Eventually, identification, selection, and implementation of advanced, non-CMOS devices and architectures for advanced information processing—Eventually, toward the end of the Roadmap or beyond, scaling of MOSFETs is likely to become ineffective and/or very costly, and advanced non-CMOS solutions will need to be implemented to continue to improve performance, power, density, etc. It is expected that such solutions will be integrated either functionally or physically with a CMOS baseline technology that takes advantage of the high-performance, cost-effective, and very dense CMOS logic that will have been developed and implemented by then.

LOGIC TECHNOLOGY REQUIREMENTS AND POTENTIAL SOLUTIONS

LOGIC TECHNOLOGY REQUIREMENTS

The technology requirements tables reflect the MOSFET transistor requirements of both high-performance and low-power digital ICs. High-performance logic refers to chips of high complexity, high performance, and high power dissipation, such as microprocessor unit (MPU) chips for desktop PCs, servers, etc. Low-power logic refers to chips for mobile systems, where the allowable power dissipation and hence the allowable leakage currents are limited by battery life. There are two major categories within low-power, low operating power (LOP) and low standby power (LSTP) logic. LOP chips are typically for relatively high-performance mobile applications, such as notebook computers, where the battery is likely to be high capacity and the focus is on reduced operating (i.e., dynamic) power dissipation. LSTP chips are typically for lower performance, lower cost consumer type applications, such as consumer cellular telephones, with lower battery capacity and an emphasis on the lowest possible static power dissipation, i.e., the lowest possible leakage current. The transistors for high-performance ICs have both the highest performance and the highest leakage current of all, and hence the physical gate length and all the other transistor dimensions are most rapidly scaled for high-performance logic. The transistors for LOP chips have somewhat lower performance and considerably lower leakage current, while the transistors for LSTP chips have both the lowest performance and the lowest leakage current of all. For LOP logic, the gate length lags behind the high-performance transistor gate length by two years, reflecting historical trends and the need for low leakage current in mobile applications. For LSTP logic, the gate length lags that of high-performance logic by four years, reflecting the ultra-low leakage current required.

For generating the entries in the logic technology requirements tables, the MASTAR MOSFET modeling software was used. T. Skotnicki, F. Boeuf and their collaborators developed MASTAR[1], [2], [3] and it contains detailed analytical MOSFET models that have been verified against literature data. It is well suited to efficiently analyzing technology tradeoffs for generating these tables. An important calculated output parameter is the intrinsic MOSFET delay, $\tau = CV/I$, where C is the total gate capacitance (including parasitic gate overlap and fringing capacitance) per micron transistor width, V is the power supply voltage (V_{dd}), and I is the saturation drive current per micron transistor width ($I_{d,sat}$). τ is a reasonable metric for the intrinsic MOSFET delay, and hence $1/\tau$, the transistor intrinsic switching frequency, is used as the key transistor performance metric. *(It should be noted that another transistor delay metric, CV/I_{eff}, where I_{eff} is a modified drain current derived from a linear superposition of currents,[4] has been developed recently and appears to be somewhat more accurate than the $CV/I_{d,sat}$ metric. We are continuing to use the latter metric because it is sufficiently accurate to follow the key scaling trends, and for consistency with previous Roadmaps.)*

To determine the projected parameter values in a table, a target is set for one of the key outputs, either leakage current (for low-power logic) or $1/\tau$ (for high-performance logic). Then the input parameters are tentatively chosen based on scaling rules, engineering judgment, and physical device principles. Using MASTAR, the input parameters are iteratively varied until the target is met, and the final set of values for the input parameters is entered into the table. The MASTAR program and the specific MASTAR process and roadmap files used to generate the tables are on the ITRS website at *http://www.itrs.net*.

In each of these tables, multiple parallel paths are followed. Planar bulk CMOS is extended as long as possible, while advanced CMOS technologies, ultra-thin body fully depleted (UTB FD) silicon-on-insulator (SOI) MOSFETs and multiple-gate, particularly double-gate (DG) MOSFETs (such as FinFETs), are implemented in 2010 or later and run in parallel with the planar bulk CMOS (for details see the logic tables). With scaling, difficulties arise with planar bulk MOSFETs because of high channel doping, inability to adequately control short channel effects, and others (for more detail see Difficult Challenges section, Item 1). The advanced CMOS technologies can be scaled more effectively, and hence are utilized later in the Roadmap. In fact, multiple-gate MOSFET scaling is superior to UTB FD MOSFET scaling, and hence the ultimate MOSFET is projected to be the multiple-gate device. For the industry as a whole, multiple paths are likely, as different companies choose different timing in extending planar bulk and then switching to the advanced CMOS technologies, depending on their needs, plans, and technological strengths. The multiple parallel paths in this roadmap are meant to reflect this. The specific set of projected parameter values in each of the tables reflects a particular scaling scenario, in which the targeted values for the key output are achieved. However, since there are numerous input parameters that can be varied, and the output parameters are complicated functions of these numerous input parameters, other sets of projected parameter values (i.e., different scaling scenarios) can be found that achieve the targeted values for the key output. For example, if, in one scenario, the equivalent oxide thickness (EOT) were scaled rapidly so that gate leakage current scales upward rapidly, requiring early introduction of high-κ gate dielectric to reduce the gate leakage current to tolerable levels, an alternate scaling scenario would scale the EOT slower. As a result, the gate leakage current would scale upward more slowly, hence delaying the required introduction of high-κ gate dielectric. However, some of the other parameters, such as the gate length, the channel doping, and/or mobility enhancement, would have to be scaled differently to compensate for the slowed scaling of the EOT and to reach the same targeted output values. Hence, the scaling scenarios in these tables constitute a good guide for the industry, and are meant to be representative, but there will be considerable variance in the actual paths that the various companies will take.

For the high-performance logic tables (see Tables PIDS2a and b), the driver is the MOSFET intrinsic performance metric, $1/\tau$. Specifically, the target is an average 17% per year increase in $1/\tau$, which matches the historic rate of improvement in device performance. Meeting this target is an important enabler for the desired rate of improvement in the chip clock speed. All the other parameter values in the table are chosen iteratively to meet this target, as explained above. Several important consequences of meeting this target are clear from the tables. The NMOSFET saturation drive current, $I_{d,sat}$, pretty steadily increases over the course of the Roadmap in order to keep $1/\tau$ increasing at the desired 17% per year rate. The subthreshold source/drain leakage current, $I_{sd,leak}$, is relatively high, at 0.34 μA/μm in 2007, and it generally increases with succeeding years, which has important consequences for the chip power dissipation (to be discussed below). Figure PIDS1 shows the scaling of $1/\tau$ for high-performance logic. Overall, the 17%/year target is met, with several important exceptions. For planar bulk, for 2009 and beyond, the $1/\tau$ curve slopes increasingly down from the 17%/year curve, mainly because of the scaling difficulties discussed in the Difficult Challenges section, Item 1. (The scaling difficulties are also indicated in the MASTAR simulations, where the required channel doping increases sharply with year, to a very high value of ~8e18 cm^{-3} in 2012.) For UTB FD SOI, a similar decrease from the 17% curve happens after 2013, although to a lesser extent, and for DG devices, the curve slopes down from the 17%/year curve after 2019. In all these cases, difficulties with controlling short-channel effects with scaling results in a slowing in the transistor performance increase in order to keep the leakage current within tolerable limits. The LSTP and LOP curves for $1/\tau$ scaling show similar effects, although the overall performance increase is ~13-14%/year for these.

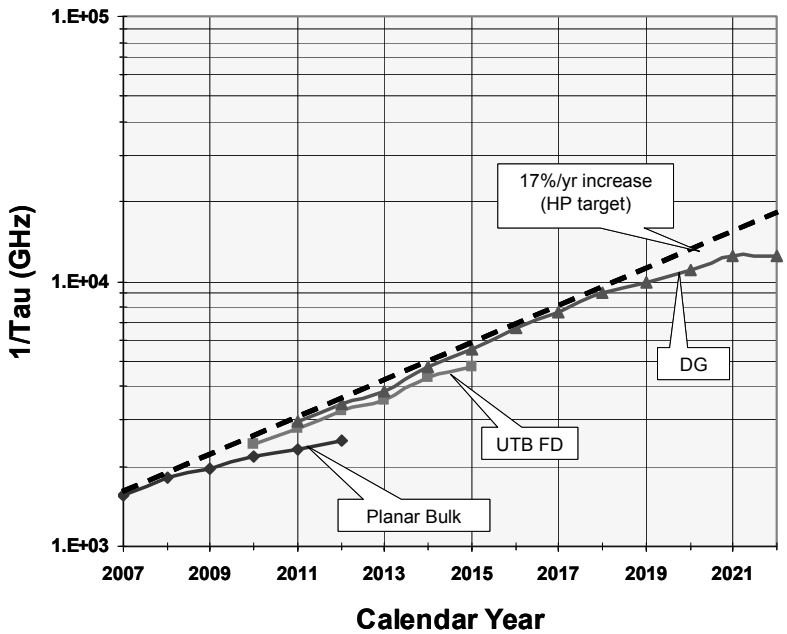

Figure PIDS1 High-Performance Logic: Scaling of Transistor Intrinsic Speed, 1/τ

The IC industry has begun to deploy architectural techniques such as multiple cores and multiple threads that exploit parallelism to improve the overall chip performance, enhance the chip functionality while maintaining chip power density and total chip power dissipation at a manageable level. With more than one central processing unit (CPU) core on chip, the cores can be clocked at a lower frequency while still getting better overall chip performance. Thus, there is a trend for system designers to emphasize integration level, which enables more cores to be put on a chip, instead of raw transistor speed in optimizing system-level performance. In addition, system designers are sweeping ever more cache memory onto the processor chip in order to minimize the system performance penalty associated with finite-cache effects. As DRAM cells are significantly smaller than SRAM cells, another high-performance system technology trend is to integrate DRAM cells onto a processor chip for use in higher-level cache memory. With scaling, it is expected that these techniques will be more and more heavily exploited. In subsequent editions of the Roadmap, the Design and PIDS Working Groups will consider the impact of these architectural techniques, and in particular whether improved architectural parallelism may allow a slackening in the 17%/year transistor performance scaling target.

For high-performance chips, the high subthreshold leakage current must be dealt with to keep chip static power dissipation within tolerable limits. One common approach is to fabricate more than one type of transistor on the chip, including the high-performance, low V_t device described above, as well as other MOSFET(s) with higher V_t and sometimes larger EOT to reduce the leakage current. These alternate, lower leakage devices will have lower saturation drive current and hence poorer device performance (i.e., lower MOSFET intrinsic switching frequency, $1/\tau$) than the high-performance devices. The high-performance device is used just in critical paths, and the low leakage devices are used everywhere else. Extensive use of the low leakage devices can significantly reduce the chip static power dissipation without seriously degrading chip performance. Current circuit/architectural techniques to curtail static power dissipation include pass gates to cut off access to power/ground rails or other techniques to power down circuit blocks. Other potential techniques include well biasing, or using electrically or dynamically adjustable V_t devices. Hence, a realistic picture of scaled high-performance ICs is that the static power dissipation is controlled by utilizing more than one type of transistor and by utilizing device/design/architectural techniques. In the technology requirements table, we have characterized only the high-performance transistor because this transistor is the technology driver.

For low-power chips, the targeted output parameter is the source/drain subthreshold leakage current, $I_{sd,leak}$, and the targets are relatively low, especially for LSTP logic, as discussed above. $I_{sd,leak}$ is 30 pA/µm and is held essentially constant for LSTP, while it is ~9 nA/µm for LOP in 2007, and it increases slowly with scaling. All the other parameter values in the tables are chosen iteratively to meet the $I_{sd,leak}$ targets, as explained above. Nevertheless, the resultant

average improvement in the device performance metric, $1/\tau$, is about 13-14% per year for both LOP and LSTP. Note that, to meet the leakage current requirements, the gate length scaling of low-power logic lags behind that of high-performance logic (see the logic tables for details). One key issue for LSTP logic is the slow scaling of V_{dd}. Refer to Tables PIDS3a and b for LSTP data. This slow scaling is a result of the relatively slow scaling of the threshold voltage, V_t, required to meet the very low subthreshold leakage current targets. V_{dd} must follow V_t in scaling slowly because, to obtain reasonable device performance, the overdrive, $(V_{dd}-V_t)$, must remain relatively large. Since dynamic power dissipation is proportional to $(V_{dd})^2$, the dynamic power dissipation for the LSTP logic scales relatively slowly, but since the activity factor for this type of logic is expected to be relatively small, the lowered static power dissipation because of the very low leakage currents more than compensates. In contrast to LSTP logic, V_{dd} scales relatively quickly for LOP logic (see technology requirements tables for LOP, Tables PIDS3c and d), where, as mentioned above, the focus is on minimizing the operating power (i.e., the dynamic power dissipation, which is proportional to V_{dd}^2). However, since $I_{sd,leak}$ is larger than for LSTP logic, the saturation threshold voltage is low enough that the overdrive, $(V_{dd}-V_t)$, is reasonable.

For low-power chips, the key goal is low power dissipation in order to enhance battery life, with a tradeoff of low performance compared to high-performance chips. This overall goal is attained through the use of transistors with low $I_{sd,leak}$ as well as through the approaches utilized for high-performance logic: multiple transistors on chip and application of circuit and architectural techniques, including power management techniques to reduce chip leakage current in the standby mode. Eventually, effective dynamic threshold voltage adjust techniques may be feasible. The nominal targets for $I_{sd,leak}$ chosen in these LSTP logic tables are quite low, and reflect a transistor design emphasizing low leakage current in the active mode. In contrast, some companies will utilize transistors with significantly higher $I_{sd,leak}$ to get higher performance, and will thus rely more heavily on circuit and architectural techniques to lower overall chip power dissipation. Finally, for LOP logic, as discussed above, V_{dd} will be scaled relatively quickly to keep the dynamic power dissipation within tolerable limits.

A critical issue is the gate leakage current, and whether the current standard silicon oxy-nitride gate dielectric can meet the gate leakage current density limit as the oxy-nitride becomes increasingly thin with scaling (Refer to Tables PIDS2a and b, and PIDS3a through d and to table notes [2] and [5]). This is an important issue, since, in the EOT regime in the Roadmap, gate leakage current is due to direct tunneling and hence the gate leakage current increases approximately exponentially with decreasing EOT. The FEP TWG and North Carolina State University performed detailed simulations of direct tunneling leakage current density through oxides, and these simulations were used to calculate the expected value of the gate leakage current density due to tunneling through oxy-nitride, using as inputs the scaled V_{dd} and EOT per the technology requirements tables. For LSTP, LOP, and high-performance logic, these calculations of the expected gate leakage current density were compared to the gate leakage current density limit ($J_{g,limit}$) from the tables. The results are shown in Figures PIDS2 through PIDS4, where "simulated J_g" is the expected value of the gate leakage current density from the simulations. For the LSTP and high-performance logic transistors, the two Jg curves cross shortly before or at 2008, and hence, for 2008 and beyond, the leakage current limit cannot be met using silicon oxy-nitride because of direct tunneling. Furthermore, for both curves the $J_{g,simulated}$ curve separates rapidly from the $J_{g,limit}$ curve after 2008, indicating that gate leakage would rapidly become completely out of specification if oxy-nitride were to continue to be used for the gate dielectric after 2008. Hence, high-κ gate dielectric (which significantly reduces gate leakage current density for a given EOT) is clearly needed for LSTP and high-performance logic by 2008; this is the leading potential solution for high gate leakage. For LOP logic, the point where the leakage current limit cannot be met using oxy-nitride is in 2009, but high κ is assumed to be implemented for LOP in 2008 as well as for the others. Note that the J_g plots in all three figures are just for planar bulk MOSFETs; the plots for UTB FD and dual gate (DG) MOSFETs have not been included in order to avoid cluttering the figures and because the implications for when high-κ gate dielectrics are needed would be unchanged if those plots were included.

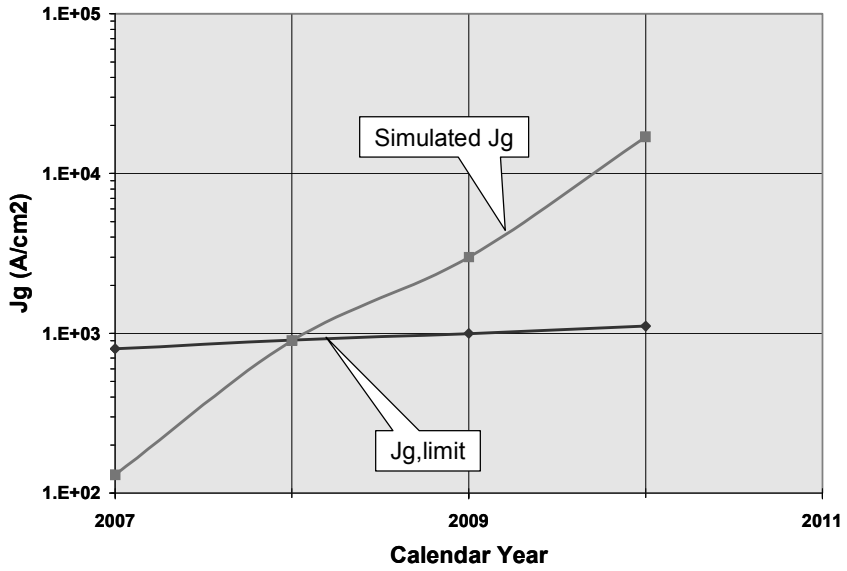

Figure PIDS2 High-Performance Logic: $J_{g,limit}$ versus Simulated Gate Leakage Current Density for SiON Gate Dielectric

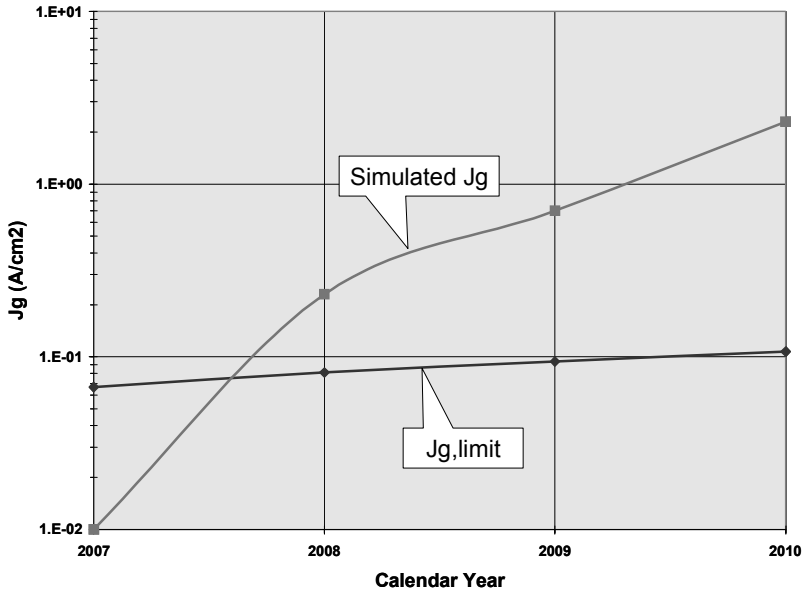

Figure PIDS3 LSTP: $J_{g,limit}$ versus Simulated Gate Leakage Current Density for SiON Gate Dielectric

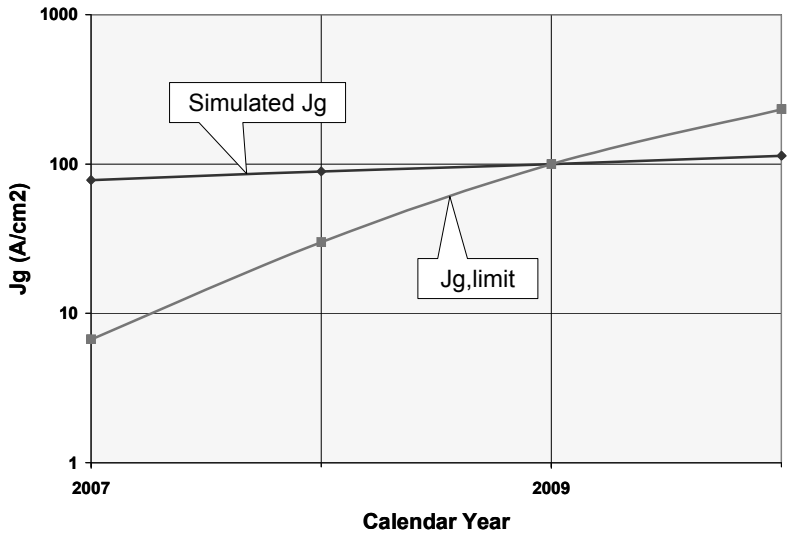

Figure PIDS4 LOP: $J_{g,limit}$ versus Simulated Gate Leakage Current Density for SiON Gate Dielectric

Table PIDS2a High-performance Logic Technology Requirements—Near-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

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*	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
L_g: Physical Lgate for High Performance logic (nm) [1]	*25*	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
EOT: Equivalent Oxide Thickness [2]									
Extended planar bulk (Å)	11	9	7.5	6.5	5.5	5			
UTB FD (Å)				7	6	5.5	5	5	5
DG (Å)					8	7	6	6	6
Gate Poly Depletion and Inversion-Layer Equivalent Thickness [3]									
Extended Planar Bulk (Å)	7.4	3.1	2.9	2.8	2.7	2.6			
UTB FD (Å)				4	4	4	4	4	4
DG (Å)					4	4	4	4	4
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]									
Extended Planar Bulk (Å)	18.4	12.1	10.4	9.3	8.2	7.6			
UTB FD (Å)				11	10	9.5	9	9	9
DG (Å)					12	11	10	10	10
$J_{g,limit}$: Maximum gate leakage current density [5]									
Extended Planar Bulk (A/cm^2)	8.00E+02	9.09E+02	1.00E+03	1.11E+03	1.25E+03	1.43E+03			
UTB FD (A/cm^2)				1.11E+03	1.25E+03	1.43E+03	1.54E+03	1.82E+03	2.00E+03
DG (A/cm^2)					1.25E+03	1.43E+03	1.54E+03	1.82E+03	2.00E+03
V_{dd}: Power Supply Voltage (V) [6]									
Extended Planar Bulk (V)	1.1	1	1	1	0.95	0.9			
UTB FD and DG (V)				1	1	0.9	0.9	0.9	0.8
$V_{t,sat}$: Saturation Threshold Voltage [7]									
Extended Planar Bulk (mV)	134	94	94	103	101	112			
UTB FD (mV)				103	89	87	93	99	99
DG (mV)					115	105	103	108	111
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]									
Extended Planar Bulk (µA/µm)	0.34	0.71	0.70	0.64	0.74	0.68			
UTB FD (µA/µm)				0.33	0.52	0.62	0.56	0.55	0.60
DG (µA/µm)					0.2	0.34	0.37	0.38	0.38
$I_{d,sat}$: NMOS Drive Current [9]									
Extended Planar Bulk (µA/µm)	1211	1513	1639	1807	1824	1762			
UTB FD (µA/µm)				1948	2000	1944	2109	2245	2030
DG (µA/µm)					1917	1943	2204	2365	2295
Mobility enhancement factor due to strain [10]	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
$I_{d,sat}$ enhancement factor due to strain [11]									
Extended Planar Bulk	1.09	1.08	1.08	1.08	1.09	1.08			
UTB FD				1.07	1.06	1.06	1.06	1.05	1.05
DG					1.04	1.04	1.04	1.03	1.03

Table PIDS2a High-performance Logic Technology Requirements—Near-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

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*	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
Effective Ballistic Enhancement Factor , Kbal [12]									
Extended Planar Bulk	1	1	1	1	1	1			
UTB FD				1.05	1.1	1.16	1.2	1.24	1.28
DG					1.17	1.25	1.31	1.37	1.53
R_{sd}: Effective Parasitic series source/drain resistance [13]									
Extended Planar Bulk (Ω-µm)	200	200	200	180	180	180			
UTB FD (Ω-µm)		-	-	180	180	180	170	160	160
DG (Ω-µm)					180	180	170	160	160
$C_{g,ideal}$: Ideal NMOS Device Gate Capacitance [14]									
Extended Planar Bulk (F/µm)	4. 70E-16	6.30E-16	6.63E-16	6.70E-16	6.71E-16	6.33E-16			
UTB FD (F/µm)				5.65E-16	5.52E-16	5.08E-16	4.98E-16	4.22E-16	3.83E-16
DG (F/µm)					4.60E-16	4.39E-16	4.48E-16	3.80E-16	3.45E-16
$C_{g,total}$: Total gate capacitance for calculation of CV/I [15]									
Extended Planar Bulk (F/µm)	7.10E-16	8.40E-16	8.43E-16	8.40E-16	8.35E-16	7.93E-16			
UTB FD (F/µm)				8.08E-16	7.22E-16	6.78E-16	6.58E-16	5.82E-16	5.43E-16
DG (F/µm)					6.50E-16	6.29E-16	6.28E-16	5.59E-16	5.25E-16
τ =CV/I: NMOSFET intrinsic delay (ps) [16]									
Extended Planar Bulk (ps)	0.64	0.55	0.51	0.46	0.43	0.4			
UTB FD (ps)				0.41	0.36	0.31	0.28	0.23	0.21
DG (ps)					0.34	0.29	0.26	0.21	0.18
1/τ: NMOSFET intrinsic switching speed (GHz) [17]									
Extended Planar Bulk (GHz)	1563	1818	1961	2174	2326	2500			
UTB FD (GHz)				2439	2778	3226	3571	4348	4762
DG (GHz)					2941	3448	3846	4762	5556

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

Table PIDS2b High-performance Logic Technology Requirements—Long-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*
MPU Physical Gate Length (nm)	*9*	*8*	*7*	*6.3*	*5.6*	*5.0*	*4.5*
L_g: Physical Lgate for High Performance logic (nm) [1]	*9*	*8*	*7*	*6*	*5.5*	*5*	*4.5*
EOT: Equivalent Oxide Thickness [2]							
Extended planar bulk (Å)							
UTB FD (Å)							
DG (Å)	**5.5**	**5.5**	**5.5**	**5**	**5**	**5**	**5**
Gate Poly Depletion and Inversion-Layer Equivalent Thickness [3]							
Extended Planar Bulk (Å)							
UTB FD (Å)							
DG (Å)	**4**	**4**	**4**	**4**	**4**	**4**	**4**
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]							
Extended Planar Bulk (Å)							
UTB FD (Å)							
DG (Å)	**9.5**	**9.5**	**9.5**	**9**	**9**	**9**	**9**
$J_{g,limit}$: Maximum gate leakage current density [5]							
Extended Planar Bulk (A/cm^2)							
UTB FD (A/cm^2)							
DG (A/cm^2)	**2.22E+03**	**2.50E+03**	**2.86E+03**	**3.33E+03**	**3.64E+03**	**4.00E+03**	**4.44E+03**
V_{dd}: Power Supply Voltage (V) [6]							
Extended Planar Bulk (V)							
UTB FD and DG (V)	**0.8**	**0.7**	**0.7**	**0.7**	**0.65**	**0.65**	**0.65**
$V_{t,sat}$: Saturation Threshold Voltage [7]							
Extended Planar Bulk (mV)							
UTB FD (mV)							
DG (mV)	**110**	**109**	**114**	**119**	**123**	**115**	**118**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)							
DG (µA/µm)	**0.44**	**0.48**	**0.45**	**0.47**	**0.43**	**0.62**	**0.60**
$I_{d,sat}$: NMOS Drive Current [9]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)							
DG (µA/µm)	**2627**	**2533**	**2804**	**2768**	**2677**	**2799**	**2786**
Mobility enhancement factor due to strain [10]	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**
$I_{d,sat}$ enhancement factor due to strain [11]							
Extended Planar Bulk							
UTB FD							
DG	**1.03**	**1.03**	**1.02**	**1.02**	**1.02**	**1.02**	**1.02**

Table PIDS2b High-performance Logic Technology Requirements—Long-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*
MPU Physical Gate Length (nm)	*9*	*8*	*7*	*6.3*	*5.6*	*5.0*	*4.5*
Effective Ballistic Enhancement Factor , Kbal [12]							
Extended Planar Bulk							
UTB FD							
DG	1.67	1.87	1.99	1.97	2.11	2.11	2.11
R_{sd}: Effective Parasitic series source/drain resistance [13]							
Extended Planar Bulk (Ω-µm)							
UTB FD (Ω-µm)							
DG (Ω-µm)	155	150	145	145	145	135	135
$C_{g,ideal}$: Ideal NMOS Device Gate Capacitance [14]							
Extended Planar Bulk (F/µm)							
UTB FD (F/µm)							
DG (F/µm)	3.27E-16	2.91E-16	2.68E-16	2.30E-16	2.11E-16	1.92E-16	1.72E-16
$C_{g,total}$: Total gate capacitance for calculation of CV/I [15]							
Extended Planar Bulk (F/µm)							
UTB FD (F/µm)							
DG (F/µm)	5.07E-16	4.81E-16	4.58E-16	4.10E-16	3.91E-16	3.62E-16	3.42E-16
τ=CV/I: NMOSFET intrinsic delay (ps) [16]							
Extended Planar Bulk (ps)							
UTB FD (ps)							
DG (ps)	0.15	0.13	0.11	0.1	0.09	0.08	0.08
1/τ: NMOSFET intrinsic switching speed (GHz) [17]							
Extended Planar Bulk (GHz)							
UTB FD (GHz)							
DG (GHz)	6667	7692	9091	1.00E+04	1.11E+04	1.25E+04	1.25E+04

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

Notes for Tables PIDS2a and b:

As described in the text, MASTAR[1,2,3], a detailed analytical MOSFET modeling software package, has been utilized to generate the parameter values in these tables. The MASTAR modeling package and user's manual are in the backup material on the ITRS website, as well as the detailed MASTAR simulations that underlay these tables. Also note that the parameters in this table are for an NMOSFET with nominal gate length at an operating temperature of 25°C. Furthermore, although there are multiple MOSFETs in a typical logic chip, with differing threshold voltages, I_{on}, I_{off}, and oxide thickness, the transistor specified here is the transistor with the lowest threshold voltage, highest I_{on} and highest I_{off}, lowest oxide thickness, and fastest intrinsic transistor switching speed. This transistor typically constitutes a small minority of the transistors on a chip; it is used mainly in critical paths, and most of the transistors on the chip have higher threshold voltage and lower leakage current. This high speed, high leakage transistor is specified in this table because it tends to drive the technology.

As explained in the text, multiple parallel options for the transistor type are included in the tables, including planar bulk CMOS extended to its practical scaling limits, ultra-thin body fully-depleted (UTB FD) SOI CMOS, also extended to its practical scaling limits, and double-gate (DG) CMOS (e.g., FinFETs). Note that the limit for planar bulk CMOS is through 2012, and for UTB FD it is through 2015, while DG continues through 2022. The challenges in scaling planar bulk are significant, resulting in an inability to attain the desired 17%/ year improvement in the transistor intrinsic switching speed beyond 2009 (see the text for further discussion). The scalability of UTB FD is superior to that of planar bulk. As a result, $I_{sd,leak}$ is higher for planar bulk than for UTB FD. Furthermore, EOT is scaled more rapidly from 2010 through 2012 for planar bulk than for UTB FD. Finally, from the MASTAR modeling results, the short channel effects such as drain induced barrier lowering (DIBL) are always larger for planar bulk than for UTB FD. In a similar vein, the scalability of DG is superior to that of UTB FD. Hence, $I_{sd,leak}$ is lower for DG than for UTB FD, while EOT is scaled more slowly for DG than for UTB FD. Furthermore, from the MASTAR modeling results, short channel effects are always lower for DG. Hence, DG is the ultimate MOSFET device, continuing through the end of the Roadmap in 2022.

For each transistor option, the scaling of the numbers in the tables reflects a particular scaling scenario in which we have attempted to optimally scale to meet the key goal for high-performance logic, 17% per year average improvement in the NMOS intrinsic switching speed, while keeping the leakage currents, the short channel effects, and other key characteristics under control. However, there are numerous parameters (such as EOT, V_{dd}, $I_{sd,leak}$, etc.) that can be varied, and different scaling scenarios are possible by making different choices on the scaling of these parameters. The scenarios in this table were selected to be as representative of the industry as possible. In particular, in this table, high-κ gate dielectric and metal gate electrode are assumed to be available in 2008. See the figures and discussion in the text for why high-κ gate dielectric is required in 2008. With the EOT=0.9 nm in 2008, metal gate electrode is needed to reduce the polysilicon depletion.

[1] L_g is the physical gate length: the final, as-etched length of the bottom of the gate electrode. Values have been set by the ORTC. The gate dimensional control requirement is set by the Lithography and FEP Etch ITWGs, and is assumed to have a three-sigma value of $\pm 12\% \times L_g$. It is expected that meeting this requirement will become increasingly difficult with scaling (refer to the Lithography chapter and the FEP chapter). Gate length variation is assumed to be a primary factor responsible for driving device parameter variation. Recent data indicate that the Lg scaling in the ITRS may be overly aggressive compared to the reality in the industry, and this will be re-examined in the 2008 ITRS.

[2] EOT: For a gate dielectric of thickness T_d and relative dielectric constant κ, EOT is defined by: EOT = T_d / (κ/3.9), where 3.9 is the relative dielectric constant of thermal silicon dioxide. For a MOSFET with the gate dielectric of thickness T_d, the ideal gate capacitance per unit area is the same as that of a similar MOSFET, but with a gate dielectric made up of thermal silicon dioxide with a thickness of EOT. It is projected that high-κ gate dielectric will be required by 2008 to control the gate leakage (see the text for further discussion on this point.) The rate of scaling of EOT has been slow from 2005 through 2007 to keep the gate leakage current within the specified limits while utilizing silicon oxy-nitride for the gate dielectric. However, there is a sharp EOT decrease in 2008, when we assume that high-κ gate dielectric will be implemented. The yellow coloring reflects the uncertainty in our high-κ solutions for EOT < 0.9nm, while the red coloring reflects the increased uncertainty of solutions for EOT < 0.7nm. Measurement of EOT is complicated, and is usually done via sophisticated MOS capacitor-voltage (CV) measurements on MOS capacitors or via optical measurements.

[3] Gate Poly Depletion and Inversion-Layer Equivalent Thickness: accounts for gate electrode depletion and inversion-layer effects, including quantum effects, both of which are calculated by MASTAR. For polysilicon gate electrodes, the portion of the electrical thickness adjustment due to gate electrode depletion is dependent on the polysilicon doping. For 2008 and beyond, there is a projected inability to adequately dope polysilicon gate electrodes to meet the gate depletion thickness adjustment requirements, and hence it is assumed that metal-gate electrodes, which reduce the gate depletion effect to zero, will be introduced. The abrupt reduction in this parameter in 2008 reflects the zero depletion. For 2008 and beyond, the difference between the parameter value for planar bulk versus the 4 nm value for DG and UTB FD reflects the light channel doping in the latter types of MOSFET and the heavy channel doping in planar bulk. For planar bulk CMOS, the metal gate work function needs to be near the silicon conduction band edge for NMOS and near the silicon valence band edge for PMOS to properly set the MOSFET threshold voltage, as with polysilicon gates. For UTB FD and DG MOSFETs, the channel is very thin and lightly doped, and the work function of the metal gates needs to be within a few hundred millivolts of the silicon midgap (i.e., "near silicon midgap" work function) to properly set the MOSFET's threshold voltage. Multiple V_t's may be obtained by tuning the work function. The yellow and red colors generally follow that of EOT (see Note [2] above), reflecting the tight coupling between the metal gate electrode and the high-κ gate dielectric. For UTB FD and DG, the color is red, reflecting the uncertainties of setting and tuning the effective work function.

[4] EOT_{elec} is the sum of EOT and Gate Poly Depletion and Inversion-Layer Thickness (see Notes [2] and [3] above). For MOSFETs in inversion, ideal gate capacitance per unit area (see Note [14]) is εox / (EOT_{elec}), where εox is the dielectric constant of thermal silicon dioxide. The equivalent electrical oxide thickness in inversion is used in calculations of the transistor intrinsic delay, CV/I (see Note [16]). Red/yellow coloring follows that of EOT and Gate Poly Depletion and Inversion-Layer Thickness.

[5] $J_{g,limit}$ is the maximum allowed gate leakage current density at 25°C, and it is measured with the gate biased to V_{dd} and the source, drain, and substrate all tied to ground. $J_{g,limit}$ is related to $I_{sd,leak,TGT}$, the nominal subthreshold leakage current target at 25°C . $I_{sd,leak,TGT}$ is set to 0.2 μA/micron device width. Specifically, $J_{g,limit}$= [Initial Factor] x [$I_{sd,leak,TGT}$ / L_g] × [Hi T Factor] / [Circuit Factor]. Hi T Factor is set to 10, and it accounts for the high operating temperature (100°C) expected for high-performance logic, by adjusting for both the rapid increase in $I_{sd,leak}$ with temperature and the insensitivity of gate leakage current (since it is due to direct tunneling) to temperature. Circuit Factor is set to 1, and it accounts for the differences between the subthreshold leakage current and the gate leakage current in logic gates compared to single isolated transistors as specified by the $J_{g,limit}$ and $I_{sd,leak}$ parameters in this table. (The reason for these differences is the different bias conditions on the various transistors in logic gates compared to the bias conditions used to define $I_{sd,leak}$ (see Note [8]) and $J_{g,limit}$ for the NMOS transistor in this table). The Initial Factor is set to 0.1, and accounts for the fact that the transistor specified in this table is the low threshold voltage transistor with high subthreshold leakage, but that the predominant transistors in typical circuits have significantly lower subthreshold leakage current. The values of Hi T Factor, Circuit Factor, and Initial Factor used here are rough estimates. The yellow and red coloring follows that of EOT (see Note [2] above).

[6] V_{dd} is the nominal power supply voltage. It has been chosen to maintain sufficient voltage over-drive [V_{dd} – saturation threshold voltage (see Note 7)] in order to meet the required saturation current drive values while still maintaining reasonable vertical gate dielectric electric field strengths. Target power supply voltage values for actual ICs may vary ± 10% (or more) from the values in this table, depending on the particular circuit design application or technology optimization.

[7] $V_{t,sat}$ is the saturation threshold voltage for a nominal gate length transistor with drain bias set equal to V_{dd}, as calculated by MASTAR. The threshold voltage values and the corresponding subthreshold leakage current values (see Note [8]) have been chosen to maintain sufficient voltage over-drive (V_{dd} – saturation threshold voltage) in order to meet the required saturation current drive values (see Note [9]). For planar bulk, the cross-hatched interim solution in 2007 indicates that because of the SiON gate dielectric, the EOT is large (1.1nm) and hence the short-channel effects are quite large (for example, DIBL is over 340 mV/V per MASTAR). The yellow color is associated with high substrate doping exceeding 4E18 cm^{-3} (from MASTAR), while the red color is associated with very high substrate doping exceeding 6E18 cm^{-3}. For UTB FD devices, the color is red in 2010 because of the challenges of controlling the very thin silicon body thickness (T_{si}) required to control $V_{t,sat}$ and short channel effects. For DG devices, the color is red right from the beginning because there are very many issues that are not understood now; in particular, defining and controlling the fin width, which is typically ~0.6 L_g, is a major challenge. The required silicon body thickness for UTB FD and the fin width for DG come from MASTAR.

[8] $I_{sd,leak}$: subthreshold leakage current is defined as the NMOSFET source current per micron of device width, at 25°C, with the drain bias set equal to V_{dd} and with the gate, source, and substrate biases set to zero volts. Total NMOS off-state leakage current (I_{off}) is the NMOSFET drain current per micron of device width at 25°C, and is the sum of the NMOS subthreshold, gate, and junction leakage currents (the latter includes band-to-band tunneling and gate induced drain leakage [GIDL] components). The subthreshold leakage current is assumed to be larger than the junction leakage current component at either 25°C or high-temperature conditions, but see Note [5] for the relation between Isd,leak and gate leakage current density. The yellow and red coloring follows that of $V_{t,sat}$ (see Note [7] above) because $V_{t,sat}$ is a critical determinant of $I_{sd,leak}$. The above subthreshold, gate, and junction leakage current scaling scenario also applies to PMOS devices.

[9] $I_{d,sat}$: saturation drive current is defined as the NMOSFET drain current per micron device width with the gate bias and the drain bias set equal to V_{dd} and the source and substrate biases set to zero. The saturation drive current values have been chosen to continue the historical 17% per year device performance scaling (see Note [16] below) as nearly as possible. PMOS saturation drive current value is assumed to be (40–50) % of the NMOS saturation drive current value. Yellow and red coloring follows that of four items: the parasitic source/drain series resistance, R_{sd} (see Note [13] below), the equivalent electrical oxide thickness in inversion (see Note [4]), the required mobility enhancement factor (see Note [10]), and the ballistic enhancement factor (see Note [12]).

[10] Mobility Enhancement Factor due to Strain (μ_{ratio}). μ_{ratio} = [enhanced mobility]/[reference mobility], where [enhanced mobility] is the actual mobility including the enhancement due to strain, and [reference mobility] is the mobility in the absence of strain. Following the literature, the value of μ_{ratio} is limited to a maximum of 1.8[5]. Mobility enhancement was implemented in product in 2004[6] to meet the required saturation drive current, and hence the coloring for extended planar bulk is initially white. However, there are numerous approaches in the literature for mobility enhancement (including global strain using thin silicon epitaxial layers on SiGe epitaxial layers[7], different process induced strain approaches such as strained thin overlayers of SiN, selective epitaxial SiGe in the PMOS S/D and selective epitaxial Si:C in the NMOS S/D, hybrid orientations, etc.[8,9]). As we continue to scale MOSFETs, it becomes more difficult to maintain the strain, and it is unclear what the optimal approach(es) will be and how to integrate them into the process flow. Consequently, the cells are colored yellow in 2009, when L_g=20 nm and the doping approaches 4.5E18 cm^{-3} according to the MASTAR modeling. For UTB FD and DG, the color is red because strain and enhanced mobility are less well understood in these structures.

[11] Mobility Enhancement Factor for $I_{d,sat}$: captures the improvement in the saturation drive current due to mobility enhancement. This factor is defined as [enhanced $I_{d,sat}$]/$I_{d,ref}$ = $I_{d,ratio}$, where [enhanced $I_{d,sat}$] is the actual saturation drive current including the impact of strain-enhanced mobility and $I_{d,ref}$ is the saturation drive current in the absence of mobility enhancement. MASTAR calculates $I_{d,ratio}$ as a function of the mobility enhancement factor, μ_{ratio} (see Note [10]). Generally, $I_{d,ratio}$ is significantly less than μ_{ratio} due to short channel effects and velocity saturation. The yellow or red coloring follows that for μ_{ratio} (Note [10]).

[12] Effective Ballistic Enhancement Factor is a multiplying factor for $I_{d,sat}$, reflecting quasi-ballistic enhanced transport (due largely to enhanced injection at the source) in highly scaled, ultra-thin body MOSFETs, both UTB FD SOI and DG MOSFETs. Planar bulk CMOS presumably does not have ballistic enhancement because of high doping in these devices. Values for this factor greater than 1 reflect quasi-ballistic enhancement. The value of this parameter is driven by the required saturation drive current to meet performance requirements. The red coloring reflects the lack of known enhanced quasi-ballistic transport solutions for transistors.

[13] R_{sd} is the maximum allowable parasitic series source plus drain resistance (i.e., total resistance for the two sides) per micron of MOSFET width. The values are scaled to allow the required saturation current drive values (see Note [9]) to be met. Yellow and red coloring reflects FEP TWG projections on contact resistance, salicide sheet resistance, and drain extension scaling. The R_{sd} values have been increased from the 2005 ITRS, based on updated information.

[14] $C_{g,ideal}$ is the ideal gate capacitance per micron device width, in inversion. $C_{g,ideal} = [\varepsilon ox / (EOT_{elec})] \times L_g$, where εox is the dielectric constant of thermal silicon dioxide, EOT_{elec} is the equivalent electrical oxide thickness in inversion (see Note [4]), and L_g is the physical gate length (see Note [1]). The red and yellow coloring follows that of EOT_{elec} (see Note [4]).

[15] $C_{g,total}$ is the total gate capacitance per micron device width in inversion. This is the sum of $C_{g,ideal}$ and the parasitic gate overlap/fringing capacitance per micron device width [including the Miller effect]. Red and yellow color here follows that of $C_{g,ideal}$.

[16] τ is the intrinsic transistor delay for NMOS devices at 25°C. $\tau = (C_{g,total} \times V_{dd}) / I_{d,sat}$. .τ for PMOSFETs is assumed to scale similarly, but with PMOS $I_{d,sat}$ ~ (0.4–0.5) × (NMOS $I_{d,sat}$). τ is a reasonable metric for the intrinsic switching delay of the device, while 1/τ is a reasonable metric for the intrinsic switching speed of the device. Red and yellow coloring follows that of both saturation drive current (see Note [9]) and $C_{g,total}$ (see Note [15]).

[17] 1/τ is the NMOS intrinsic switching speed. Maintenance of the historical 17% per year improvement in this parameter is the key scaling goal for high-performance logic. Red and yellow coloring follows that of τ.

Table PIDS3a Low Standby Power Technology Requirements—Near-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in 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*	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
L_g: Physical gate length for LSTP [1]									
Extended Planar Bulk and DG (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**	**18**	**16**
UTB FD (nm)						**22**	**20**	**18**	**17**
EOT: Equivalent Oxide Thickness [2]									
Extended planar bulk (Å)	**19**	**16**	**15**	**14**	**13**	**12**	**11**		
UTB FD (Å)						**13**	**12**	**11**	**10**
DG (Å)						**14**	**13**	**12**	**11**
Gate Poly Depletion and Inversion-Layer Equivalent Thickness [3]									
Extended planar bulk (Å)	**6.2**	**3.3**	**3.4**	**3.3**	**3.2**	**3.1**	**3.1**		
UTB FD (Å)						**4**	**4**	**4**	**4**
DG (Å)						**4**	**4**	**4**	**4**
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]									
Extended planar bulk (Å)	**25.2**	**19.3**	**18.4**	**17.3**	**16.2**	**15.1**	**14.1**		
UTB FD (Å)						**17**	**16**	**15**	**14**
DG (Å)						**18**	**17**	**16**	**15**
$J_{g,limit}$: Maximum gate leakage current density [5]									
Extended Planar Bulk (A/cm^2)	**6.67E-02**	**8.11E-02**	**9.38E-02**	**1.07E-01**	**1.20E-01**	**1.36E-01**	**1.50E-01**		
UTB FD (A/cm^2)						**1.36E-01**	**1.50E-01**	**1.67E-01**	**1.76E-01**
DG (A/cm^2)						**1.36E-01**	**1.50E-01**	**1.67E-01**	**1.88E-01**
V_{dd}: Power Supply Voltage (V) [6]									
Extended Planar Bulk (V)	**1.1**	**1.1**	**1**	**1**	**1**	**1**	**0.95**		
UTB FD (V)						**0.9**	**0.9**	**0.9**	**0.85**
DG (V)						**0.85**	**0.85**	**0.85**	**0.8**
$V_{t,sat}$: Saturation Threshold Voltage [7]									
Extended Planar Bulk (mV)	**534**	**567**	**535**	**535**	**544**	**552**	**547**		
UTB FD (mV)						**395**	**399**	**401**	**404**
DG (mV)						**380**	**362**	**361**	**366**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]									
Extended Planar Bulk (µA/µm)	**3.03E-05**	**3.03E-05**	**3.05E-05**	**3.07E-05**	**3.02E-05**	**3.02E-05**	**3.03E-05**		
UTB FD (µA/µm)						**3.14E-05**	**3.09E-05**	**3.17E-05**	**3.02E-05**
DG (µA/µm)						**1.15E-05**	**2.44E-05**	**2.82E-05**	**2.65E-05**
$I_{d,sat}$: NMOS Drive Current [9]									
Extended Planar Bulk (µA/µm)	**465**	**569**	**501**	**528**	**542**	**560**	**519**		
UTB FD (µA/µm)						**608**	**669**	**744**	**786**
DG (µA/µm)						**550**	**612**	**674**	**702**
Mobility enhancement factor due to strain [10]									
Extended Planar Bulk	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**		
UTB FD and DG						**1.4**	**1.8**	**1.8**	**1.8**

Table PIDS3a Low Standby Power Technology Requirements—Near-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in 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*	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
$I_{d,sat}$ enhancement factor due to strain [11]									
Extended Planar Bulk	**1.19**	**1.17**	**1.16**	**1.17**	**1.17**	**1.16**	**1.17**		
UTB FD						**1.04**	**1.07**	**1.07**	**1.07**
DG						**1.03**	**1.05**	**1.05**	**1.05**
Effective Ballistic Enhancement Factor [12]									
Extended Planar Bulk	**1**	**1**	**1**	**1**	**1**	**1**	**1**		
UTB FD						**1**	**1**	**1**	**1.1**
DG						**1**	**1**	**1**	**1.1**
R_{sd}: Effective Parasitic series source/drain resistance [13]									
Extended Planar Bulk (Ω-µm)	**180**	**180**	**180**	**180**	**180**	**180**	**180**		
UTB FD (Ω-µm)						**200**	**200**	**180**	**160**
DG (Ω-µm)						**210**	**210**	**200**	**200**
$C_{g,ideal}$: Ideal NMOS Device Gate Capacitance [14]									
Extended Planar Bulk (F/µm)	**6.17E-16**	**6.62E-16**	**6.01E-16**	**5.58E-16**	**5.32E-16**	**5.02E-16**	**4.90E-16**		
UTB FD (F/µm)						**4.46E-16**	**4.31E-16**	**4.14E-16**	**4.19E-16**
DG (F/µm)						**4.22E-16**	**4.06E-16**	**3.88E-16**	**3.68E-16**
$C_{g,total}$: Total gate capacitance for calculation of CV/I [15]									
Extended Planar Bulk (F/µm)	**8.57E-16**	**9.02E-16**	**8.21E-16**	**7.68E-16**	**7.32E-16**	**6.92E-16**	**6.70E-16**		
UTB FD (F/µm)						**6.86E-16**	**6.71E-16**	**6.54E-16**	**6.39E-16**
DG (F/µm)						**6.62E-16**	**6.46E-16**	**6.28E-16**	**6.08E-16**
τ =CV/I: NMOSFET intrinsic delay (ps) [16]									
Extended Planar Bulk (ps)	**2.03**	**1.74**	**1.64**	**1.46**	**1.35**	**1.24**	**1.23**		
UTB FD (ps)						**1.02**	**0.9**	**0.79**	**0.69**
DG (ps)						**1.02**	**0.9**	**0.79**	**0.69**
1/τ: NMOSFET intrinsic switching speed (GHz) [17]									
Extended Planar Bulk (GHz)	**493**	**575**	**610**	**685**	**741**	**806**	**813**		
UTB FD (GHz)						**980**	**1111**	**1266**	**1449**
DG (GHz)						**980**	**1111**	**1266**	**1449**

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

Table PIDS3b Low Standby Power Technology Requirements—Long-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*
MPU Physical Gate Length (nm)	*9*	*8*	*7*	*6.3*	*5.6*	*5*	*4.5*
L_g: Physical gate length for LSTP [1]							
Extended Planar Bulk and DG (nm)	**14**	**13**	**12**	**11**	**10**	**9**	**8**
UTB FD (nm)	**16**	**15**					
EOT: Equivalent Oxide Thickness [2]							
Extended planar bulk (Å)							
UTB FD (Å)	**9**	**8**					
DG (Å)	**11**	**10**	**10**	**9**	**9**	**8**	**8**
Gate Poly Depletion and Inversion-Layer Equivalent Thickness [3]							
Extended planar bulk (Å)							
UTB FD (Å)	**4**	**4**					
DG (Å)	**4**	**4**	**4**	**4**	**4**	**4**	**4**
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]							
Extended planar bulk (Å)							
UTB FD (Å)	**13**	**12**					
DG (Å)	**15**	**14**	**14**	**13**	**13**	**12**	**12**
$J_{g,limit}$: Maximum gate leakage current density [5]							
Extended Planar Bulk (A/cm^2)							
UTB FD (A/cm^2)	**1.88E-01**	**2.00E-01**					
DG (A/cm^2)	**2.14E-01**	**2.31E-01**	**2.50E-01**	**2.73E-01**	**3.00E-01**	**3.33E-01**	**3.75E-01**
V_{dd}: Power Supply Voltage (V) [6]							
Extended Planar Bulk (V)							
UTB FD (V)	**0.8**	**0.8**					
DG (V)	**0.8**	**0.8**	**0.8**	**0.75**	**0.75**	**0.7**	**0.7**
$V_{t,sat}$: Saturation Threshold Voltage [7]							
Extended Planar Bulk (mV)							
UTB FD (mV)	**404**	**405**					
DG (mV)	**366**	**371**	**365**	**374**	**378**	**369**	**376**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**3.10E-05**	**3.27E-05**					
DG (µA/µm)	**2.97E-05**	**2.55E-05**	**3.38E-05**	**2.62E-05**	**2.39E-05**	**3.38E-05**	**2.89E-05**
$I_{d,sat}$: NMOS Drive Current [9]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**771**	**838**					
DG (µA/µm)	**738**	**839**	**889**	**895**	**935**	**934**	**946**

Table PIDS3b Low Standby Power Technology Requirements—Long-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*
MPU Physical Gate Length (nm)	*9*	*8*	*7*	*6.3*	*5.6*	*5*	*4.5*
Mobility enhancement factor due to strain [10]							
Extended Planar Bulk							
UTB FD and DG	1.8	1.8	1.8	1.8	1.8	1.8	1.8
$I_{d,sat}$ enhancement factor due to strain [11]							
Extended Planar Bulk							
UTB FD	1.08	1.07					
DG	1.04	1.04	1.04	1.04	1.04	1.04	1.04
Effective Ballistic Enhancement Factor [12]							
Extended Planar Bulk							
UTB FD	1.15	1.18					
DG	1.15	1.22	1.27	1.4	1.45	1.5	1.55
R_{sd}: Effective Parasitic series source/drain resistance [13]							
Extended Planar Bulk (Ω-µm)							
UTB FD (Ω-µm)	150	150					
DG (Ω-µm)	200	180	180	170	160	140	140
$C_{g,ideal}$: Ideal NMOS Device Gate Capacitance [14]							
Extended Planar Bulk (F/µm)							
UTB FD (F/µm)	4.25E-16	4.31E-16					
DG (F/µm)	3.22E-16	3.20E-16	2.96E-16	2.92E-16	2.65E-16	2.59E-16	2.30E-16
$C_{g,total}$: Total gate capacitance for calculation of CV/I [15]							
Extended Planar Bulk (F/µm)							
UTB FD (F/µm)	6.25E-16	6.31E-16					
DG (F/µm)	5.62E-16	5.60E-16	5.26E-16	5.02E-16	4.65E-16	4.49E-16	4.20E-16
τ =CV/I: NMOSFET intrinsic delay (ps) [16]							
Extended Planar Bulk (ps)							
UTB FD (ps)	0.65	0.6					
DG (ps)	0.61	0.53	0.47	0.42	0.37	0.34	0.31
1/τ: NMOSFET intrinsic switching speed (GHz) [17]							
Extended Planar Bulk (GHz)							
UTB FD (GHz)	1538	1667					
DG (GHz)	1639	1887	2128	2381	2703	2941	3226

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

Table PIDS3c Low Operating Power Technology Requirements—Near-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in 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*	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
L_g: Physical gate length for LOP (nm) [1]	***32***	***28***	***25***	***22***	***20***	***18***	***16***	***14***	***13***
EOT: Equivalent Oxide Thickness [2]									
Extended planar bulk (Å)	**12**	**11**	**10**	**9**	**8**	**8**			
UTB FD (Å)					**9**	**9**	**8**	**8**	**8**
DG (Å)					**9**	**9**	**9**	**8**	**8**
Gate Poly Depletion and Inversion-Layer Equivalent Thickness [3]									
Extended planar bulk (Å)	**6.4**	**3.4**	**3.3**	**3.4**	**3.3**	**3.2**			
UTB FD (Å)					**4**	**4**	**4**	**4**	**4**
DG (Å)					**4**	**4**	**4**	**4**	**4**
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]									
Extended planar bulk (Å)	**18.4**	**14.4**	**13.3**	**12.4**	**11.3**	**11.2**			
UTB FD (Å)					**13**	**13**	**12**	**12**	**12**
DG (Å)					**13**	**13**	**13**	**12**	**12**
$J_{g,limit}$: Maximum gate leakage current density [5]									
Extended Planar Bulk (A/ cm^2)	**78**	**89**	**100**	**114**	**125**	**139**			
UTB FD (A/ cm^2)					**125**	**139**	**156**	**179**	**192**
DG (A/ cm^2)					**125**	**139**	**156**	**179**	**192**
V_{dd}: Power Supply Voltage (V) [6]									
Extended Planar Bulk (V)	**0.8**	**0.8**	**0.8**	**0.7**	**0.7**	**0.7**			
UTB FD (V)					**0.7**	**0.7**	**0.6**	**0.6**	**0.6**
DG (V)					**0.7**	**0.7**	**0.6**	**0.6**	**0.6**
$V_{t,sat}$: Saturation Threshold Voltage [7]									
Extended Planar Bulk (mV)	**294**	**296**	**289**	**259**	**246**	**249**			
UTB FD (mV)					**218**	**209**	**195**	**202**	**202**
DG (mV)					**207**	**202**	**203**	**201**	**202**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]									
Extended Planar Bulk (µA/µm)	**9.08E-03**	**7.35E-03**	**8.96E-03**	**1.83E-02**	**2.55E-02**	**3.57E-02**			
UTB FD (µA/µm)					**8.32E-03**	**1.19E-02**	**2.02E-02**	**1.86E-02**	**1.98E-02**
DG (µA/µm)					**5.84E-03**	**7.73E-03**	**8.61E-03**	**1.07E-02**	**1.10E-02**
$I_{d,sat}$: NMOS Drive Current [9]									
Extended Planar Bulk (µA/µm)	**563**	**705**	**760**	**682**	**754**	**760**			
UTB FD (µA/µm)					**766**	**788**	**747**	**763**	**810**
DG (µA/µm)					**780**	**821**	**754**	**826**	**893**
Mobility enhancement factor due to strain [10]	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**
$I_{d,sat}$ enhancement factor due to strain [11]									
Extended Planar Bulk	**1.15**	**1.11**	**1.11**	**1.11**	**1.1**	**1.09**			
UTB FD					**1.07**	**1.07**	**1.07**	**1.06**	**1.06**
DG					**1.05**	**1.05**	**1.05**	**1.05**	**1.04**

Table PIDS3c Low Operating Power Technology Requirements—Near-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in 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*	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
Effective Ballistic Enhancement Factor [12]									
Extended Planar Bulk	1	1	1	1	1	1			
UTB FD					1	1	1.09	1.1	1.15
DG					1	1	1.15	1.18	1.25
R_{sd}: Effective Parasitic series source/drain resistance [13]									
Extended Planar Bulk (Ω-µm)	190	190	190	190	190	190			
UTB FD (Ω-µm)					190	190	180	170	165
DG (Ω-µm)					190	180	180	180	170
$C_{g,ideal}$: Ideal NMOS Device Gate Capacitance [14]									
Extended Planar Bulk (F/µm)	6.01E-16	6.70E-16	6.48E-16	6.13E-16	6.12E-16	5.54E-16			
UTB FD (F/µm)					5.31E-16	4.78E-16	4.60E-16	4.02E-16	3.74E-16
DG (F/µm)					5.31E-16	4.78E-16	4.25E-16	4.02E-16	3.74E-16
$C_{g,total}$: Total gate capacitance for calculation of CV/I [15]									
Extended Planar Bulk (F/µm)	8.41E-16	9.10E-16	8.78E-16	8.13E-16	8.12E-16	7.54E-16			
UTB FD (F/µm)					7.51E-16	6.88E-16	6.60E-16	6.02E-16	5.54E-16
DG (F/µm)					7.71E-16	7.18E-16	6.65E-16	6.43E-16	6.14E-16
$\tau = CV/I$: NMOSFET intrinsic delay (ps) [16]									
Extended Planar Bulk (ps)	1.19	1.03	0.92	0.83	0.75	0.69			
UTB FD (ps)					0.69	0.61	0.53	0.47	0.41
DG (ps)					0.69	0.61	0.53	0.47	0.41
$1/\tau$: NMOSFET intrinsic switching speed (GHz) [17]									
Extended Planar Bulk (GHz)	840	971	1087	1205	1333	1449			
UTB FD (GHz)					1449	1639	1887	2128	2439
DG (GHz)					1449	1639	1887	2128	2439

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

Table PIDS3d Low Operating Power Technology Requirements—Long-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*
MPU Physical Gate Length (nm)	*9*	*8*	*7*	*6.3*	*5.6*	*5*	*4.5*
L_g: Physical gate length for LOP (nm) [1]	***11***	***10***	***9***	***8***	***7***	***6.5***	***6***
EOT: Equivalent Oxide Thickness [2]							
Extended planar bulk (Å)							
UTB FD (Å)	**7**						
DG (Å)	**8**	**7**	**7**	**7**	**7**	**6**	**6**
Gate Poly Depletion and Inversion-Layer Equivalent Thickness [3]							
Extended planar bulk (Å)							
UTB FD (Å)	**4**						
DG (Å)	**4**	**4**	**4**	**4**	**4**	**4**	**4**
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]							
Extended planar bulk (Å)							
UTB FD (Å)	**11**						
DG (Å)	**12**	**11**	**11**	**11**	**11**	**10**	**10**
$J_{g,limit}$: Maximum gate leakage current density [5]							
Extended Planar Bulk (A/ cm^2)							
UTB FD (A/ cm^2)	**227**						
DG (A/ cm^2)	**227**	**250**	**278**	**313**	**357**	**385**	**417**
V_{dd}: Power Supply Voltage (V) [6]							
Extended Planar Bulk (V)							
UTB FD (V)	**0.5**						
DG (V)	**0.6**	**0.5**	**0.5**	**0.5**	**0.5**	**0.45**	**0.45**
$V_{t,sat}$: Saturation Threshold Voltage [7]							
Extended Planar Bulk (mV)							
UTB FD (mV)	**187**						
DG (mV)	**202**	**188**	**194**	**190**	**195**	**190**	**201**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**3.71E-02**						
DG (µA/µm)	**1.31E-02**	**2.23E-02**	**1.94E-02**	**2.55E-02**	**2.41E-02**	**3.26E-02**	**2.40E-02**
$I_{d,sat}$: NMOS Drive Current [9]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**716**						
DG (µA/µm)	**916**	**808**	**850**	**900**	**919**	**874**	**876**
Mobility enhancement factor due to strain [10]	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.8**
$I_{d,sat}$ enhancement factor due to strain [11]							
Extended Planar Bulk							
UTB FD	**1.06**						
DG	**1.04**	**1.04**	**1.04**	**1.04**	**1.03**	**1.03**	**1.03**

Table PIDS3d Low Operating Power Technology Requirements—Long-term Years

Grey cells delineate one of two time periods: either before initial production ramp has started for ultra-thin body fully depleted (UTB FD) SOI or double-gate (DG) MOSFETs, or beyond when planar bulk or UTB FD MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*22*	*20*	*18*	*16*	*14*	*13*	*11*
MPU Physical Gate Length (nm)	*9*	*8*	*7*	*6.3*	*5.6*	*5*	*4.5*
Effective Ballistic Enhancement Factor [12]							
Extended Planar Bulk							
UTB FD	1.24						
DG	1.27	1.35	1.43	1.45	1.5	1.6	1.68
R_{sd}: Effective Parasitic series source/drain resistance [13]							
Extended Planar Bulk (Ω-µm)							
UTB FD (Ω-µm)	160						
DG (Ω-µm)	170	155	150	140	140	140	140
$C_{g,ideal}$: Ideal NMOS Device Gate Capacitance [14]							
Extended Planar Bulk (F/µm)							
UTB FD (F/µm)	3.14E-16						
DG (F/µm)	3.16E-16	3.14E-16	2.82E-16	2.51E-16	2.20E-16	2.24E-16	2.07E-16
$C_{g,total}$: Total gate capacitance for calculation of CV/I [15]							
Extended Planar Bulk (F/µm)							
UTB FD (F/µm)	5.15E-16						
DG (F/µm)	5.56E-16	5.24E-16	4.82E-16	4.41E-16	4.10E-16	4.14E-16	3.97E-16
τ =CV/I: NMOSFET intrinsic delay (ps) [16]							
Extended Planar Bulk (ps)							
UTB FD (ps)	0.36						
DG (ps)	0.36	0.32	0.28	0.24	0.22	0.21	0.2
1/τ: NMOSFET intrinsic switching speed (GHz) [17]							
Extended Planar Bulk (GHz)							
UTB FD (GHz)	2778						
DG (GHz)	2778	3125	3571	4167	4545	4762	5000

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

Notes for Tables PIDS3a through d:

As described in the text, MASTAR[1,2,3], a detailed analytical MOSFET modeling software package, has been utilized to generate the parameter values in these tables. The MASTAR modeling package and user's manual are in the backup material on the ITRS website, as well as the detailed MASTAR simulations that underlay these tables. Also note that the parameters in this table are for an NMOSFET with nominal gate length at an operating temperature of 25°C. Furthermore, although there are multiple MOSFETs in a typical logic chip, with differing threshold voltages, Ion, Ioff, and oxide thickness, for LSTP logic the transistor specified in this table is the transistor with the highest threshold voltage, lowest Ion and Ioff, highest oxide thickness, and slowest intrinsic transistor speed. The majority of the transistors on the chip are of this type, in order to keep the leakage and static power dissipation within tolerable limits. This transistor is specified here because it drives the technology. In contrast, for LOP logic, the transistor specified in this table is the "standard" transistor, with intermediate threshold voltage, I_{on}, and I_{off}. The majority of the transistors on the chip are of this type, because the performance requirements are critical, and standby power dissipation is less critical than for LSTP. Dynamic power dissipation is critical here, and V_{dd} is rapidly scaled to keep this within tolerable limits. This transistor is specified here because it drives the technology.

As explained in the text, multiple parallel options for the transistor type are included in the tables, including planar bulk CMOS extended to its practical scaling limits, ultra-thin body fully-depleted (UTB FD) SOI CMOS, also extended to its practical scaling limits, and double-gate (DG) CMOS (e.g., FinFETs). Note that, for LOP, the limit for planar bulk CMOS is through 2012, and for UTB FD is through 2016. UTB FD and DG start in 2011, with overlap of the three options from 2011 through 2012. In contrast, for LSTP, the limit for planar bulk CMOS is through 2013, and UTB FD continues through 2017. UTB FD and DG start in 2012, with overlap of the three options from 2012 through 2013.

For each transistor option, the scaling of the numbers in the tables reflects a particular scaling scenario in which we have attempted to optimally scale to meet the key goals while keeping the performance, short channel effects, and other key characteristics under control. For LSTP, the key goal is ultra-low leakage current, while for LOP the goal is relatively high speed and low dynamic power dissipation, along with low leakage current (but not as low as for LSTP). However, there are numerous parameters (such as EOT, V_{dd}, $I_{sd,leak}$, etc.) that can be varied, and different scaling scenarios are possible by making different choices on the scaling of these parameters. The scenarios in this table were selected to be as representative of the industry as possible. In particular, in these tables, high-κ gate dielectric and metal gate electrode are assumed to be available in 2008. See the figures and discussion in the text for why high-κ gate dielectric is required in 2008.

[1] L_g is the physical gate length: the final, as-etched length of the bottom of the gate electrode. The values here lag behind the gate length values for high- performance logic by two years (LOP) or four years (LSTP) in order to meet the stringent leakage current requirements. For LSTP, for UTB FD devices, late in the ITRS L_g scaling lags slightly behind that for DG MOSFETs. This is because of the difficulty in scaling UTB FD MOSFETs and keeping the leakage current within tolerance for such short devices. The gate dimensional control requirement is set by the Lithography and FEP Etch ITWGs, and is assumed to have a three-sigma value of $\pm 12\% \times L_g$. It is expected that meeting this requirement will become increasingly difficult with scaling (refer to the Lithography chapter and the FEP Chapter). Gate length variation is assumed to be a primary factor responsible for driving device parameter variation.

[2] EOT: for a gate dielectric of thickness T_d and relative dielectric constant κ EOT is defined by: EOT = T_d / (κ /3.9), where 3.9 is the relative dielectric constant of thermal silicon dioxide. For a MOSFET with the gate dielectric of thickness T_d, the ideal gate capacitance per unit area is the same as that of a similar MOSFET, but with a gate dielectric made up of thermal silicon dioxide with a thickness of EOT. It is projected that high-κ gate dielectric will be required by 2008 to control the gate leakage (see the text for further discussion on this point.) The yellow coloring reflects the uncertainty in our high-κ solutions for EOT < 0.9nm, while the red coloring reflects the increased uncertainty of solutions for EOT < 0.7nm. Measurement of EOT is complicated, and is usually done via sophisticated MOS capacitor-voltage (CV) measurements on MOS capacitors or via optical measurements.

[3] Gate Poly Depletion and Inversion-Layer Equivalent Thickness: accounts for gate electrode depletion and inversion-layer effects, including quantum effects, both of which are calculated by MASTAR. For polysilicon gate electrodes, the portion of the electrical thickness adjustment due to gate electrode depletion is dependent on the polysilicon doping. For 2008 and beyond, it is assumed that metal-gate electrodes, which reduce the gate depletion effect to zero, will be introduced. The abrupt reduction in this parameter for 2008 reflects the zero depletion. For 2008 and beyond, the difference between the parameter value for planar bulk versus the 4 nm value for DG and UTB FD reflects the light channel doping in the latter types of MOSFET and the heavy channel doping in planar bulk. For planar bulk CMOS, the metal gate work function needs to be near the silicon conduction band edge for NMOS and near the silicon valence band edge for PMOS to properly set the MOSFET threshold voltage, as with polysilicon gates. For UTB FD and DG MOSFETs, the channel is very thin and lightly doped, and the work function of the metal gates needs to be within a few hundred millivolts of the silicon midgap (i.e., "near silicon midgap" work function) to properly set the MOSFET's threshold voltage. Multiple V_t's may be obtained by tuning the work function. The yellow and red colors generally follow that of EOT (see Note [2] above), reflecting the tight coupling between the metal gate electrode and the high-κ gate dielectric. For UTB FD and DG, the color is red, reflecting the uncertainties of setting and tuning the effective work function.

[4] EOT_{elec} is the sum of EOT and gate poly depletion and inversion-layer thickness (see Notes [2] and [3] above). For MOSFETs in inversion, ideal gate capacitance per unit area (see Note [14]) is $\varepsilon_{ox} / (EOT_{elec})$, where ε_{ox} is the dielectric constant of thermal silicon dioxide. The equivalent electrical oxide thickness in inversion is used in calculations of the transistor intrinsic delay, CV/I (see Note [16]). Red/yellow coloring follows that of EOT and gate poly depletion and inversion-layer thickness.

[5] $J_{g,limit}$ is the maximum allowed gate leakage current density at 25°C, and it is measured with the gate biased to V_{dd} and the source, drain, and substrate all set to ground. $J_{g,limit}$ is related to $I_{sd,leak,TGT}$, the nominal subthreshold leakage current target at 25°C . $I_{sd,leak,TGT}$ is set to 30 pA/micron device width for LSTP and to 5nA/micron device width for LOP. Specifically, $J_{g,limit} = [I_{sd,leak,TGT} / L_g] \times$ [Hi T Factor] / [Circuit Factor]. For LOP, Hi T Factor is set to 5, and it accounts for the high operating temperature (well over room temperature, but not as high as the 100°C for high-performance logic, where Hi T Factor = 10). Hi T Factor accounts for both the rapid increase in $I_{sd,leak}$ with temperature and the insensitivity of gate leakage current (since it is due to direct tunneling) to temperature. For LSTP, where the operating temperature is expected to be room temperature, Hi T Factor = 1. The Circuit Factor is set to 1, and it accounts for the differences between the subthreshold leakage current and the gate leakage current in logic gates compared to single isolated transistors as specified by the $J_{g,limit}$ and $I_{sd,leak}$ parameters in this table. (The reason for these differences is the different bias conditions on the various transistors in logic gates compared to the bias conditions used to define $I_{sd,leak}$ (see Note [8]) and $J_{g,limit}$ for the NMOS transistor in this table). The values of Hi T Factor and Circuit Factor used here are rough estimates. The yellow and red coloring follows that of EOT (see Note [2] above).

[6] V_{dd} is the nominal power supply voltage. It has been chosen to maintain sufficient voltage over-drive [V_{dd} – saturation threshold voltage (see Note [7]) in order to meet the required saturation current drive values while still maintaining reasonable vertical gate dielectric electric field strengths. Target power supply voltage values for actual ICs may vary ± 10% (or more) from the values in this table, depending on the particular circuit design application or technology optimization. Note that V_{dd} is relatively high and scales slowly for LSTP, because the saturation threshold voltage is high here to keep the subthreshold leakage current very low. On the other hand, for LOP V_{dd} scales down rapidly in order to keep the dynamic power dissipation low.

[7] $V_{t,sat}$ is the saturation threshold voltage for a nominal gate length transistor with drain bias set equal to V_{dd}, as calculated by MASTAR. The threshold voltage values and the corresponding subthreshold leakage current values (see Note [8]) have been chosen to maintain sufficient voltage over-drive (V_{dd} – saturation threshold voltage) in order to meet the required saturation current drive values (see Note [9]). For planar bulk, the yellow color is associated with high substrate doping exceeding 4E18cm^{-3} (from MASTAR), while the red color is associated with very high substrate doping exceeding 6E18 cm^{-3}. These doping levels are required to set the threshold voltage to the desired level and to keep short channel effects under control. For UTB FD devices, the color is red from the beginning because of the challenges of controlling the very thin silicon body thickness (right from the beginning, ~7 nm for LOP and < 7 nm for LSTP) required to control $V_{t,sat}$ and short channel effects. For DG devices, the color is red right from the beginning because there are numerous issues that are not understood here; in particular, defining and controlling the fin width, which is typically ~0.6 L_g, is a major challenge. The required silicon body thickness for UTB FD and the fin width for DG come from MASTAR.

[8] $I_{sd,leak}$: subthreshold leakage current is defined as the NMOSFET source current per micron of device width, at 25°C, with the drain bias set equal to V_{dd} and with the gate, source, and substrate biases set to zero volts. Total NMOS off-state leakage current (I_{off}) is the NMOSFET drain current per micron of device width at 25°C, and is the sum of the NMOS subthreshold, gate, and junction leakage current (which includes band-to-band tunneling and gate induced drain leakage [GIDL]) components). The subthreshold leakage current is assumed to be larger than the junction leakage current component at either 25°C or high-temperature conditions, but see Note [5] for the relation between $I_{sd,leak}$ and gate leakage current density. The yellow and red coloring follows that of the $V_{t,sat}$ (see Note [7] above) because $V_{t,sat}$ is a critical determinant of $I_{sd,leak}$. The above subthreshold, gate, and junction leakage current scaling scenario also applies to PMOS devices. For LSTP, meeting the $I_{sd,leak}$ target of ~30pA/µm is the key scaling goal.

[9] $I_{d,sat}$: saturation drive current is defined as the NMOSFET drain current per micron device width with the gate bias and the drain bias set equal to V_{dd} and the source and substrate biases set to zero. PMOS saturation drive current value is assumed to be (40–50)% of the NMOS saturation drive current value. Yellow/red coloring follows that of four items: the parasitic source/drain series resistance, R_{sd} (see Note [13] below), the equivalent electrical oxide thickness in inversion (see Note [4]), the mobility enhancement factor (see Note [10]), and the ballistic enhancement factor (see Note [12]).

[10] Mobility Enhancement Factor due to Strain (μ_{ratio}). μ_{ratio} = [enhanced mobility]/[reference mobility], where [enhanced mobility] is the actual mobility including the enhancement due to strain, and [reference mobility] is the mobility in the absence of enhancement. Following the literature, the value of μ_{ratio} is limited to a maximum of 1.8[10]. Mobility enhancement was implemented in product in 2004[11] to meet the required saturation drive current, and hence the coloring for extended planar bulk is initially white. However, there are numerous approaches in the literature for mobility enhancement (including global strain using thin silicon epitaxial layers on SiGe epitaxial layers[12], different process induced strain approaches such as strained thin overlayers of SiN, selective epitaxial SiGe in the PMOS S/D and selective epitaxial Si:C in the NMOS S/D, hybrid orientations, etc.[13,14]). As we continue to scale MOSFETs, it becomes more difficult to maintain the strain, and it is unclear what the optimal approach(es) will be and how to integrate them into the process flow. The cells are colored yellow when L_g=20 nm. For UTB FD and DG, the color is red because strain and enhanced mobility are less well understood in these structures.

[11] Mobility Enhancement Factor for $I_{d,sat}$: captures the improvement in the saturation drive current due to mobility enhancement from strain. This factor is defined as [enhanced $I_{d,sat}$]/$I_{d,ref}$ = $I_{d,ratio}$, where [enhanced $I_{d,sat}$] is the actual saturation drive current including the impact of enhanced mobility and $I_{d,ref}$ is the saturation drive current in the absence of mobility enhancement. MASTAR calculates $I_{d,ratio}$ as a function of the mobility enhancement factor, μ_{ratio} (see Note [10]). Generally, $I_{d,ratio}$ is significantly less than μ_{ratio} due to short channel effects and velocity saturation. The yellow or red coloring follows that for μ_{ratio} (Note [10]).

[12] Effective Ballistic Enhancement Factor is a multiplying factor for $I_{d,sat}$, reflecting quasi-ballistic enhanced transport (due largely to enhanced injection at the source) in highly scaled, ultra-thin body MOSFETs, both UTB FD SOI and DG MOSFETs. Planar bulk CMOS presumably does not have ballistic enhancement because of high doping in these devices. Values for this factor greater than 1 reflect quasi-ballistic enhancement. The value of this parameter is driven by the required saturation drive current to meet performance requirements. The red coloring reflects the lack of known enhanced quasi-ballistic transport solutions for transistors.

[13] R_{sd} is the maximum allowable parasitic series source plus drain resistance (i.e., total resistance for the two sides) per micron of MOSFET width. The values are scaled to allow the required saturation current drive values (see Note [9]) to be met. Yellow/red coloring reflects FEP TWG projections on contact resistance, salicide sheet resistance, and drain extension scaling. The R_{sd} values have been increased from the 2005 ITRS, based on updated information.

[14] $C_{g,ideal}$ is the ideal gate capacitance per micron device width, in inversion. $C_{g,ideal} = [\varepsilon_{ox}/(EOT_{elec})] \times L_g$, where ε_{ox} is the dielectric constant of thermal silicon dioxide, EOT_{elec} is the equivalent electrical oxide thickness in inversion (see Note [4]), and L_g is the physical gate length (see Note [1]). The red/yellow coloring follows that of EOT_{elec} (see Note [4]).

[15] $C_{g,total}$ is the total gate capacitance per micron device width in inversion. This is the sum of $C_{g,ideal}$ and the parasitic gate overlap/fringing capacitance per micron device width [including the Miller effect]. Red/yellow color here follows that of $C_{g,ideal}$.

[16] τ is the intrinsic transistor delay for NMOS devices at 25°C. $\tau = (C_{g,total} \times V_{dd}) / I_{d,sat}$. τ for PMOSFETs is assumed to scale similarly, but with PMOS $I_{d,sat}$ ~ (0.4–0.5) × (NMOS $I_{d,sat}$). τ is a reasonable metric for the intrinsic switching delay of the device, while $1/\tau$ is a reasonable metric for the intrinsic switching speed of the device. Red/yellow coloring follows that of both saturation drive current (see Note [9]) and $C_{g,total}$ (see Note [15]).

[17] $1/\tau$ is the NMOS intrinsic switching speed. Red/yellow coloring follows that of τ.

LOGIC POTENTIAL SOLUTIONS

There is a strong correlation between the challenges indicated by the colors in the technology requirements tables and the potential solutions (see Figure PIDS5). In many cases, red coloring (manufacturable solutions are not known) in the technology requirements tables corresponds to the projected year of introduction for a potential solution to the challenge indicated by these colors. Another important general point is that each potential solution highlighted in the Potential Solutions figure involves significant technological innovation. The qualification/pre-production interval has been set to one and a half years in order to understand and deal with any new and different reliability, yield, and process integration issues associated with these innovative solutions. Most of the potential solutions, with the exception of high-κ gate dielectric and metal gate electrodes, are required first for high-performance logic. Finally, the industry faces a major overall challenge due to the sheer number of major technological innovations required over the next five years: enhanced mobility (already implemented but requiring continuous improvement with scaling), high-κ gate dielectric, metal gate electrodes, ultra-thin body, fully depleted SOI and multiple-gate MOSFETs with quasiballistic enhanced transport.

The first potential solution, enhanced mobility due to strain, is needed to enhance the saturation current drive to meet transistor performance targets. (Note that, in the Logic Technology Requirements tables, significantly enhanced mobility is assumed in the projections.) It was first implemented in 2004 for high-performance logic. There are numerous techniques to implement enhanced mobility, including via various types of process-induced local strain or by globally induced strain in a thin strained silicon layer, either on relaxed SiGe layers with controlled percentages of Ge or in SOI substrates. Other approaches include use of hybrid orientations (e.g., PMOSFET mobility is highest for the (110) substrate orientation) or use of strained SiGe or (eventually) strained Ge channels. The potential solutions figure indicates that continuous improvement will be needed here, to increase the mobility enhancement to the maximum extent possible for both NMOSFET and PMOSFET transistors, to integrate mobility enhancement optimally with the overall process flow, and eventually to utilize mobility enhancement for advanced MOSFETs such as UTB SOI and multiple-gate MOSFETs. In addition, continuous improvement will be needed to deal with the reduced effectiveness of process-induced strain techniques with scaling: as the spacing between transistors is reduced with scaling, techniques such as embedded SiGe or Si:C in the source/drain and the addition of stressed thin film silicon nitride liner layers over the top of the transistor tend to becomes less effective at inducing stress in the channel.

In order to scale the basic MOSFET structure significantly beyond 2007 (corresponding to physical gate length of 25 nm for high-performance logic), key technology issues involving the current standard device gate stack (silicon oxy-nitride gate dielectric and doped polysilicon gate electrode) need to be addressed. As the physical gate length is scaled, ideally the gate dielectric equivalent oxide thickness (EOT) is scaled correspondingly to control short-channel effects and to increase the saturation current drive. However, continued thinning of the silicon oxy-nitride results in a significant increase in gate leakage current due to an approximately exponential increase in the direct tunneling current(see Figures PIDS2-4). In addition, the effectiveness of continued EOT reduction becomes limited due to the non-scalability of gate electrode depletion and inversion layer effects, which both increase the equivalent electrical oxide thickness in inversion. High-κ gate dielectric material is a potential solution to solve the problem of high gate leakage current, since the gate leakage current density corresponding to a given EOT is much smaller for high-κ than for oxy-nitride gate dielectric. For all three logic types, it is projected that high-κ gate dielectric will be required by 2008 (see the discussion in the logic technology requirements section for more detail.) For all three logic types, metal gate electrodes are also projected for 2008, in order to effectively prevent gate electrode depletion and hence allow acceptable scaling of the equivalent electrical oxide thickness in inversion. To set the threshold voltage correctly for planar bulk CMOS, the gate electrode work function needs to be near the silicon valence band edge for PMOSFETs and near the silicon conduction band edge for NMOSFETs. Hence, different metals will probably be needed for the PMOSFET and NMOSFET.

As scaling proceeds, it will become increasingly difficult to effectively scale planar bulk CMOS devices. In particular, adequately controlling short channel effects is projected to become especially problematical for such short channel devices. Furthermore, the channel doping will need to be increased to exceedingly high values, which will tend to reduce the mobility and to cause high leakage current due to band-to-band tunneling between the drain and the body. Finally, the total number of dopants in the channel for such small MOSFETs becomes relatively small, which results in large random fluctuations in the dopant placement and number, and hence unacceptably large statistical variation of the threshold voltage. These difficulties become worse with further scaling. A potential solution is to utilize ultra-thin body, fully depleted (UTB FD) SOI MOSFETs. The channel doping is relatively light, and for such devices, the threshold voltage can be set by adjusting the work function of the gate electrode, rather than by doping the channel as in planar bulk MOSFETs. Metal gate electrodes with near-midgap work functions will be needed to set the threshold voltage to the desired values. Because of the different work functions in this case, the electrode material will presumably be different than those utilized for planar bulk MOSFETs. In fact, one electrode material with work function tunable within several hundred meV on either side of midgap may be possible. Due to the lightly doped and fully depleted channel, the threshold

voltage control by the work function of the gate electrode, and the ultra-thin body, these SOI MOSFETs are considerably more scalable and develop more saturation drive current than comparable planar bulk MOSFETs. Single gate SOI MOSFETs are projected for 2010 for high-performance logic. Multiple-gate, ultra-thin body, fully depleted MOSFETs are both more complex and more scalable, and are projected to be implemented in 2011 for high-performance logic. As the gate length is scaled well below 20 nm, the fully depleted, lightly doped MOSFETs are likely to require enhanced quasi-ballistic transport to meet the performance requirements (see Row 12 of the Logic Technology Requirements tables for detailed numbers). These enhancements will be obtained through reduced scattering and particularly through improved injection at the source. Eventually, late in the Roadmap, more forward-looking solutions, such as utilization of high transport materials for the channel (e.g., Ge or III-V or silicon-based nanowire structures or carbon nanotubes) to further enhance the transport, may be adopted.

Finally, at the end of the Roadmap or beyond, MOSFET scaling will likely become ineffective and/or very costly, and novel, non-CMOS (emerging research) devices and/or circuits/architectures are a potential solution then (see Emerging Research Devices section for detailed discussion of these). Such solutions may be integrated, functionally or physically, with a CMOS baseline technology that takes advantage of the high-performance, cost-effective, and very dense CMOS logic that will have been developed and implemented by then.

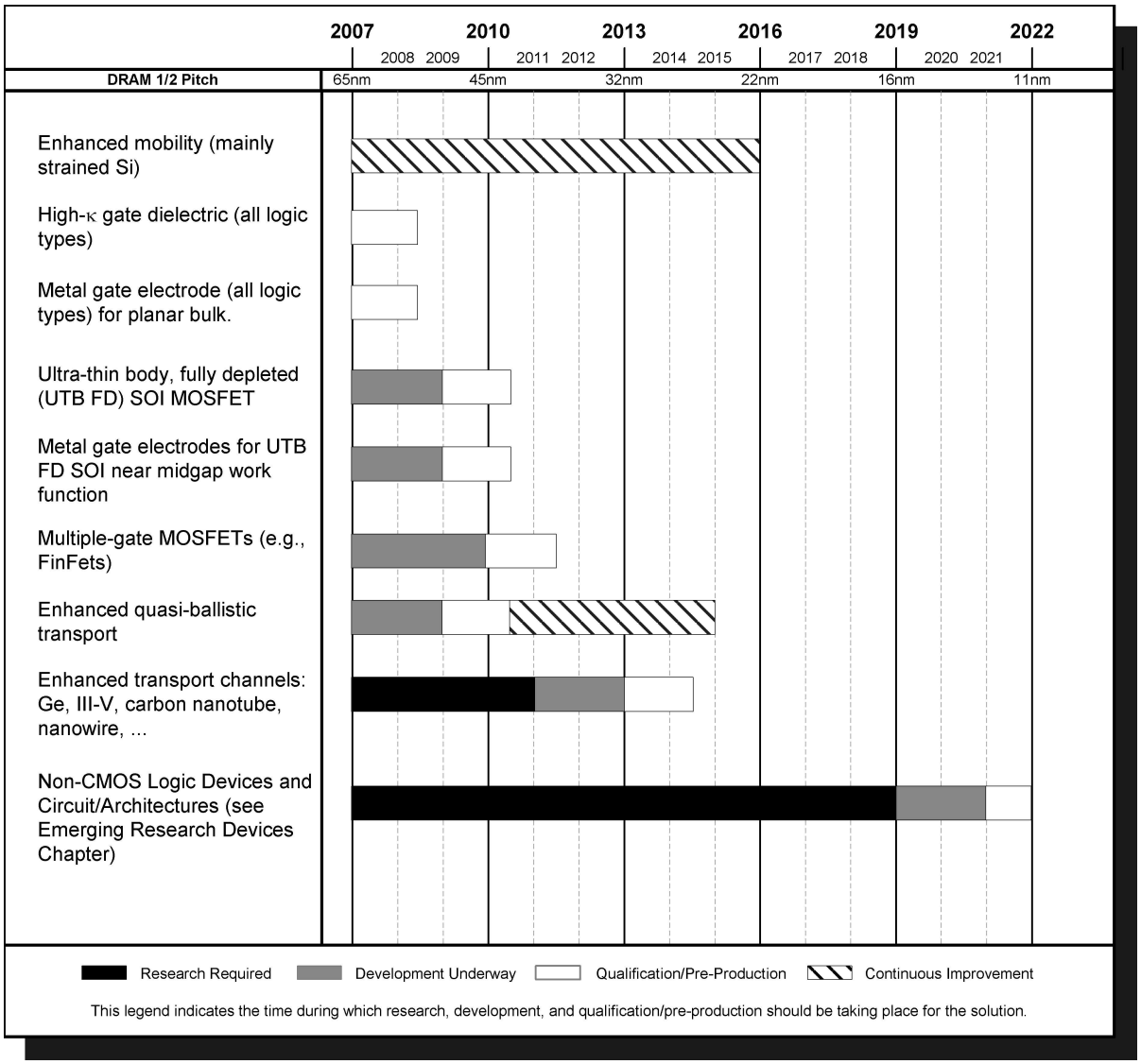

Figure PIDS5 Logic Potential Solutions

MEMORY TECHNOLOGY REQUIREMENTS AND POTENTIAL SOLUTIONS

DRAM

Technical requirements for DRAMs become more difficult with scaling (see Table PIDS4a and b). The process associated with 193 nm argon fluoride (ArF) immersion lithography technology is the key for 60 nm or smaller half pitch DRAMs. However there exist several significant process flow issues for both trench and stack capacitor structures from a production standpoint. Process steps such as capacitor formation or high aspect ratio contact etches require photoresists that can stand up for a prolonged etch time. To overcome these challenges, the technology related to photoresists with a hard mask layer for pattern transfer is gaining in importance. Furthermore, continuous improvements in lithography and etch will be needed.

On the other hand, with the scaling of peripheral CMOS devices, a low temperature process flow is required for process steps after formation of these devices. This is a challenge for DRAM cells with stack capacitors, which are typically constructed after the CMOS devices are formed, and which will therefore be limited to low temperature processing. In addition, the planar access device (cell FET) for the one transistor-one capacitor (1T-1C) cell is getting difficult to design due to the need to maintain a low level of both subthreshold leakage and junction leakage current to meet the retention time requirements. Recessed channel cell FET is being adapted for 80 nm or smaller half pitch DRAMs. Furthermore,

Fin-type cell FET is needed for the 50 nm half pitch and beyond to meet the high drive current requirement with low voltage operation. Another challenge is a highly reliable gate insulator. A highly boosted gate voltage is required to drive higher drain current with the relatively high threshold voltage adopted for the cell FET to suppress the subthreshold leakage current. The scaling of the DRAM cell FET dielectric, maximum word line (WL) level, and the electric field in the cell FET dielectric is plotted in Figure PIDS6. Because of the gate insulator reliability concerns, the EOT is relatively large, and the equivalent electric field in the dielectric is held approximately constant with scaling, as shown in the figure. Process requirements for DRAMs such as front end isolation, low resistance materials for the word lines, self-aligned and high aspect ratio etches, planarization, and Cu interconnection are all needed for future high density DRAMs.

Since the DRAM storage capacitor gets physically smaller with scaling, the EOT must scale down sharply to maintain adequate storage capacitance with scaling. To scale the EOT, dielectric materials having high relative dielectric constant (κ) will be needed. Several manufacturers have introduced MIM (Metal Insulator Metal) capacitors using HfSiO and Al_2O_3 (κ~ 10–25) for DRAMs with 70 nm ½ pitch in 2006. Eventually, in 2010, MIM structures and dielectric materials with even higher κ values such as ZrO_2 will be adopted.. Finally, it is expected that very high κ values of 50 and greater will be needed later in the Roadmap (See Figure PIDS8, DRAM Potential Solutions, for details). Also, the physical thickness of the high-κ insulator should be scaled down to fit the minimum feature size. All in all, maintaining sufficient storage capacitance will pose an increasingly difficult requirement for continued scaling of DRAM devices. The scaling of DRAM storage node cell dielectric, DRAM storage node capacitor voltage, and equivalent electric field of the storage capacitor dielectric is plotted in Figure PIDS7. As shown in the figure, the electric field in the capacitor dielectric is expected to increase sharply with scaling, due to aggressive scaling of EOT.

Keeping the chip size approximately constant as the DRAM capacity (number of bits per chip) is increased with scaling is very important from a chip cost point of view. In order to do so, the cell size factor (*a*) scaling (where a = [DRAM cell size]/[DRAM half pitch]) is critically important. Along with the overall technology scaling, some companies have started production of DRAMs with an *a* of 6 in 2006, but the other companies are staying with an *a* of 8 for now. It is expected that more companies will likely move to use an *a* of 6 to shrink chip size further. When *a* is decreased from 8 to 6, the array area efficiency (the ratio of cell storage array to total chip area) is decreased from 0.63 to 0.56 because the peripheral circuit area stays the same. Table PIDS4a and b does not include $a = 4$ because a $4F^2$ memory storage cell structure is not considered feasible yet, even though there are some research papers published.

Table PIDS4a DRAM Technology Requirements—Near-term Years

Year in Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ Pitch (nm) [1]	*68*	*58*	*50*	*45*	*40*	*36*	*32*	*30*	*25*
DRAM cell size (μm^2) [2]	**0.0277**	**0.0202**	**0.0150**	**0.0122**	**0.0096**	**0.0078**	**0.0061**	**0.0054**	**0.0038**
DRAM storage node cell capacitor dielectric: equivalent oxide thickness EOT (nm) [3]	**1.2**	**0.90**	**0.80**	**0.60**	**0.50**	**0.40**	**0.30**	**0.30**	**0.30**
DRAM storage node cell capacitor voltage (V) [4]	**0.65**	**0.65**	**0.60**	**0.60**	**0.55**	**0.55**	**0.5**	**0.50**	**0.45**
Equivalent Electric field of capacitor dielectric, (MV/cm) [5]	**5.7**	**7.2**	**7.5**	**10.0**	**11.0**	**13.8**	**16.7**	**16.7**	**15.0**
DRAM cell FET structure [6]	**RCAT**	**RCAT**	**RCAT**	**FinFET**	**FinFET**	**FinFET**	**FinFET**	**FinFET**	**FinFET**
DRAM cell FET dielectric: equivalent oxide thickness, EOT (nm) [7]	**5.0**	**5.0**	**4.5**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**
Maximum Wordline (WL) level (V) [8]	**3.0**	**2.8**	**2.7**	**2.7**	**2.7**	**2.7**	**2.6**	**2.6**	**2.4**
Negative Wordline (WL) use [9]	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
Equivalent Electric field of cell FET device dielectric (MV/cm) [10]	**6.00**	**5.60**	**6.00**	**6.75**	**6.75**	**6.75**	**6.50**	**6.50**	**6.00**
Cell Size Factor: a [11]	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**
Array Area Efficiency [12]	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**
Minimum DRAM retention time (ms) [13]	**64**	**64**	**64**	**64**	**64**	**64**	**64**	**64**	**64**
DRAM soft error rate (fits) [14]	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**
V_{int} (support FET voltage) [V] [15]	**1.3**	**1.2**	**1.1**	**1.1**	**1.1**	**1.1**	**1.1**	**1.0**	**0.9**
Support nMOS EOT [nm] [16]	**3.2**	**3**	**2.6**	**2.6**	**2.5**	**2.2**	**2**	**1.8**	**1.6**
Support PMOS Gate Electrode [17]	**P+Poly/W**	**P+Poly/W**	**P+Poly/W**	**P+Poly/W**	**P+Poly/W**	**P+Poly/W**	**P+Poly/W**	**TiN**	**TiN**
Support Gate Oxide [18]	**SiON**	**SiON**	**SiON**	**SiON**	**SiON**	**SiON**	**HfSiON**	**HfSiON**	**HfSiON**
Support min. L_{gate} for NMOS FET, physical [nm] [19]	**100**	**90**	**75**	**75**	**65**	**60**	**50**	**48**	**40**
Support $I_{sat\text{-}n}$ [μA/μm] (25C, V_g=V_d=V_{int}) [20]	**500**	**465**	**470**	**450**	**410**	**430**	**450**	**445**	**440**
Support min. V_{tn} (25C, $G_{m,max}$, V_d=55mV) [21]	**0.40**	**0.40**	**0.38**	**0.37**	**0.37**	**0.33**	**0.33**	**0.31**	**0.31**
Support $I_{sat\text{-}p}$ [μA/μm] (25C, Vg=V_d=-V_{int}) [22]	**220**	**210**	**220**	**210**	**165**	**170**	**175**	**170**	**190**
Support min. V_{tp} (25C, $G_{m,max}$, V_d=55mV) [23]	**-0.45**	**-0.40**	**-0.38**	**-0.38**	**-0.38**	**-0.34**	**-0.34**	**-0.32**	**-0.32**

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

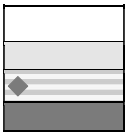

Table PIDS4b DRAM Technology Requirements—Long-term Years

Year in Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
DRAM ½ Pitch (nm) [1]	*22*	*20*	*18*	*16*	*14*	*13*	*12*
DRAM cell size (μm^2) [2]	**0.0029**	**0.0024**	**0.0019**	**0.00154**	**0.00118**	**0.00101**	**0.00086**
DRAM storage node cell capacitor dielectric: equivalent oxide thickness EOT (nm) [3]	**0.30**	**0.30**	**0.30**	**0.25**	**0.20**	**0.15**	**0.12**
DRAM storage node cell capacitor voltage (V) [4]	**0.45**	**0.4**	**0.4**	**0.35**	**0.35**	**0.35**	**0.35**
Equivalent Electric field of capacitor dielectric, (MV/cm) [5]	**15.0**	**13.3**	**13.3**	**14.0**	**17.5**	**23.3**	**29.2**
DRAM cell FET structure [6]	**FinFET**	**FinFET**	**FinFET**	**FinFET**	**FinFET**	**FinFET**	**FinFET**
DRAM cell FET dielectric: equivalent oxide thickness, EOT (nm) [7]	**3.5**	**3.5**	**3.5**	**3.5**	**3.5**	**3.5**	**3.5**
Maximum Wordline (WL) level (V) [8]	**2.3**	**2.3**	**2.3**	**2.0**	**2.0**	**2.0**	**2.0**
Negative Wordline (WL) use [9]	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
Equivalent Electric field of cell FET device dielectric (MV/cm) [10]	**6.57**	**6.57**	**6.57**	**5.71**	**5.71**	**5.71**	**5.71**
Cell Size Factor: a [11]	**6**	**6**	**6**	**6**	**6**	**6**	**6**
Array Area Efficiency [12]	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**
Minimum DRAM retention time (ms) [13]	**64**	**64**	**64**	**64**	**64**	**64**	**64**
DRAM soft error rate (fits) [14]	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**
Vint (support FET voltage) [V] [15]	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.7**	**0.7**
Support nMOS EOT [nm] [16]	**1.5**	**1.4**	**1.4**	**1.3**	**1.3**	**1.3**	**1.2**
Support PMOS Gate Electrode [17]	**TiN**	**TiN**	**TiN**	**TiN**	**TiN**	**TiN**	**TiN**
Support Gate Oxide [18]	**HfSiON**	**HfSiON**	**HfSiON**	**HfSiON**	**HfSiON**	**HfSiON**	**HfSiON**
Support min. L_{gate} for NMOS FET, physical [nm] [19]	**35**	**31**	**28**	**25**	**23**	**21**	**19**
Support $I_{sat\text{-}n}$ [$\mu A/\mu m$] (25C, V_g=V_d=V_{int}) [20]	**480**	**550**	**550**	**550**	**550**	**550**	**550**
Support min. V_{tn} (25C, $G_{m,max}$, V_d=55mV) [21]	**0.31**	**0.31**	**0.31**	**0.31**	**0.31**	**0.31**	**0.31**
Support $I_{sat\text{-}p}$ [$\mu A/\mu m$] (25C, V_g=V_d=-V_{int}) [22]	**215.00**	**215.00**	**215.00**	**215.00**	**215.00**	**215.00**	**215.00**
Support min. V_{tp} (25C, $G_{m,max}$, V_d=55mV) [23]	**-0.32**	**-0.32**	**-0.32**	**-0.32**	**-0.32**	**-0.32**	**-0.32**

Notes for Table PIDS4a and b:

[1] From ORTC (Overall Roadmap Technology Characteristics) Table 1a and b. These DRAM half pitch numbers are the same as those in the 2006 ITRS due to no further speed up in the pace of DRAM half pitch scaling during 2006.

[2] The DRAM cell size is driven by the values for DRAM capacity (bits per chip) and chip size, as discussed in more detail in the Front End Process chapter. The capacity and chip size numbers are based on the ORTC Tables 1a and 1b. Since the DRAM capacity and chip size numbers are quite aggressive, the cell size must also be scaled aggressively. The difficulty will lie in reducing the value of the cell size factor "a", where "a" equals (cell size /F^2) and F is the DRAM half pitch. The required values of "a" are 6 for 68 nm and beyond.

[3] Storage node cell dielectric EOT is defined as (dielectric physical thickness / [κ/3.9]), where κ is the relative dielectric constant of the storage node cell dielectric and 3.9 is the relative dielectric constant of thermal SiO_2. The value of EOT is driven by the values for DRAM capacity (bits per chip) and chip size, as discussed in more detail in the Front End Process chapter. The capacity and the chip size numbers used by FEP are from the ORTC Tables 1a and 1b. Since the values of DRAM capacity and chip size from FEP are quite aggressive, the EOT must also be scaled very aggressively. Up to 2009, the dielectric material is based on Al_2O_3 or HfO_2, and hence the color is white. Beyond 2009, breakthroughs such as MIM structure with higher κ insulator material of epsilon more than 40 and physical thickness of less than 9nm are needed, so the color is yellow. Finally, for 2012 and beyond, there are no known solutions with demonstrated credibility, and hence the color is red. The actual EOT required for each year also depends on other factors such as cell height and/or 3D structure, film leakage current and contact formation.

[4] The DRAM storage node capacitor voltage must be low enough that the resulting electric field in the dielectric (see Note [5]) is within acceptable limits.

[5] The equivalent electric field in the capacitor dielectric is (DRAM storage node capacitor voltage / DRAM storage node dielectric equivalent oxide thickness, EOT). The equivalent field is the electric field if the dielectric is silicon dioxide; if the dielectric is high-κ, the actual electric field is [equivalent field]/[κ/3.9]. Note the sharp increase in the equivalent field with scaling. The color turns yellow in 2009, when the field is 7.5 MV/cm, and red in 2012, when the field becomes 13.75 MV/cm..

[6] DRAM cell MOSFET structure migrates from RCAT (recessed channel array transistor) to FinFET. RCAT is a technology to improve retention time characteristics by introducing recessed channel structure. FinFET is used to increase the drive current in the limited cell FET area and also to improve retention time characteristics.

[7] DRAM cell FET dielectric EOT is defined as (dielectric physical thickness / [κ/3.9]), where κ is the relative dielectric constant of the DRAM cell FET dielectric and 3.9 is the relative dielectric constant of thermal SiO_2. The EOT values here are large, mainly because of the high word line voltage levels (see Note 8) and the need to keep the electric field in the dielectric within tolerable limits (see Note 9)

[8] Maximum word line level is the (highly boosted) gate voltage for cell FET devices. The high gate voltage is required to get enough device drive current with high threshold voltage due to back gate voltage at the operating condition.

[9] Negative word line is used to suppress sub-threshold leakage current of cell transistor even in the case of lower level of V_t value of cell FET. The low V_t is preferable to get higher drive current of cell FET.

[10] The equivalent electric field in the cell FET device dielectric is (maximum word line level / DRAM cell FET dielectric equivalent oxide thickness, EOT). The equivalent field is the electric field if the dielectric is silicon dioxide; if the dielectric is high-κ, the actual field is [equivalent field]/[κ/3.9].

[11] Cell size factor = a = (DRAM cell size/F^2), where F is the DRAM ½ pitch. The required values of a are 6 for 2007 and beyond. In contrast, the 2005 version of the DRAM table has a = 8 in 2005, 2006 and 2007.

[12] Array area efficiency is the ratio of cell array area to total chip area. Hence, array area efficiency = 1 / (1 + [peripheral circuit area]/NaF^2), where N is the DRAM capacity (number of bits per chip), F is the DRAM ½ pitch, and a is the cell size factor (see Note 9). For a = 8, array area efficiency is estimated to be 0.63, so when a is decreased to 6 from 2007, the array area efficiency of 0.56 is made in conjunction with a = 6, assuming the same relative peripheral circuit area.

[13] Retention time is defined at 85ºC, and is the minimum time during which the data from memory can still be sensed correctly without refreshing a row bit line. The 64 ms specified here is the value needed for PC applications. The retention time depends on the combined interaction of device leakage current, signal strength, and signal sensing circuit sensitivity, and also depends on operational frequency and temperature.

[14] This is a typical FIT rate and depends on cycle time and the quality of cell capacitor and sensing circuits.

[15] V_{int} is the nominal power supply voltage for DRAM support FET in peripheral circuit area. It has been chosen to maintain sufficient voltage over-drive in order to meet the required saturation current drive values while still maintaining reasonable vertical gate dielectric electric field strengths.

[16] DRAM support MOS FET dielectric EOT is defined as (dielectric physical thickness / [κ/3.9]), where κ is the relative dielectric constant of the DRAM cell FET dielectric and 3.9 is the relative dielectric constant of thermal SiO_2.

[17] Support PMOS FET Gate electrode material migrates from P+Poly/W to TiN.

[18] DRAM support MOS FET dielectric material migrates from SiON to HfSiON in order to leakage current to meet.

[19] Physical support min. L_{gate} for NMOS FET is the final, as-etched length of the bottom of the gate electrode.

[20] Support $I_{sat\text{-}n}$ (the saturation drive current for support NMOS FET) is defined as the NMOSFET drain current per micron device width with the gate bias and the drain bias set equal to V_{int} (see Note [15]) and the source and substrate biases set to zero at 25°C, namely $V_g=V_d=V_{int}$).

[21] Support min V_{tn} is the saturation threshold voltage measured at 25 ºC, $G_{m\ max}$,V_d=55mV.

[22] Support Isat-p (the saturation drive current for support PMOS FET) is defined as the PMOSFET drain current per micron device width with the gate bias and the drain bias set equal to -V_{int} (see Note [15]) and the source and substrate biases set to zero at 25°C ,namely, $V_g=V_d=-V_{int}$).

[23] Support min V_{tp} is the saturation threshold voltage measured at 25 ºC, $G_{m\ max}$,V_d=55mV.

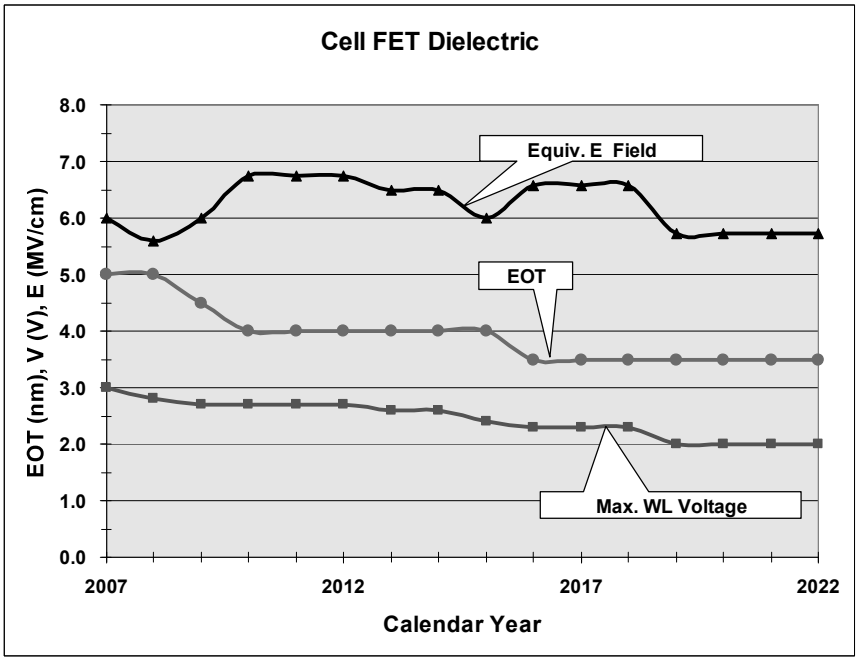

Figure PIDS6 Cell FET Devices

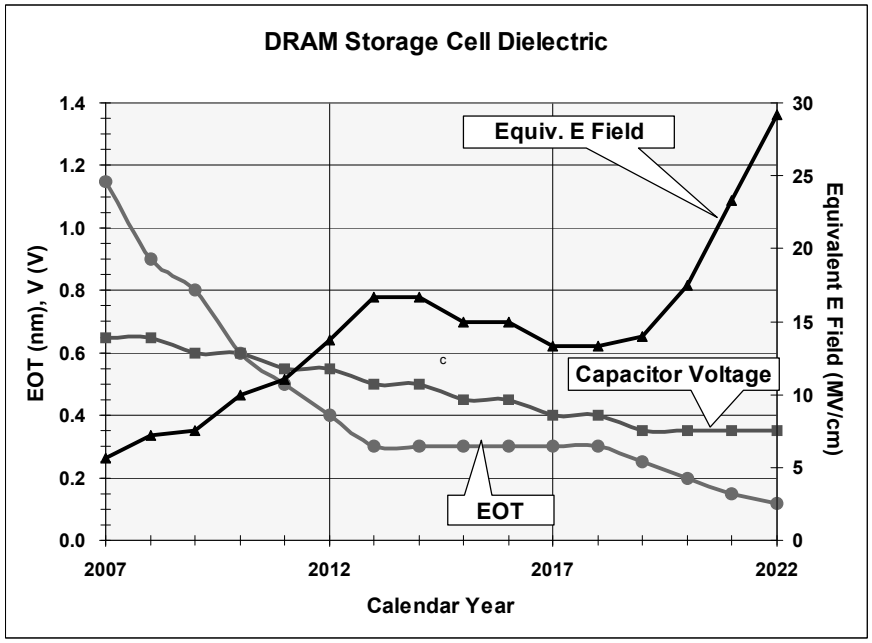

Figure PIDS7 Storage Node Capacitor

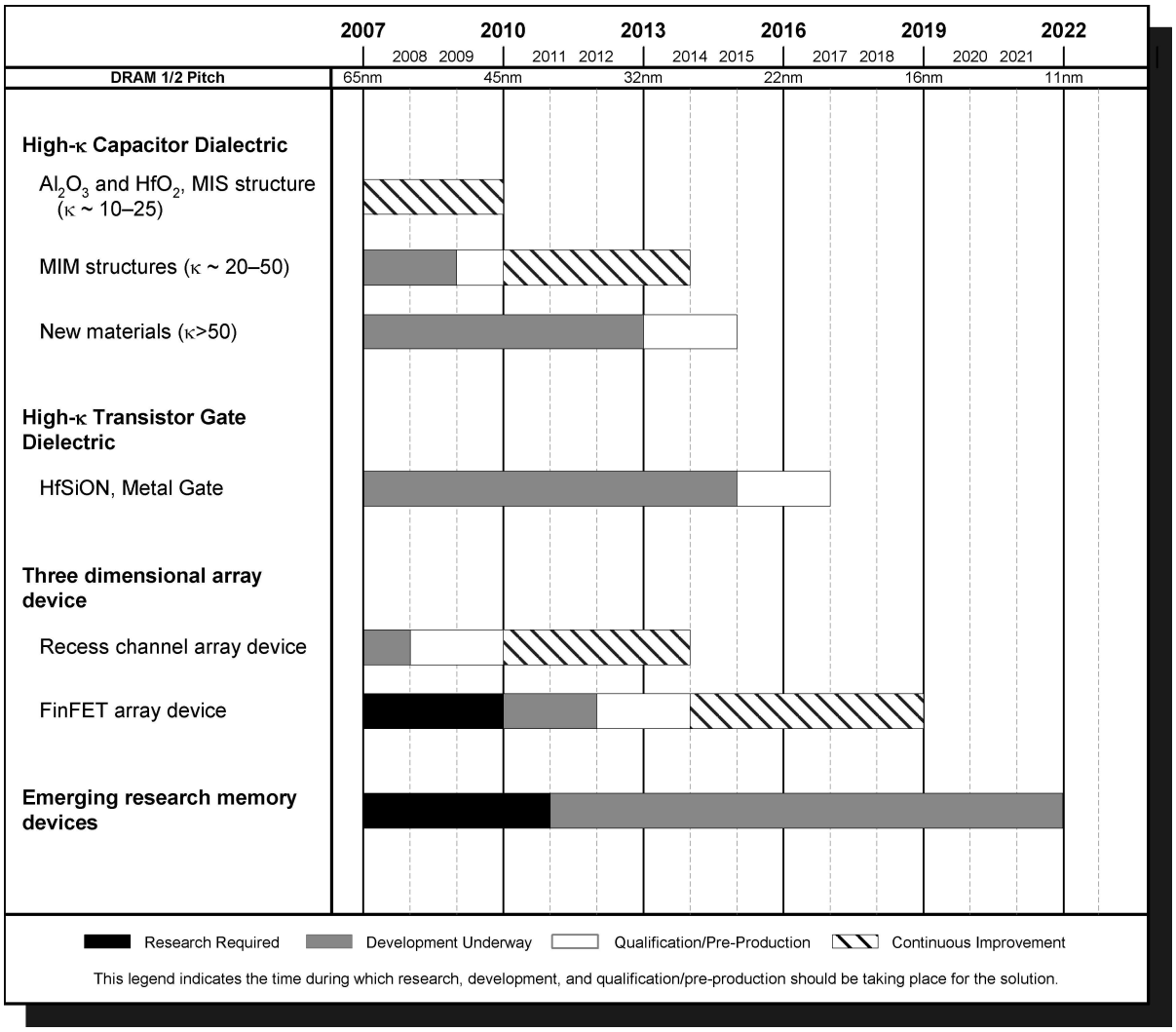

Figure PIDS8 DRAM Potential Solutions

NON-VOLATILE MEMORY

NON-VOLATILE MEMORY TECHNOLOGY REQUIREMENTS

Non-volatile memory consists of several intersecting technologies that share one common trait – non-volatility. The requirements and challenges differ according to the applications, ranging from RFIDs that only require KB of storage to high-density storage of tens of Gb in a chip. The requirements tables are divided into three categories—NAND Flash, NOR Flash, and non-charge-storage memories. Each category may contain more than one approach. For example, NOR Flash memories are fabricated using both floating gate device and nitride charge trapping device, each with their own design rules and scaling trend. Overlapping and/or succession of different technologies for the same application are indicated as appropriate, since as technology evolves best density and performance may be achieved through multiple paths.

Information on each technology is organized into three categories. The requirements tabulation for each technology first treats the issue of density. The applicable feature size "F" is identified and the expected area factor "a" is given (cell size in terms of the number of F^2 units required). Second, the tabulation presents a number of parameters important to each specific technology such as gate lengths, write-erase voltage maxima, key material parameters, etc. These parameters have significance because they are important to the scaling model and/or identify key challenge areas. Third, the endurance (erase-write cycle or read-write cycle) ratings and the retention ratings are presented. Endurance and retention

are requirements unique to NVM technologies and determine whether the device has adequate utility to be of interest to an end customer.

Table PIDS5a and b show technology requirements for NAND Flash, NOR Flash and non-charge-storage memories for near term years and long term years, respectively. The tables identify both the current CMOS half-pitch and the feature size actually used to form the NVM cells (i.e., the NVM technology "F" in nanometers). Until recently NVM half-pitches have lagged those for DRAM or CMOS logic devices in the same year. Rapid progress in NAND technology has recently reversed this trend, and some design rules are now more aggressively driven by NAND than DRAM. This trend has not spread to other NVM applications yet, however.

Table PIDS5a Non-volatile Memory Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ Pitch (nm) (contacted)	*68*	*58*	*50*	*45*	*40*	*36*	*32*	*30*	*25*
NAND Flash poly ½ Pitch (nm)	*51*	*45*	*40*	*36*	*32*	*28*	*25*	*22*	*20*
NAND Flash									
NAND Flash technology – F (nm) [1]	**51**	**45**	**40**	**36**	**32**	**28**	**25**	**22**	**20**
Number of word lines in one NAND string [2]	**32**	**32**	**64**	**64**	**64**	**64**	**64**	**64**	**64**
Cell type (FG, CT, 3D, etc.) [3]	**FG**	**FG**	**FG**	**FG/CT**	**CT**	**CT**	**CT-3D**	**CT-3D**	**CT-3D**
3D NAND number of memory layers	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**2**	**2**
A. Floating Gate NAND Flash									
Cell size – area factor a in multiples of F^2 SLC/MLC [4]	**4.0/2.0**	**4.0/2.0**	**4.0/1.3**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**
Tunnel oxide thickness (nm) [5]	**6-7**	**6-7**	**6-7**	**6-7**	**6-7**	**6-7**	**6-7**	**6-7**	**6-7**
Interpoly dielectric material [6]	**ONO**	**ONO**	**ONO**	**ONO**	**ONO**	**High-κ**	**High-κ**	**High-κ**	**High-κ**
Interpoly dielectric thickness (nm)	**10-13**	**10-13**	**10-13**	**10-13**	**10-13**	**9-10**	**9-10**	**9-10**	**9-10**
Gate coupling ratio (GCR) [7]	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**
Control gate material [8]	**n-Poly**	**n-Poly**	**n-Poly**	**n-Poly**	**n-Poly**	**Poly/metal**	**Poly/metal**	**Poly/metal**	**Metal**
Highest W/E voltage (V) [9]	**17-19**	**17-19**	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**
Endurance (erase/write cycles) [10]	**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**
Nonvolatile data retention (years) [11]	**10-20**	**10-20**	**10-20**	**10-20**	**10-20**	**10-20**	**10-20**	**10-20**	**10-20**
Maximum number of bits per cell (MLC) [12]	**2**	**2**	**3**	**4**	**4**	**4**	**4**	**4**	**4**
B. Charge trapping NAND Flash (MANOS or Barrier Engineering) [13]									
Cell size—area factor a in multiples of F^2 SLC/MLC				**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**
Tunnel dielectric material [14]				**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**
Tunnel dielectric thickness EOT (nm)				**3-4**	**3-4**	**3-4**	**3-4**	**3-4**	**3-4**
Blocking dielectric material [15]				**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**
Blocking dielectric thickness EOT (nm)				**6–8**	**6–8**	**6–8**	**6–8**	**6–8**	**6–8**
Charge trapping layer material [16]				**SiN**	**SiN**	**SiN**	**SiN**	**SiN**	**SiN**
Charge trapping layer thickness (nm) [17]				**5–7**	**5–7**	**5–7**	**5–7**	**5–7**	**4–6**
Gate material [18]				**p-Poly/Metal**	**p-Poly/Metal**	**p-Poly/Metal**	**p-Poly/Metal**	**p-Poly/Metal**	**Metal**
Highest W/E voltage (V)				**15-17**	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**
Endurance (erase/write cycles) [19]				**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**	**1.E+05**
Nonvolatile data retention (years) [20]				**10-20**	**10-20**	**10-20**	**10-20**	**10-20**	**10-20**
Maximum number of bits per cell (MLC)				**4**	**4**	**4**	**4**	**4**	**4**

Table PIDS5a Non-volatile Memory Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ Pitch (nm) (contacted)	*68*	*58*	*50*	*45*	*40*	*36*	*32*	*30*	*25*
NAND Flash poly ½ Pitch (nm)	*51*	*45*	*40*	*36*	*32*	*28*	*25*	*22*	*20*
NOR Flash									
NOR Flash technology – F (nm) [21]	**65**	**57**	**50**	**45**	**40**	**35**	**32**	**28**	**25**
A. Floating gate NOR Flash									
Cell size—area factor a in multiples of F^2 [22], [23], [24], [25]	**9-11**	**9-11**	**9-11**	**9-11**	**9-11**	**9-11**	**9-11**	**9-11**	**9-11**
Gate length L_g, physical (nm) [26]	**130**	**120**	**100**	**90**	**80**	**70**	**64**	**56**	**50**
Tunnel oxide thickness (nm) [27]	**8–9**	**8–9**	**8–9**	**8**	**8**	**8**	**8**	**7 - 8**	**7 - 8**
Interpoly dielectric material [28]	**ONO**	**ONO**	**ONO**	**ONO**	**ONO**	**ONO**	**High-κ**	**High-κ**	**High-κ**
Interpoly dielectric thickness EOT (nm)	**13-15**	**13-15**	**13-15**	**13-15**	**13-15**	**13-15**	**8-10**	**8-10**	**8-10**
Gate coupling ratio [29]	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**
Highest W/E voltage (V) [30]	**7-9**	**7-9**	**7-9**	**7-9**	**7-9**	**7-9**	**6-8**	**6-8**	**6-8**
I_{read} (μA) [31]	**25-34**	**23-31**	**21-27**	**20-26**	**19-25**	**17-22**	**15-20**	**14-19**	**13-18**
Endurance (erase/write cycles) [32]	**1.00E+05**	**1.00E+05**	**1.00E+05**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**
Nonvolatile data retention (years) [33]	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**20**	**20**	**20**
Maximum number of bits per cell (MLC) [34]	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Array architecture (with cell contact (CC) or virtual ground (VG))[35]	**CC**	**CC**	**CC**	**CC**	**CC**	**CC**	**CC/VG**	**CC/VG**	**CC/VG**
B. Charge trapping NOR Flash (SONOS/NROM) [36]									
SONOS/NROM technology, F (nm)	**65**	**57**	**50**	**45**	**40**	**35**	**32**	**28**	**25**
SONOS/NROM cell size - area factor a in multiples of F^2	**6-7**	**6-7**	**6-7**	**6-7**	**7-8**	**7-8**	**7-8**	**7-8**	**8-9**
Cell size (per bit) – area factor a in multiples of F^2 (SLC/MLC) [37]	**3.3/1.6**	**3.3/1.6**	**3.3/1.6**	**3.3/1.6**	**3.7/1.9**	**3.7/1.9**	**3.7/1.9**	**3.7/1.9**	**4.3/2.2**
Gate length L_g, physical (nm) [38]	**140**	**130**	**120**	**110**	**110**	**100**	**100**	**90**	**90**
Tunnel oxide thickness (nm) [39]	**5**	**5**	**5**	**4.5**	**4.5**	**4.5**	**4**	**4**	**4**
Charge trapping layer thickness (nm) [40]	**5-7**	**5-7**	**5-7**	**4-6**	**4-6**	**4-6**	**4-6**	**4-6**	**4-5**
Blocking (top) dielectric thickness EOT (nm) [41]	**7–9**	**7–9**	**7–9**	**6–8**	**6–8**	**6–8**	**6–8**	**6–8**	**5–7**
Highest W/E voltage (V)	**7-9**	**7-9**	**7-9**	**6-8**	**6-8**	**6-8**	**6-8**	**5–7**	**5–7**
I_{read} (μA) [31]	**25-34**	**23-31**	**21-27**	**20-26**	**19-25**	**17-22**	**15-20**	**14-19**	**13-18**
Endurance (erase/write cycles) [32]	**1.00E+05**	**1.00E+05**	**1.00E+05**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**
Nonvolatile data retention (years) [33]	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**
Maximum number of bits per cell (physical 2-bit/cell + MLC) [37]	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**6**
Non-charge-storage NVM									
A. FeRAM (Ferroelectric RAM)									
FeRAM technology – F (nm) [42]	**180**	**180**	**180**	**130**	**130**	**130**	**90**	**90**	**90**
FeRAM cell size – area factor a in multiples of F^2 [43]	**22**	**22**	**22**	**20**	**20**	**20**	**16**	**16**	**16**
FeRAM cell size (μm^2)	**0.713**	**0.713**	**0.713**	**0.450**	**0.450**	**0.450**	**0.270**	**0.270**	**0.270**
FeRAM cell structure [44]	**2T2C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**

Table PIDS5a Non-volatile Memory Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ Pitch (nm) (contacted)	*68*	*58*	*50*	*45*	*40*	*36*	*32*	*30*	*25*
NAND Flash poly ½ Pitch (nm)	*51*	*45*	*40*	*36*	*32*	*28*	*25*	*22*	*20*
FeRAM capacitor structure [45]	**stack**	**stack**	**stack**	**stack**	**stack**	**stack**	**stack**	**stack**	**stack**
FeRAM capacitor footprint (μm^2) [46]	**0.330**	**0.330**	**0.330**	**0.200**	**0.200**	**0.200**	**0.106**	**0.106**	**0.106**
FeRAM capacitor active area (μm^2) [47]	**0.330**	**0.330**	**0.330**	**0.200**	**0.200**	**0.200**	**0.106**	**0.106**	**0.106**
FeRAM cap active area/footprint ratio [48]	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Ferro capacitor voltage (V) [49]	**1.50**	**1.50**	**1.50**	**1.20**	**1.20**	**1.20**	**1.20**	**1.20**	**1.20**
FeRAM minimum switching charge density ($\mu C/cm^2$) [50]	**13.5**	**13.5**	**13.5**	**20**	**20**	**20**	**34**	**34**	**34**
FeRAM endurance (read/write cycles) [51]	**1.0E+14**	**1E+14**	**1E+14**	**1E+14**	**1E+14**	**1E+14**	**1E+15**	**1E+15**	**1E+15**
FeRAM nonvolatile data retention (years)	**10 Years**	**10 Years**	**10 Years**	**10 Years**	**10 Years**	**10 Years**	**10 Years**	**10 Years**	**10 Years**
B. MRAM (Magnetic RAM)									
MRAM technology F (nm) [52]	**90**	**65**	**65**	**45**	**45**	**45**	**32**	**32**	**32**
MRAM cell size area factor a in multiples of F^2	**20**	**22**	**19**	**20**	**18**	**18**	**19**	**17**	**17**
MRAM typical cell size (μm^2)	**0.16**	**0.09**	**0.08**	**0.041**	**0.036**	**0.036**	**0.019**	**0.017**	**0.017**
MRAM switching field (Oe) [53]	**35**	**35**	**35**	**35**	**35**	**35**	**35**	**35**	**35**
MRAM write energy (pJ/bit) [54]	**70**	**35**	**35**	**25**	**25**	**25**	**20**	**20**	**20**
MRAM active area per cell (μm^2) [55]	**0.05**	**0.025**	**0.025**	**0.013**	**0.013**	**0.013**	**0.009**	**0.009**	**0.009**
MRAM resistance-area product (KOhm-(μm^2) [56]	**2**	**1.1**	**1**	**0.8**	**0.8**	**0.8**	**0.6**	**0.6**	**0.6**
MRAM magnetoresistance ratio (%) [57]	**70**	**70**	**70**	**70**	**70**	**70**	**70**	**70**	**70**
MRAM nonvolatile data retention (years)	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**
MRAM write endurance (read/write cycles)	**>3e16**	**>3e16**	**>3e16**	**>3e16**	**>3e16**	**>3e16**	**>3e16**	**>3e16**	**>3e16**
MRAM endurance – tunnel junction reliability (years at bias) [58]	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**
C. PCRAM (Phase-Change RAM)									
PCRAM technology F (nm) [58]	**72**	**58**	**46**	**40**	**35**	**32**	**28**	**25**	**22**
PCRAM cell size area factor a in multiples of F^2 (BJT access device) [59]	**4.8**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**
PCRAM cell size area factor a in multiples of F^2 (nMOSFET access device) [60]	**15.0**	**14.0**	**12.0**	**11.0**	**10.0**	**8.9**	**8.8**	**8.4**	**7.4**
PCRAM typical cell size (nm^2) (BJT access device) [61]	**24883**	**13456**	**8464**	**6400**	**4900**	**4096**	**3136**	**2500**	**1936**
PCRAM typical cell size (nm^2) (nMOSFET access device) [62]	**77760**	**47096**	**25392**	**17600**	**12250**	**9114**	**6899**	**5250**	**3582**
PCRAM number of bits per cell (MLC) [63]	**1**	**1**	**2**	**2**	**2**	**4**	**4**	**4**	**4**
PCRAM typical cell area per bit size (μm^2) (BJT access device) [64]	**24883**	**13456**	**4232**	**3200**	**2450**	**1024**	**784**	**625**	**484**
PCRAM typical cell area per bit size (μm^2) (nMOSFET access device) [65]	**77760**	**47096**	**12696**	**8800**	**6125**	**2278**	**1725**	**1313**	**895**
PCRAM storage element CD (nm) [66]	**45**	**36**	**30**	**25**	**22**	**20**	**18**	**16**	**14**
PCRAM phase change volume (nm^3) [67]	**373,000**	**195,000**	**112,000**	**64,000**	**43,000**	**33,000**	**25,000**	**18,000**	**12,000**

Table PIDS5a Non-volatile Memory Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ Pitch (nm) (contacted)	*68*	*58*	*50*	*45*	*40*	*36*	*32*	*30*	*25*
NAND Flash poly ½ Pitch (nm)	*51*	*45*	*40*	*36*	*32*	*28*	*25*	*22*	*20*
PCRAM reset current (µA) [68]	**235**	**170**	**130**	**100**	**80**	**70**	**62**	**52**	**43**
PCRAM set resistance (KOhm) [69]	**3.54**	**4.57**	**5.68**	**7.08**	**8.29**	**9.21**	**10.20**	**11.66**	**13.56**
PCRAM BJT current density (A/cm^2) [70]	**1.50E+07**	**1.50E+07**	**1.50E+07**	**1.50E+07**	**1.50E+07**	**1.50E+07**	**1.50E+07**	**1.60E+07**	**1.70E+07**
PCRAM BJT emitter area (nm^2) [71]	**4072**	**2642**	**1662**	**1257**	**962**	**804**	**616**	**491**	**380**
PCRAM nMOSFET apparent current density for reset (µA/nm) [72]	**1.5**	**1.5**	**1.8**	**1.8**	**1.8**	**2.1**	**2.1**	**2.1**	**2.4**
PCRAM nMOSFET apparent device width (nm) [73]	**239**	**171**	**108**	**82**	**68**	**51**	**43**	**36**	**26**
PCRAM nonvolatile data retention (years) [74]	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**
PCRAM write endurance (read/write cycles) [75]	**1.0E+08**	**1.0E+08**	**1.0E+10**	**1.0E+10**	**1.0E+10**	**1.0E+12**	**1.0E+12**	**1.0E+12**	**1.0E+15**

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

Table PIDS5b Non-volatile Memory Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
DRAM ½ Pitch (nm) (contacted)	*22*	*20*					
NAND Flash poly ½ Pitch (nm)	*19*	*18*	*16*	*14*	*13*	*11*	*10*
NAND Flash							
NAND Flash technology – F (nm) [1]	**19**	**18**	**16**	**14**	**13**	**11**	**10**
Number of word lines in one NAND string [2]	**64**	**64**	**64**	**64**	**64**	**64**	**64**
Cell type (FG, CT, 3D, etc.) [3]	**CT-3D**	**CT-3D**	**CT-3D**	**CT-3D**	**CT-3D**	**CT-3D**	**CT-3D**
3D NAND number of memory layers	**2**	**2**	**4**	**4**	**4**	**4**	**4**
A. Floating Gate NAND Flash							
Cell size – area factor a in multiples of F^2 SLC/MLC [4]	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**
Tunnel oxide thickness (nm) [5]	**4**	**4**	**4**	**4**	**4**	**4**	**4**
Interpoly dielectric material [6]	**High-κ**	**High-κ**	**High-κ**	**High-κ**	**High-κ**	**High-κ**	**High-κ**
Interpoly dielectric thickness (nm)	**9-10**	**9-10**	**9-10**	**9-10**	**9-10**	**9-10**	**9-10**
Gate coupling ratio (GCR) [7]	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6-0.7**	**0.6-0.7**	**0.6-0.7**	**0.6-0.7**
Control gate material [8]	**Metal**	**Metal**	**Metal**	**Metal**	**Metal**	**Metal**	**Metal**
Highest W/E voltage (V) [9]	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**
Endurance (erase/write cycles) [10]	**1.E+04**	**1.E+04**	**1.E+04**	**1.E+04**	**1.E+04**	**1.E+04**	**1.E+04**
Nonvolatile data retention (years) [11]	**5-10**	**5-10**	**5-10**	**5-10**	**5-10**	**5-10**	**5-10**
Maximum number of bits per cell (MLC) [12]	**4**	**4**	**4**	**4**	**4**	**4**	**4**
B. Charge trapping NAND Flash (MANOS or Barrier Engineering) [13]							
Cell size – area factor a in multiples of F^2 SLC/MLC	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**	**4.0/1.0**
Tunnel dielectric material [14]	**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**	**SiO_2 or ONO**
Tunnel dielectric thickness EOT (nm)	**3-4**	**3-4**	**3-4**	**3-4**	**3-4**	**3-4**	**3-4**
Blocking dielectric material [15]	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**	**SiO_2 or Al_2O_3**
Blocking dielectric thickness EOT (nm)	**6–8**	**6–8**	**6–8**	**6–8**	**6–8**	**6–8**	**6–8**
Charge trapping layer material [16]	**SiN / High-κ**	**SiN / High-κ**	**SiN / High-κ**	**SiN / High-κ**	**SiN / High-κ**	**SiN / High-κ**	**SiN / High-κ**
Charge trapping layer thickness (nm) [17]	**4–6**	**4–6**	**4–6**	**4–6**	**4–6**	**3-4**	**3-4**
Gate material [18]	**Metal**	**Metal**	**Metal**	**Metal**	**Metal**	**Metal**	**Metal**
Highest W/E voltage (V)	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**	**15-17**
Endurance (erase/write cycles) [19]	**1.E+04**	**1.E+04**	**1.E+04**	**1.E+04**	**1.E+04**	**1.E+04**	**1.E+04**
Nonvolatile data retention (years) [20]	**5-10**	**5-10**	**5-10**	**5-10**	**5-10**	**5-10**	**5-10**
Maximum number of bits per cell (MLC)	**4**	**4**	**4**	**4**	**4**	**4**	**4**

Table PIDS5b Non-volatile Memory Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
DRAM ½ Pitch (nm) (contacted)	*22*	*20*					
NAND Flash poly ½ Pitch (nm)	*19*	*18*	*16*	*14*	*13*	*11*	*10*
NOR Flash							
NOR Flash technology – F (nm) [21]	**22**	**20**	**18**	**16**	**14**	**12**	**10**
A. Floating gate NOR Flash							
Cell size – area factor a in multiples of F^2 [22], [23], [24], [25]	**10-13**	**10-13**	**10-13**	**10-13**	**10-13**	**10-13**	**10-13**
Gate length L_g, physical (nm) [26]	**44**	**40**	**36**	**32**	**28**	**24**	**20**
Tunnel oxide thickness (nm) [27]	**7 - 8**	**7 - 8**	**7 - 8**	**7 - 8**	**7 - 8**	**7 - 8**	**7 - 8**
Interpoly dielectric material [28]	**High-κ**	**High-κ**	**High-κ**	**High-κ**	**High-κ**	**High-κ**	**High-κ**
Interpoly dielectric thickness EOT (nm)	**8-10**	**8-10**	**7-9**	**6-8**	**6-8**	**6-8**	**6-8**
Gate coupling ratio [29]	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6-0.7**	**0.6-0.7**	**0.6-0.7**	**0.6-0.7**
Highest W/E voltage (V) [30]	**6-8**	**6-8**	**6-8**	**6-8**	**6-8**	**6-8**	**6-8**
I_{read} (μA) [31]	**12–17**	**11–16**	**10–15**	**9-14**	**8-13**	**7-12**	**6-10**
Endurance (erase/write cycles) [32]	**1.00E+07**	**1.00E+07**	**1.00E+07**	**1.00E+07**	**1.00E+07**	**1.00E+07**	**1.00E+07**
Nonvolatile data retention (years) [33]	**20**	**20**	**20**	**20**	**20**	**20**	**20**
Maximum number of bits per cell (MLC) [34]	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Array architecture (with cell contact (CC) or virtual ground (VG))[35]	**CC/VG**	**CC/VG**	**CC/VG**	**CC/VG**	**CC/VG**	**CC/VG**	**CC/VG**
B. Charge trapping NOR Flash (SONOS/NROM) [36]							
SONOS/NROM technology, F (nm)	**22**	**20**	**18**	**16**	**14**	**12**	**10**
SONOS/NROM cell size - area factor a in multiples of F^2	**8-9**	**8-9**	**8-9**	**9-10**	**9-10**	**9-10**	**9-10**
Cell size (per bit) – area factor a in multiples of F^2 (SLC/MLC) [37]	**4.3/2.2**	**4.3/2.2**	**4.3/2.2**	**4.8/2.4**	**4.8/2.4**	**4.8/2.4**	**4.8/2.4**
Gate length L_g, physical (nm) [38]	**80**	**80**	**70**	**70**	**70**	**60**	**60**
Tunnel oxide thickness (nm) [39]	**4**	**4**	**4**	**3.5**	**3.5**	**3.5**	**3.5**
Charge trapping layer thickness (nm) [40]	**4-5**	**4-5**	**4-5**	**4**	**4**	**4**	**4**
Blocking (top) dielectric thickness EOT (nm) [41]	**5–7**	**5–7**	**5–7**	**5–7**	**5–7**	**5–7**	**5–7**
Highest W/E voltage (V)	**5–7**	**5–7**	**5–7**	**5–7**	**5–7**	**4-6**	**4-6**
I_{read} (μA) [31]	**12–17**	**11–16**	**10–15**	**9-14**	**8-13**	**7-12**	**6-10**
Endurance (erase/write cycles) [32]	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**
Nonvolatile data retention (years) [33]	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**
Maximum number of bits per cell (physical 2-bit/cell + MLC) [37]	**6**	**6**	**6**	**6**	**6**	**6**	**6**

Table PIDS5b Non-volatile Memory Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
DRAM ½ Pitch (nm) (contacted)	*22*	*20*					
NAND Flash poly ½ Pitch (nm)	*19*	*18*	*16*	*14*	*13*	*11*	*10*
Non-charge-storage NVM							
A. FeRAM (Ferroelectric RAM)							
FeRAM technology – F (nm) [42]	**90**	**90**	**90**	**65**	**65**	**65**	**65**
FeRAM cell size – area factor a in multiples of F^2 [43]	**14**	**14**	**14**	**12**	**12**	**12**	**12**
FeRAM cell size (μm^2)	**0.113**	**0.113**	**0.113**	**0.051**	**0.051**	**0.051**	**0.051**
FeRAM cell structure [44]	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**
FeRAM capacitor structure [45]	**3D**	**3D**	**3D**	**3D**	**3D**	**3D**	**3D**
FeRAM capacitor footprint (μm^2) [46]	**0.041**	**0.041**	**0.041**	**0.016**	**0.016**	**0.016**	**0.016**
FeRAM capacitor active area (μm^2) [47]	**0.100**	**0.100**	**0.100**	**0.069**	**0.069**	**0.069**	**0.069**
FeRAM cap active area/footprint ratio [48]	**2.46**	**2.46**	**2.46**	**4.25**	**4.25**	**4.25**	**4.25**
Ferro capacitor voltage (V) [49]	**1.00**	**1.00**	**1.00**	**0.70**	**0.70**	**0.70**	**0.70**
FeRAM minimum switching charge density ($\mu C/cm^2$) [50]	**30**	**30**	**30**	**30**	**30**	**30**	**30**
FeRAM endurance (read/write cycles) [51]	**>1.0E16**	**>1.0E16**	**>1.0E16**	**>1.0E16**	**>1.0E16**	**>1.0E16**	**>1.0E16**
FeRAM nonvolatile data retention (years)	**10 Years**	**10 Years**	**10 Years**	**10 Years**	**10 Years**	**10 Years**	**10 Years**
B. MRAM (Magnetic RAM)							
MRAM technology F (nm) [52]	**22**	**22**	**22**	**16**	**16**	**16**	**16**
MRAM cell size area factor a in multiples of F^2	**18**	**16**	**16**	**17**	**16**	**17**	**16**
MRAM typical cell size (μm^2)	**0.009**	**0.0077**	**0.0077**	**0.0044**	**0.0041**	**0.0044**	**0.0041**
MRAM switching field (Oe) [53]	**35**	**35**	**35**	**35**	**35**	**35**	**35**
MRAM write energy (pJ/bit) [54]	**20**	**20**	**20**	**20**	**20**	**20**	**20**
MRAM active area per cell (μm^2) [55]	**0.007**	**0.007**	**0.007**	**0.005**	**0.005**	**0.005**	**0.005**
MRAM resistance-area product (KOhm-(μm^2) [56]	**0.6**	**0.6**	**0.6**	**0.6**	**0.6**	**0.6**	**0.6**
MRAM magnetoresistance ratio (%) [57]	**70**	**70**	**70**	**70**	**70**	**70**	**70**
MRAM nonvolatile data retention (years)	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**
MRAM write endurance (read/write cycles)	**>3e16**	**>3e16**	**>3e16**	**>3e16**	**>3e16**	**>3e16**	**>3e16**
MRAM endurance – tunnel junction reliability (years at bias) [58]	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**

Table PIDS5b Non-volatile Memory Technology Requirements—Long-term Years

Year of Production	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*
DRAM ½ Pitch (nm) (contacted)	*22*	*20*					
NAND Flash poly ½ Pitch (nm)	*19*	*18*	*16*	*14*	*13*	*11*	*10*
C. PCRAM (Phase-Change RAM)							
PCRAM technology F (nm) [58]	**20**	**18**	**16**	**14**	**12**	**11**	**9**
PCRAM cell size area factor a in multiples of F^2 (BJT access device) [59]	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**	**4.0**
PCRAM cell size area factor a in multiples of F^2 (nMOSFET access device) [60]	**7.3**	**7.3**	**6.0**	**6.0**	**6.0**	**5.5**	**5.5**
PCRAM typical cell size (nm^2) (BJT access device) [61]	**1600**	**1296**	**1024**	**784**	**576**	**480**	**340**
PCRAM typical cell size (nm^2) (nMOSFET access device) [62]	**2920**	**2365**	**1536**	**1176**	**864**	**650**	**450**
PCRAM number of bits per cell (MLC) [63]	**4**	**4**	**4**	**4**	**4**	**4**	**4**
PCRAM typical cell area per bit size (μm^2) (BJT access device) [64]	**400**	**324**	**256**	**196**	**144**	**120**	**85**
PCRAM typical cell area per bit size (μm^2) (nMOSFET access device) [65]	**730**	**591**	**384**	**294**	**216**	**162**	**112**
PCRAM storage element CD (nm) [66]	**13**	**12**	**10**	**9**	**8**	**8**	**7**
PCRAM phase change volume (nm^3) [67]	**9,000**	**6,700**	**4,700**	**3,200**	**2,000**	**1,300**	**900**
PCRAM reset current (µA) [68]	**37**	**32**	**27**	**22**	**18**	**15**	**13**
PCRAM set resistance (KOhm) [69]	**15.17**	**17.18**	**19.74**	**23.11**	**27.72**	**31.00**	**35.00**
PCRAM BJT current density (A/cm^2) [70]	**1.90E+07**	**2.00E+07**	**2.10E+07**	**2.20E+07**	**2.40E+07**	**2.50E+07**	**2.70E+07**
PCRAM BJT emitter area (nm^2) [71]	**314**	**254**	**201**	**154**	**113**	**91**	**73**
PCRAM nMOSFET apparent current density for reset (µA/nm) [72]	**2.4**	**2.4**	**2.88**	**2.88**	**2.88**	**2.88**	**2.88**
PCRAM nMOSFET apparent device width (nm) [73]	**23**	**21**	**16**	**14**	**12**	**10**	**9**
PCRAM nonvolatile data retention (years) [74]	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**
PCRAM write endurance (read/write cycles) [75]	**1.0E+15**	**1.0E+15**	**1.0E+15**	**1.0E+15**	**1.0E+15**	**1.0E+15**	**1.0E+15**

Notes for Table PIDS5a and b:

[1] NAND Flash has surpassed CMOS and DRAM technology since 2005. This entry provides the F value for designs in the indicated time period.

[2] NAND Flash architecture consists of bit line strings of a number of storage devices. Long bit line strings reduce the overhead for bit line transistors and increase the packing density, however, at the expense of higher overall resistance and consequently lower read current. The number of word lines in a bit line string has increased from 16 to 32 nm.

[3] Because of the difficulty in maintaining high gate coupling ratio and preventing cross talk between neighboring cells, NAND technology is forecasted to migrate gradually from floating gate devices (FG) to charge trapping devices (CT). (Ref: K. Kim, "Technology for sub 50nm DRAM and NAND Flash manufacturing," in Tech. Digest International Electron Devices Meeting, pp. 539-543, 2005.) The statistical fluctuation limit of storing too few electrons imposes new challenges to data retention (Ref: G. Molas, et al., "Impact of few electron phenomena on floating gate memory reliability," Tech. Digest 2004 International Electron Devices Meeting, pp. 877-880, 2004.) and 3D integration of multiple layers of devices may be required to continue the scaling. (Refs: S.H. Lee et al., "Three dimensionally stacked NAND Flash memory technology using stacking single crystal Si layers on ILD and TANOS structure for beyond 30nm node," Tech. Digest 2006 International Electron Devices Meeting, pp. 37-40, 2006. E-K. Lai, et al., "A multi-layer stackable thin-film transistor (TFT) NAND-type Flash memory," Tech. Digest 2006 International Electron Devices Meeting, pp. 41-44, 2006.)

[4] The area factor "a" = cell area per bit/F^2, so this entry presents the expected range for Flash cell area in multiples of the implementation technology F^2. It is possible to store more than 1 bit of information in a Flash cell but increasing the logic levels from (1, 0) to (11, 10, 00, 01), etc. by using multi-level cell (MLC). Therefore, the area factor includes both single-level cell (SLC) and multi-level cell (MLC) devices.

[5] The scaling of tunnel oxide for NAND Flash faces the same challenge as that for NOR Flash. However, the use of error code correction (ECC) in NAND tolerates tunnel oxide defects to a higher level than that for NOR, thus allowing tunnel oxide of 6–7 nm. Currently there are no known solutions to scale tunnel oxide significantly below 6 nm.

[6] ONO has been used as the interpoly dielectric (IPD) up to now and will continue in the near future. However, below 40 nm the spacing between floating gates becomes too narrow to fill effectively with ONO and word line polysilicon, and the loss of sidewall control gate to floating gate coupling will severely degrade the gate coupling ratio (GCR) and the device becomes inoperable. Since it is impossible to create additional space, higher dielectric constant IPD or charge trapping (CT) device must be used. Here the path of migrating to high-κ IPD is shown. It is shown in red color since engineering solutions have not been demonstrated yet.

[7] Gate coupling ratio (GCR) is defined as (control gate to floating gate capacitance)/(total floating gate capacitance). GCR represents the fraction of voltage drop across the tunnel oxide and must be higher than 0.6 for the device to function during write and erase operations. High GCR is normally achieved by wrapping the control gate around the sidewalls of the floating gate. This requires tall floating gate and the cross talk along the bit line direction with neighboring cells is a challenge for MLC operation. At below 40 nm , the spacing between floating gates may become too narrow for the IPD and control gate to wrap around and maintaining sufficiently high GCR is a difficult challenge. Planar device with high-K IPD is a potential solution.

[8] n-type polysilicon (and polycide) gate has been used for the control gate so far and will continue in the near future. The introduction of high-κ IPD, with a lower barrier height to Si, will cause severe gate injection during erase operation and high work function material such as p-type polysilicon or metal gate may have to be adopted.

[9] Low write and erase voltage is desirable but EOT for tunnel oxide and IPD must be decreased to allow lower W/E voltage without compromising W/E speed.

[10] Write and erase cycling endurance reflects the tunnel oxide damage caused by repeated passing of charges under high electric field. Scaling does not worsen the oxide damage, however, larger array strains the ECC capability and the tolerance for defects is thus reduced. High-κ IPD may also trap charge and cause degradation. Current projection is gradually reduced cycling endurance for future technology. Note that this is not suitable for certain applications, e.g., solid-state drive storage that requires high cycling endurance.

[11] Data retention is controlled by both tunnel oxide integrity and statistical distribution of the number of stored electrons. Both thinner tunnel oxide and fewer number of stored electrons in the long term contribute to the shorter retention forecast.

[12] Multi-level cell (MLC) with 4 logic levels (2-bit/cell) is commonly used for NAND Flash today and devices with 8 logic levels (3-bit/cell) and 16 logic levels (4-bit/cell) are being developed. 8-bit/cell MLC device requires 256 logic levels and so far seems beyond reach of current technology, thus no forecast is made for 8-bit/cell MLC even in the long-term years.

[13] MANOS (Metal-Al2O3-Nitride-Oxide-Si) devices use a relatively thin tunnel oxide, a SiN trapping layer for charge storage, Al_2O_3 to increase voltage drop across the tunnel oxide and high work function metal gate to stop gate injection. (Ref: C.H. Lee, et al., "A novel SONOS structure of $SiO_2/SiN/Al_2O_3$ with TaN metal gate for multi-giga bit Flash memories," Tech. Digest 2003 International Electron Devices Meeting, pp. 613-616, 2003.) Barrier engineering uses composite tunneling barriers to allow easy erase operation by substrate hole tunneling, yet prevents low field hole direct tunneling during retention. (Ref: H.T. Lue, et al., "BE-SONOS: a bandgap engineered SONOS with excellent performance and reliability," Tech. Digest 2005 International Electron Devices Meeting, pp. 555-558, 2005.) Without floating gate GCR and cross talk issues, these two CT approaches promise to help NAND scaling below 30 nm.

[14] Tunnel dielectric for MANOS type device is a relatively thin silicon oxide (3-4nm). Tunnel dielectric for barrier engineered (BE) device may be a composite ONO (e.g., 2 nm/2 nm/2 nm) or other composites such as OAO.

[15] For MANOS device Al_2O_3 is the preferred blocking oxide since it has a high barrier height. Other composite blocking layers such as AN, ANO, or AHO are also potential candidates. For BE device the blocking oxide may be SiO_2, although Al_2O_3 and other composite layers may also be used.

[16] SiN is the most common and best known charge trapping layer with relatively deep electron traps that provide good data retention. Other exotic high-κ material with even deeper traps may be used in the long term years. (Ref: A. Chin, et al., "Low voltage high speed SiO2/AlGaN/AlLaO3/TaN memory with good retention," Tech. Digest 2005 International Electron Devices Meeting, pp. 165-168, 2005.) Note that for CT device the charge loss mechanism may be mainly substrate hole tunneling in low field and thus deeper traps do not necessarily improve data retention except for high temperature applications.

[17] Trapping efficiency in SiN seems thickness dependent. (Ref: H.T. Lue, et al., Proc. 2007 International Reliability Physics Symposium, 2007). Therefore, thinner SiN or other high-κ trapping layers forecasted for long-term years may suffer from reduced programming efficiency.

[18] High work function metal gate is the best to prevent gate injection. However, p-type polysilicon may become an interim solution because of its easy processing, low cost, and reasonably good performance.

[19] The mechanism for endurance degradation for CT devices is not well understood yet. Unlike floating gate device, CT devices are not sensitive to tunnel oxide damage since the charge is stored in discrete traps and one weak spot does not cause all stored charge to leak out as in floating gate devices. Published endurance data seem to indicate similar endurance as floating gate device.

[20] The mechanism for data retention loss for BE device seems understood. (Ref: H.T. Lue, et al., "Reliability model of bandgap engineered SONOS (BE-SONOS)," Tech. Digest 2006 International Electron Devices Meeting, pp. 495-498, 2006.) Data retention mechanism for MANOS device is still not well understood. Published data suggest that data retention is comparable to floating gate devices under certain conditions.

[21] NOR Flash traditionally falls behind CMOS and DRAM but has caught up in recent years and is now on par with DRAM.

[22] High-κ interpoly dielectric is projected at 32 nm and beyond to achieve gate coupling ratio of >0.6 but this offers only limited help to the area factor. (Ref: 2005 Symposium on VLSI Technology, 11B-3, E. S. Cho, et al.," Hf-silicate Inter-Poly Dielectric Technology for sub 70 nm Body Tied FinFET Flash Memory," pp. 208-209.)

[23] Although virtual ground (VG) array may significantly decrease the cell size in the near term years (Ref: 2005 Symposium on VLSI Technology, 11B-1, R. Koval, et al. "Flash ETOX Virtual Ground Architecture: A Future Scaling Direction," pp. 204-205.), this effect has not been included in the current table because VG is radically different from the conventional array and large development effort is needed to implement it and so far there is no industrial consensus to take up this endeavor.

[24] Although non-planar devices (such as FinFET) are being developed for future Flash scaling, their impact has not been included in the current table. The dilemma of filling ONO and control and floating gates in the narrow gap between adjacent vertical devices has not been completely resolved yet.

[25] Both the cell size and the gate length for NOR Flash have been more aggressively scaled recently with an area factor of approximately 10F2. (Refs: 2005 ISSCC, "A 90 nm 512Mb 166 MHz Multilevel Cell Flash Memory with 1.5MBytes/s Programming," pp. 54-55. 2003 Symposium on VLSI Technology, "Highly Manufacturable 90nm NOR Flash Technology with 0.081μm^2 Cell Size," pp. 91-92. 2004 Symposium on VLSI Technology, "A 70nm NOR Flash Technology with 0.049μm2 Cell Size," pp. 238-239.)

[26] NOR Flash uses channel hot electron programming which requires steep junction profile to generate, with the consequence of difficulties in controlling the short channel effect, but yet the NOR architecture is vulnerable to device leakage. Since the tunnel oxide thickness cannot be scaled, controlling the short channel effect imposes a very difficult challenge for scaling. The gate length generally is substantially larger than F in recent years. Despite this difficulty, the area factor has been maintained at 10F2 in recent years (see note 25).

[27] 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 pose fundamental problems for retention reliability.

[28] ONO has been used as the interpoly dielectric up to now and most likely in the near future. However, at 32 nm and beyond high-κ IPD may be necessary to maintain GCR at 0.6 or above. Currently, the GCR is achieved by wrapping the control gate over the sidewalls of the floating gate thus increasing the control gate to floating gate capacitor area. At 32 nm or below, the gap between adjacent gloating gates becomes too narrow for the ONO and control gate to fill in and this semi-vertical structure will cease to function.

[29] The gate coupling ratio (GCR) is the (control gate to floating gate capacitance)/(total floating gate capacitance). GCR must be greater than about 0.6 for proper device operation.

[30] This is the highest voltage relative to ground seen in the cell array, usually supplied by on-chip charge pumping circuits. Low voltage is desired to reduce the charge pumping circuit overhead and simplify processing. The introduction of high-κ IPD will help to reduce the erase voltage.

*[31] In principle the read current decreases with scaling at a rate W/(L*Cox) to prevent voltage overdrive (read disturb). Since access time depends critically on the read current and is an important performance parameter for NOR Flash the read current decreases slower than W/(L*Cox) to maintain performance. This may cause read disturb issues in long term years.*

[32] E/W endurance requirements vary with the specifics of an application, but 1E5 cycles have been accepted as the historical minimum acceptable level for a useful NOR product.

[33] Retention is a defect related parameter rather than an intrinsic device characteristic. Improvement in defect control and accumulation of device history is expected to eventually allow specification of 20 years retention. Also, it should become possible to accept a reduced retention specification as a tradeoff for increased E/W endurance.

[34] Cell read out distinguishes between four levels of charge storage to provide two storage bits (Multilevel cell MLC). Progression to 8 or 16 levels is potentially possible but maintaining reasonable V_t, read speed and array efficiency beyond 2-bit/cell are challenging. Unlike NAND Flash where density is a key competitive advantage, performance and reliability tend to hold higher importance than density for NOR Flash, thus the pressure to higher level MLC is not as strong as for NAND Flash.

[35] Virtual ground array uses junction isolation and buried diffusion for bit line, thus requires no STI isolation or bit line contact in the cell. In principle the area factor for VG array can be as small as $4F^2$, compared to ~ $10F^2$ for an array with cell STI isolation and contact. The scaling of buried diffusion is a difficult challenge and junction isolation produces more leakage paths and complicates the design. Large R&D effort is needed to implement VG array and overcome its shortcomings to take advantage of its smaller cell size.

[36] Charge trapping device for NOR application uses mainly a SONOS structure and thus often is confused with the conventional SONOS device. Conventional SONOS is more suitable for a NAND array. The device is programmed by Fowler-Nordheim tunneling of electrons from the substrate and the charge is stored in the SiN layer of SONOS. Since the electrons are stored in deep SiN traps it is difficult to de-trap by Fowler-Nordheim tunneling. Instead, a very thin (2–3 nm) tunnel oxide is used to allow substrate hole tunneling into the SiN to erase the device. However, such thin tunnel oxide also allows direct tunneling of holes from the substrate even under weak electric field produced by the stored electrons and good data retention is difficult to achieve. NROM is a device proposed to solve the SONOS problem in a NOR array[15]. Using channel hot electron to program the cell, electrons are stored in the SiN layer near the edge of S/D junctions. To erase, band-to-band tunneling generated hot holes are injected into the SiN. A relatively thick (4–5 nm) tunnel oxide is adopted and this solves the data retention issue. NROM also has the advantage of storing two bits of information in one device (source side and drain side) and it applies a reverse read method to distinguish the different states. NROM is built on a virtual ground array architecture and has a relatively small cell size. It is used not only for NOR Flash, but also for some data storage applications even though it is not a NAND structure.

[37] CT NOR SLC Flash stores one bit on the source side and one bit on the drain side, or 2-bit/cell. The MLC Flash stores 2 bits on the source side and two bits on the drain side, thus 4-bit/cell.

[38] Although physically storing two bits in the same device, the gate length scaling is not limited by left-right bit interference. The scaling is limited by the same factors for floating gate device - junction breakdown voltage and short channel effect. 4-bit/cell MLC device window, on the other hand, is affected by left-right bit interference and the L_g scaling for MLC may be more gradual than SLC.

[39] Because electrons are trapped in deep levels in SiN, the tunnel oxide can be scaled more aggressively than for floating gate device.

[40] Reducing the thickness of charge trapping SiN will reduce the EOT but will degrade trapping efficiency.

[41] Interpoly dielectric must be thick enough to assure retention.

[42] This entry is the critical dimension "F" within the FeRAM cell for stand-alone memory devices (not embedded devices).

[43] This is the area factor "a" = cell size/F^2. FeRAM cell size is presented in terms of F^2 multiples of the FeRAM implementation technology.

[44] FeRAM cell structures have migrated to one transistor, one capacitor (1T1C) formats. (Refs. J.H. Park, et al., "Fully Logic Compatible (1.6V Vcc, 2 Additional FRAM Masks) Highly Reliable Sub 10F2 Embedded FRAM with Advanced Direct Via Technology and Robust 100 nm thick MOCVD PZT Technology", 2004 IEDM, 23.7.1, pp. 591-594. Y. M. Kang et al., "Sub-1.2V Operational, 0.15µm/12F^2 Cell FRAM Technologies for Next Generation SoC Applications", 2005 Symposium on VLSI Technology, 6B-4, pp. 102-103.) Other alternative configurations are under investigation such as Chain-FeRAM. (Refs. H. Kanaya et al., "A 0.602µm^2 Nestled Chain Cell Structure Formed by One Mask Etching Process for 64Mbit FeRAM," 2004 Symposium for VLSI Technology, pp. 150-151. N. Nagel et al., "New Highly Scalable 3 Dimensional Chain FeRAM Cell with Vertical Capacitor," 2004 Symposium on VLSI Technology, pp. 146-147.)

[45] The geometry of the capacitor is a key factor in determining cell size. Stacked planar films are expected to be replaced by more efficient 3D structures.

[46] This is the footprint of the capacitor in micrometers squared. It is this area that constitutes the capacitor area contribution to the cell size. For 2005–2006 ~19F^2, for 2007 - 2009 ~16F^2, and for 2010–2020 ~10F^2 (3D capacitor) are assumed.

[47] This is the actual effective area of the capacitor. It is larger than the footprint for 3D capacitor because of the utilization of area in the third dimension.

[48] This ratio of the effective area to the footprint gives a measure of the impact of utilization of the third dimension.

[49] This is the operating voltage (V_{op}) applied to the capacitor. Low voltage operation is a difficult key design issue. Generally the ferroelectric film thickness needs to be decreased in order to reduce the V_{op}, with great technological challenges. (Ref. D. C. Yoo et al., "Highly Reliable 50nm-thick PZT Capacitor and Low Voltage FRAM Device Using Ir/$SrRuO_3$/MOCVD PZT Capacitor Technology", 2005 Symposium on VLSI Technology, 6B-3, pp. 100-101.)

[50] The minimum switching charge density in µC/cm^2 is a useful design parameter. It is equal to the cell minimum switching charge divided by the capacitor actual effective area. The capacitor voltage is taken as V_{op}.

[51] FeRAM is a destructive read-out technology, so every read is accompanied by a write to restore the data. Endurance cycles are taken as the sum of all read and all write cycles. For FeRAM to compete with DRAM and SRAM the cycle endurance should be about 1E15. Testing time is a serious concern. Note that operation at 100 MHz for 3 years would accumulate 1E16 cycles.

[52] MRAM devices are expected to lag the CMOS current technology up until 45 nm half pitch in 2010. This entry provides the F value for designs in the indicated time period.

[53] The MRAM switching field is the magnetic intensity H required to change the direction of magnetization of the cell.

*[54] MRAM switching energy per bit is calculated as (write current * power supply voltage * write time). It is preferred to use the median value of switching energy measured on a multi-megabit array. A good estimate of power drain is (switching energy * number of writes per second).*

*[55] MRAM active bit area is the area of the magnetic material stack within the cell. It represents the "A" in the R*A product.*

*[56] MRAM resistance-area product (i.e., the R*A product) is an intrinsic property of the magnetic material stack that provides a convenient basis for comparing cells of different sizes. The R*A product can be computed by measuring the effective low state resistance (R_{low}) of the magnetic tunnel junction and multiply it by the active bit area of the magnetic stack.*

[57] MRAM magnetoresistive ratio is calculated as $100*(R_{high} - R_{low})/R_{low}$. This ratio summarizes the difference between a logic ONE and a logic ZERO, and as such it represents the intrinsic capability of the magnetic stack. The magnetic tunnel junction resistance values are to be measured at low currents.

[58]This is the critical dimension, F.

[59] The area factor "a" = cell area/F^2. This entry is the expected PCRAM cell area in multiples of the implementation technology F^2. PCRAM requires significant reset current to change the phase-change element from crystalline to amorphous. A BJT transistor is capable of providing more current per unit area compared to a MOSFET, thus helps to reduce the cell size. Both BJT and nMOSFET access device cells are represented in this table. PCRAM is capable of MLC multi-bit per cell. This area factor is per cell, not per bit.

[60] The area factor "a" = cell area/F^2. This entry is the expected PCRAM cell area in multiples of the implementation technology F^2. PCRAM requires significant reset current to change the phase-change element from crystalline to amorphous. A BJT transistor is capable of providing more current per unit area compared to a MOSFET, thus helps to reduce the cell size. An nMOSFET transistor has larger cell size in the near term years, but offers simple process and low voltage operation. Both BJT and nMOSFET access device cells are represented in this table. PCRAM is capable of MLC multi-bit per cell. This area factor is per cell, not per bit.

[61] The expected "typical" PCRAM cell size with BJT access device is presented in micrometers squared.

[62] The expected "typical" PCRAM cell size with nMOSFET access device is presented in micrometers squared.

[63] PCRAM is capable of MLC multi-bit/cell operation since the resistance ratio between amorphous and crystalline state is typically 100–1,000. This entry is the expected number of MLC bits per cell.

[64] The expected cell size per MLC bit for the PCRAM with BJT cell. It is the physical cell size divided by the number of MLC bits per cell.

[65] The expected cell size per MLC bit for the PCRAM with nMOSFET cell. It is the physical cell size divided by the number of MLC bits per cell.

[66] PCRAM phase change element must be substantially smaller than the technology F to have efficiency reset operation with reasonable current. This entry is the expected dimension for the phase change element in nanometers.

[67] PCRAM phase change volume is a key factor for device design and peak power requirement. This entry is the expected phase change volume in nanometer cubed.

[68] This entry is the expected reset current for PCRAM in microamperes.

[69] The set resistance is a key design factor for PCRAM read speed.

[70] This entry is the expected current density output from the BJT access device required to reset the PCRAM cell (from crystalline to amorphous state). It is a compromise between larger area BJT (which causes larger cell size) and higher output current (which requires higher operation voltage).

[71] This entry is the expected BJT emitter area that can provide the needed reset current, assuming the BJT current density is met.

[72] This entry is the expected current density output from the nMOSFET access device required to reset the PCRAM cell (from crystalline to amorphous state). It is a compromise between larger width nMOSFET (which causes larger cell size) and higher output current (which requires higher operation voltage or less reliable device).

[73] This entry is the expected nMOSFET gate width that can provide the needed reset current, assuming the MOSFET output current density is met.

[74] This entry is the expected PCRAM data retention that will allow it to be used as a nonvolatile memory. Data retention mechanism for PCRAM is not yet thoroughly studied. Recent published data indicate >10 years of retention at elevated temperatures. (Refs. S. J. Ahn et al., "Highly Manufacturable High Density Phase Change Memory of 64Mb and Beyond," 2004 IEDM, 37.2, pp. 907-910. A. L. Lacaita et al., "Electrothermal and Phase Change Dynamics in Chalcogenide-Based Materials," 2004 IEDM, 37.3, pp. 911-914.)

[75] This entry is the expected PCRAM W/E cycling endurance. Recent published data indicate cycling endurance from 1E+9 to 1E+13. (Refs. S.J. Ahn et al., "Highly Manufacturable High Density Phase Change Memory of 64Mb and Beyond," 2004 IEDM, 37.2, pp. 907-910. S. Lai et al., "Current Status of Phase Change Memory and Its Future," 2003 IEDM, pp. 255-258.)

NON-VOLATILE MEMORY POTENTIAL SOLUTIONS

Nonvolatile memory (NVM) technologies combine CMOS peripheral circuitry with a memory array. The memory array generally requires additional, but CMOS compatible, processes to implement the non-volatility. Non-volatile memories are used in a wide range of applications, some standalone and some embedded, with varying requirements that depend on the application. The memory array architecture and signal sensing method also differ for different applications. The technical challenges are difficult, and in some cases fundamental physics limitations may be reached before the end of the current roadmap. For charge storage devices, the number of electrons in the storage node, whether for single level logic cells (SLC) or multi-level logic cells (MLC), needs to be sufficiently high to maintain stable threshold voltage against statistical fluctuation, and cross talk between neighboring bits must be reduced while the spacing between neighbors decreases. Meanwhile, data retention and cycling endurance requirements must be maintained, and in some cases even increased for new applications. Non-charge-storage devices also may face fundamental limitations when the storage volume becomes small such that random thermal noise starts to interfere with signal.

FLOATING GATE FLASH DEVICES

Floating gate Flash devices achieve non-volatility by storing and sensing the charge on a floating gate. The conventional memory transistor vertical stack consists of a refractory polycide control gate, an interpoly dielectric (IPD) that usually consists of triple oxide-nitride-oxide (ONO) layers, a polysilicon floating gate, a tunnel dielectric, and the silicon substrate. The tunnel dielectric must be thin enough to allow charge transfer to the floating gate at reasonable voltage levels and thick enough to avoid charge loss when in read or off modes. The interpoly dielectric thickness must scale with the tunnel dielectric to maintain adequate coupling of applied erase or write pulses to the tunnel dielectric. The gate coupling ratio (GCR), defined as the capacitance ratio of the control gate to floating gate capacitor to the total floating gate capacitance (control gate to floating gate + floating gate to substrate), is a critical scaling parameter, and must be ≥ 0.6. In most structures, to achieve a GCR ≥ 0.6, the word line (control gate) wraps around the sidewall of the floating gate to provide extra capacitance. Since IPD (ONO) is $\geq$ 15 nm in thickness, it is difficult to achieve the wrap around structure when the bit line spacing becomes 30 nm or less. Therefore, maintaining the GCR is a major challenge for floating gate Flash device scaling.

The tunnel oxide thickness for the floating gate device poses a great scaling challenge, and there is no currently recognized solution. This inhibits the scaling of gate length for NOR Flash, since leakage from short channel effects introduces severe program disturb. The use of high-κ dielectric in the IPD dielectric (ONO) will be helpful to both reduce the total EOT while maintaining or even increasing the gate coupling ratio. However, planar devices must be used at half-pitch of 30nm or less, and the lost geometrical advantage (wrap around) must be compensated by the higher dielectric constant of high-κ. Therefore, the permittivity must be substantially higher (> 3×) than that of SiO_2.

NOR AND NAND FLOATING GATE FLASH

A NOR Flash cell consists of a single MOS transistor serving both as the cell isolation device and the storage node. The threshold voltage of the transistor is modulated by charges stored in the floating gate and is used as an indication of the storage status. The storage cell may store single level logic (SLC, actually means bi-level logic 1 and 0) or multiple logic levels (MLC, e.g., (11), (10), (00), and (01)). The memory array is an X-Y cross wire structure, thus allowing random access. Programming is by channel hot electron or other variations of hot electron generation, and erasing is by Fowler-Nordheim tunneling of electrons out of the floating gate. The generation of hot electrons requires high lateral electric field under the device and is provided by steep junction profiles. This in turn causes short channel effect and leakage current

that produces program disturb. Halo implants are used to control device leakage, and this subsequently reduces the junction breakdown voltage and limits the scaling capability. If virtual ground array, which requires no bit line contact within the array, is successfully developed then the NOR Flash cell size can be drastically reduced in the near term years.

A NAND Flash cell also consists of a single MOS transistor, serving mainly as the storage device. The NAND array consists of bit line strings of 32 devices or more with a selection device at each end. This architecture requires no direct bit line contact to the cell, thus allows the smallest cell size. During programming or reading, the unselected cells in the selected bit line string must be turned on and serve as "pass" devices, thus the data stored in each device cannot be accessed randomly conveniently. Both programming and erasing are by Fowler-Nordheim tunneling of electrons into and out of the floating gate. The low Fowler-Nordheim tunneling current allows the simultaneous programming of many bits, thus gives high programming speed. Since devices in the same bit line string serve as pass transistors their leakage current does not affect programming or reading operation, and without the need for hot electrons junctions can be shallow. Thus even with the same tunnel oxide limitation NAND devices are easier to scale than NOR devices. Designed to provide storage and access to large quantities of data but not to instantly execute program codes, NAND Flash generally employs error correction code (ECC) algorithms, and is more fault tolerant than NOR Flash. How to maintain a GCR > 0.6 and to avoid floating gate to floating gate cross talk are two difficult challenges when scaling to 32 nm and below. The timely development of high-κ IPD in a planar device will be a key milestone for continued scaling. Eventually, the few-electron limitation will cause unacceptable retention time distribution, for which there is no currently recognized solution.

Charge Trapping (CT) Devices in NOR Architecture

The threshold voltage of a device may also be affected by charges stored in a charge trapping layer, such as SiN. Charge trapping devices using a SiN as the trapping layer are usually called SONOS, since the device has a SONOS stack—a Si (polycide) gate, a blocking oxide, a nitride storage layer, and a tunnel oxide. The prevailing SONOS device using a relatively thick tunnel oxide in a NOR architecture is commonly known as NROM.[15] NROM uses channel hot electron for programming, and band-to-band tunneling of hot hole for erasing. Since charges injected into the nitride storage layer are well localized near the junctions two bits of information can be stored, one on the source side and one on the drain side, in the same device. The threshold voltage of the device can be read out by shielding the drain side bit with a drain bias and "reverse read" the source side information.

NROM NOR array can be implemented in a virtual ground architecture for which buried diffusion serves as the bit line and the device channel lies along the word line (polycide) direction. This structure requires neither bit line contact nor STI in the cell, thus offering a substantially smaller cell than the conventional NOR array. The cross talk between the two storage nodes in the same device cannot be completely eliminated. This so-called "second bit effect" restricts the threshold voltage window each storage node can carry, and the implementation of MLC in NROM poses a higher level of challenge than for floating gate devices. However, NROM is intrinsically 2-bit/cell and a 4-level MLC implementation results in 4-bit/cell, compared to 16-level MLC required for floating gate device for the same density. The virtual ground array offers a factor of 1.5× to 2× density advantage over conventional NOR architecture using the same design rules, and the single poly process reduces the mask layers.

Charge trapping devices do not have the gate coupling ratio issue floating gate devices face; however, the scaling challenges are otherwise quite similar. The virtual ground array and 2-bit/cell operation are sensitive to device leakage and the use of hot carriers for programming and erasing increases the vulnerability to reliability failures. The scaling limitation is similar to floating gate NOR—leakage from short channel effect and junction breakdown.

Charge Trapping Devices in NAND Architecture

Currently most NAND products are fabricated with floating gate devices. The difficult challenges of maintaining or increasing the GCR and reducing the neighboring cell cross talk may be bypassed by using charge trapping devices. The single gate controls the MOS device channel directly and thus there is no GCR issue, and the cross talk between thin nitride storage layers is insignificant. Nitride trapping devices may be implemented in a number of variations of a basic SONOS type device. SONOS using a simple tunnel oxide, however, is not suitable for NAND application—once electrons are trapped in deep SiN trap levels they are difficult to detrap even under high electric field. In order to erase the device quickly holes in the substrate must be injected into the SiN to neutralize the electron charge. Since the hole barrier for SiO_2 is high (~4.1 eV), hole injection efficiency is poor and sufficient hole current is only achievable by using very thin tunnel oxide (~ 2 nm). Such thin tunnel oxide, however, results in poor data retention because direct hole tunneling from the substrate under the weak retention built-in field cannot be stopped.

Several variations of SONOS have been proposed recently. Tunnel dielectric engineering concepts are used to modify the tunneling barrier properties to create "variable thickness" tunnel dielectric. For example, triple ultra-thin (1–2 nm) layers

of ONO are introduced to replace the single oxide (BE-SONOS).[16] Under high electric field, the upper two layers of oxide and nitride are offset above the Si valence band, and substrate holes readily tunnel through the bottom thin oxide and inject into the thick SiN trapping layer above. In data storage mode, the weak electric field does not offset the triple layer and both electrons in the SiN and holes in the substrate are blocked by the total thickness of the triple layer. In MANOS (metal-Al_2O_3-nitride-oxide-Si),[17] a high-κ blocking dielectric and metal gate are combined to both prevent gate injection during erase operation, and to boost the electric field at tunnel oxide. A thicker (3–4 nm) tunnel oxide may be used to prevent substrate hole direct tunneling during retention mode.

Non-Planar and Multi-Gate Devices

Non-planar and multi-gate devices such as FinFET and surround-gate devices provide better channel control and allow further scaling of both floating gate and nitride trapping devices. However, the vertical structure also presents new challenges. For example, the space between fins must be sufficiently wide to allow room for tunnel oxide, floating gate and IPD (for floating gate device) and may forbid scaling beyond 30 nm if innovative solutions are not found.

3D Stacking of Memory Arrays

When the number of stored electrons reaches statistical limits, even if devices can be further scaled and smaller cells achieved, the threshold voltage distribution of all devices in the memory array will become uncontrollable and logic states unpredictable. The memory density cannot be increased by continued scaling, but may be increased by stacking memory layers vertically. Successful stacking of memory arrays vertically has been demonstrated recently. One approach uses single crystal Si by lateral epitaxial growth.[18] Another uses polycrystalline Si thin-film transistor (TFT) device.[19] The processing temperature and thermal budget must be such that the layers fabricated earlier are not degraded by the additional thermal processes. This imposes a significant challenge to either achieve identical devices in different layers that experience different thermal processes, or design circuits that can handle devices that are slightly different in each layer. Although 3D stacking can help increase the memory density beyond conventional scaling, its effectiveness diminishes after several layers are stacked. The complexity in interconnects increases and the array efficiency decreases with the number of layers. In addition, the complex processing and the large number of masks cumulatively affect the yield. Recently, a "punch and plug" approach is proposed to fabricate the bit line string vertically to simplify the processing steps.[20]

Non-Charge-Storage Non-Volatile Memories

Since the ultimate scaling limitation for charge storage devices is too few electrons, devices that provide memory states without electric charges are promising to scale further. Several non-charge-storage memories have been extensively studied and some commercialized, and each has its own merits and unique challenges. Some of these are uniquely suited for special applications and may follow a scaling path independent of NOR and NAND Flash. Some may eventually replace NOR or NAND in the longer term. Logic states that do not depend on charge storage eventually also run into fundamental physics limits. For example, small storage volume may be vulnerable to random thermal noise, such as the case of superparamagnetism limitation for MRAM.

FeRAM devices achieve non-volatility by switching and sensing the polarization state of a ferroelectric capacitor. To read the memory state the hysteresis loop of the ferroelectric capacitor must be traced and the data must be written back after reading. Because of this "destructive read," it is a challenge to find ferroelectric and electrode materials that provide both adequate change in polarization and the necessary stability over extended operating cycles. The ferroelectric materials are foreign to the normal complement of CMOS fabrication materials, and can be degraded by conventional CMOS processing conditions. Thus the ferroelectric materials, buffer materials, and process conditions are still being refined. So far the most advanced FeRAM[21] is less dense than NOR and NAND Flash using the same design rules, fabricated at least one technology generation behind NOR and NAND Flash, and not capable of MLC. Thus the hope for near term replacement of NOR or NAND Flash has faded. However, FeRAM is fast, low power, and low voltage and thus is suitable for RFID, smart card, ID card, and other embedded applications. In order to achieve density goals with further scaling, the basic geometry of the cell must be modified while maintaining the desired isolation. Recent progress in electrode materials shows promise to thin down the ferroelectric capacitor and extends the viability of 2D stacked capacitor through most of the near term years. Beyond this the need for 3D capacitor still poses steep challenges.

MRAM devices employ a magnetic tunnel junction (MTJ) as the memory element. An MTJ cell consists of two ferromagnetic materials separated by a thin insulating layer that acts as a tunnel barrier. When the magnetic moment of one layer is switched to align with the other layer (or to oppose the direction of the other layer) the effective resistance to current flow through the MTJ changes. The magnitude of the tunneling current can be read to indicate whether a ONE or a ZERO is stored. MRAM probably is the closest to an ideal "universal memory" since it is non-volatile and fast and can be cycled indefinitely, thus may be used as NVM as well as SRAM and DRAM. Control and development of the MTJ

dimensions and material properties is the major challenge. Also, the tunnel Al_2O_3 layer must endure current stressing at each read/write operation. Management of the material sensitivities to IC processing temperatures and conditions is also an issue. In the long term, the challenge will be the achievement of adequate magnetic intensity H fields to accomplish switching in scaled cells, where electromigration limits the current density that can be used. Recent advances in "spin transfer torque" approach where the spin polarity is transferred from one material directly to another through a "spin current" without resorting to an external magnetic field offer a new potential solution.[22]

PCRAM devices use the resistivity difference between the amorphous and the crystalline states of chalcogenide glass (the most commonly used compound is $Ge_2Sb_2Te_5$, or GST) to store the logic ONE and logic ZERO levels. The device consists of a top electrode, the chalcogenide phase change layer, and a bottom electrode. The leakage path is cut off by an access transistor in series with the phase change element. The phase change write/erase consist of two operations: (1) RESET, for which the chalcogenide glass is momentarily melted by a short electric pulse and then quickly quenched into amorphous solid with high resistivity, and (2) SET, for which a lower amplitude but longer pulse (10ns–100ns) anneals the amorphous phase into low resistance crystalline state. The 1T1R cell is comparable in size to NOR Flash, but the device may be programmed to any final state without erasing the previous state, thus provides substantially faster programming speed. The simple resistor structure and the low voltage operation also make PCRAM attractive for embedded NVM applications. The major challenges for PCRAM are the high current (fraction of mA) required to reset the phase change element, and the relatively long set time. Since the volume of phase change material decreases rapidly with each technology generation, both above issues become easier with scaling. Interaction of phase change material with electrodes may pose long-term reliability issues and limit the cycling endurance and is a major challenge for the maturity of PCRAM.

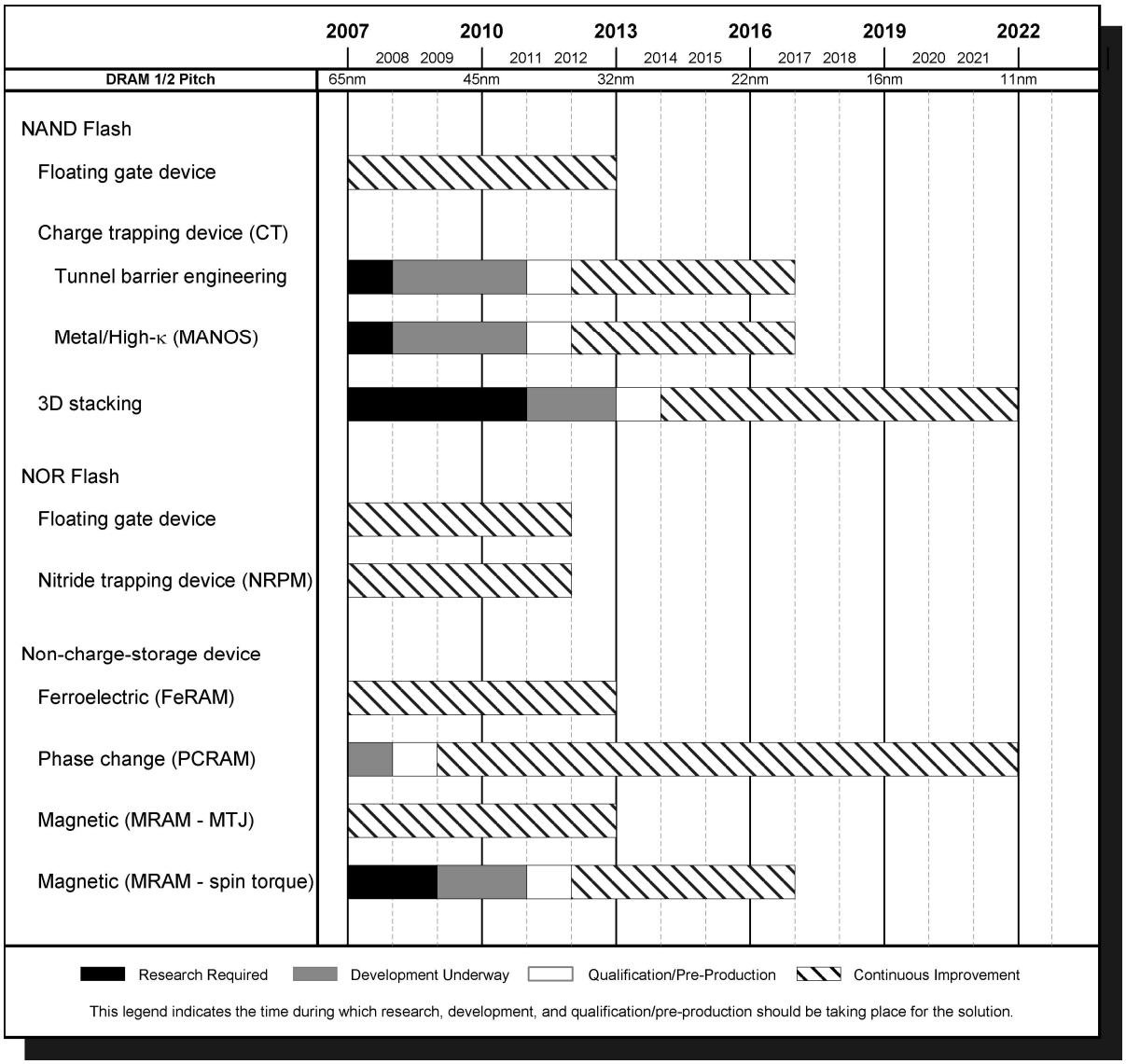

Figure PIDS9 Non-volatile Memory Potential Solutions

RELIABILITY TECHNOLOGY REQUIREMENTS AND POTENTIAL SOLUTIONS

INTRODUCTION

Reliability is an important requirement for almost all users of integrated circuits. The challenge of realizing the required levels of reliability is increasing due to scaling, the introduction of new materials and devices, increasing stresses (fields, current densities, temperatures), and increasing constraints of time and money. Scaling produces ICs with more transistors and more interconnections, both on-chip and in the package. This leads to an increasing number of potential failure sites. Failure mechanisms are impacted by scaling. For example, the time dependent dielectric breakdown (TDDB) of silicon oxy-nitride gate insulators has changed from electric-field-driven to voltage-driven as the insulator thickness has been scaled below 5 nm. In addition, negative bias temperature instability (NBTI) in p-channel devices, which used to be a minor effect when threshold voltages were larger, is now a great concern at the smaller threshold voltages of state-of-the-art devices.

Scaling also leads to increases in the stresses that cause failures. First, the current density is increasing and this increase impacts interconnect reliability. Second, voltages are often scaled down more slowly than dimensions, leading to increased electric fields that impact insulator reliability. Third, scaling has led to increasing power dissipation that results

in higher temperatures, larger temperature cycles, and increased thermal gradients, all of which impact multiple failure mechanisms. The temperature effects are further aggravated by the reduced thermal conductivity that accompanies the reduction in the dielectric constant of the dielectrics between metal lines.

There are even more profound reliability challenges associated with revolutionary changes associated with new materials and new devices. Recognized failure mechanisms can change. For example, aluminum is stable after being deposited and the preferred path for electromigration is along grain boundaries. In contrast, there is grain boundary growth in copper after electroplating that can lead to stress voiding failures when a single via is connected to a wide metal line. In addition, in copper the preferred electromigration path is along the surface, making copper electromigration and stress voiding much more sensitive to the properties of the intermetal dielectric. This makes the reliability of copper lines much more sensitive to interfaces compared to aluminum. The electromigration in copper will also become worse as the cross section of the copper lines is reduced with scaling. New materials and devices can introduce new failure mechanisms. For example, the poor mechanical and thermal properties of low-κ intermetal dielectrics can lead to mechanical failure mechanisms not seen in silicon dioxide intermetal dielectrics. The impact of an unrecognized failure mechanism that made it into end products would be significant.

These reliability challenges will be exacerbated by the need to introduce multiple major technology changes in a brief period of time. Interactions between changes can increase the difficulty of understanding and controlling failure modes. Furthermore, having to deal simultaneously with several major issues will tax limited reliability resources.

TOP RELIABILITY CHALLENGES

Table PIDS6 indicates the SEMATECH Reliability Council's consensus view of the top five near-term difficult reliability challenges. It expands on the PIDS overall Difficult Challenge 3, "Timely assurance for the reliability of multiple and rapid material, process, and structural changes," described at the beginning of this chapter.

The first near-term reliability challenge concerns failure mechanism associated with the MOS transistor. The time to a first breakdown event is decreasing with scaling. This first event is often a "soft" breakdown. However, depending on the circuit it may take more than one soft breakdown to produce an IC failure. Negative bias temperature instability is a gradual degradation in the properties of p channel transistors. It has grown in importance as threshold voltages have been scaled down and as silicon oxy-nitride has replaced silicon dioxide as the gate insulator. Burn-in may be impacted, as it may accelerate NBTI shifts. High κ will impact insulator failure modes (e.g., breakdown and stability) as well as transistor failure modes such as hot carrier effects and negative bias temperature instability. To put this change into perspective, even after decades of study, there are still issues with silicon dioxide reliability that need to be resolved. The replacement of polysilicon with metal gates also impacts insulator reliability and raises new thermo-mechanical issues. The simultaneous introduction of high κ and metal will make it more difficult to determine reliability than if high κ were first introduced with poly gates.

As mentioned above, the move to copper and low κ has raised issues with electromigration, stress voiding, the poorer mechanical, interface adhesion and thermal conductivity of low-κ dielectrics and the porosity of low-κ dielectrics. The change from Al to Cu has changed electromigration (from grain boundary to surface diffusion) and stress voiding (from thin lines to vias over wide lines). Reliability in the Cu/low-κ system is very sensitive to interface issues. The poorer mechanical properties of low κ also impact wafer probing and packaging. The poorer thermal conductivity of low-κ dielectrics leads to higher on-chip temperatures and higher localized thermal gradients, which impact reliability. The porosity of low-κ dielectrics can trap and transport process chemicals and moisture leading to corrosion and other failure mechanisms

There are additional reliability challenges associated with advanced packaging for higher performance, higher power integrated circuits. Increasing power, increasing pin count, and increasing environmental regulations (e.g., lead-free) all impact package reliability. The interaction between the package and die will increase, especially with the introduction of low-κ intermetal dielectrics. The move to multi-chip packaging and/or heterogeneous integration makes reliability even more challenging. As currents increase and the size of balls/bumps decreases, there is an increased risk of failures due to electromigration.

ICs are used in a variety of different applications. There are some special applications for which reliability is especially challenging. First, there are the applications in which the environment subjects the ICs to stresses much greater than found in typical consumer or office applications. For example, automotive, military, and aerospace applications subject ICs to extremes in temperature and shock. In addition, aviation and space-based applications also have a more severe radiation environment. Second, there are important applications (e.g., implantable electronics, safety systems) for which the consequences of an IC failure are much greater than in mainstream IC applications. In general scaled-down ICs are less "robust" and this makes it harder to meet the reliability requirements of these special applications.

At the heart of reliability engineering is the fact that there is a distribution of lifetimes for each failure mechanism. With low failure rate requirements we are interested in the early-time edge of the failure time distributions. There has been an increase in process variability with scaling (e.g., distribution of dopant atoms, CMP variations, line-edge roughness). At the same time the size of a critical defect decreases with scaling. These trends will translate into an increasing spread in the time to failure distributions and, thus, a decreasing time to first failure. We need to develop reliability engineering tools (e.g., screens, qualification, Design for Reliability) software that can handle this increase in variability.

The single long-term Reliability Difficult Challenge concerns novel, disruptive changes in devices, structures, materials, and applications. For example, at some point there will be a need to implement non-Copper interconnect (e.g., optical interconnect). For the selected non-Copper solution it is likely that there will be little, if any, reliability knowledge (as least as far as use in ICs is concerned). This will involve a significant effort to discover, model (both a statistical model of lifetime distributions and a physical model of how lifetime depends on stress, geometries, and materials), and apply (new building-in reliability, designing-in reliability, screens, and tests). It also seems likely that there will be less-than-historic amounts of time and money to develop these new reliability capabilities. Disruptive material or devices lead to disruption in reliability capabilities and it will take considerable resources to develop those capabilities.

Table PIDS6 Reliability Difficult Challenges

Difficult Challenges ≥ 22 nm	*Summary of Issues*
Transistor Reliability	Time dependent dielectric breakdown Negative bias temperature instability Threshold voltage shifts due to traps, carrier injection, program or erase Mobility degradation due to mechanical stress relaxation or interface state density change New or changed failure mechanisms (TDDB, PBTI, NBTI< moisture absorption, etc.) resulting from high κ/metal gate
Interconnect Reliability	Copper electromigration and stress voiding in scaled interconnects (lines and vias) Electrical breakdown of interconnect dielectrics, especially low κ and ultra low κ Moisture absorption/transport due to voids in porous low κ dielectrics Cu (ionic) migration through cracked or thin barrier metals
Packaging Reliability	New failure mechanisms associated with Pb-free solders and new mold compounds Electromigration in package traces, vias, and bumps Impact of multichip modules and stacked dies on failure rate Solder ball electromigration, for example in CSP and flip chip Radioactive contaminants in packaging materials
Reliability in Extreme and/or Critical Applications	Automotive (define mission profile for HOT underhood versus passenger and substantial cycling) Military (rugged versus shock and dust, highly diverse environmental requirements) Space, i.e., radiation hard Aeronautical (singe event effects tolerant and large, fast temperature swings) Medical (corrosive, hermeticity, and safety)
Impact of Variability on Reliability	Statistic variation growing larger and defect size is comparable to feature size: Distribution of dopant atoms; subtle ultra-thin gate oxide defects; line edge roughness and other litho "fidelity" issues; surface scattering How to cope with cost-effective screens and qualifications that capture some "good" units Design for Reliability in face of large percentage process variability How to use yield to drive reliability
Difficult Challenges<22 nm	*Summary of Issues*
Reliability of novel devices, structures, materials and applications	ITRS proposes many new materials and structures, yet currently very little known about failure mechanisms Need to have reliability characterization in place well in advance of application to develop appropriate reliability requirements and qualification procedures Design for Reliability tools

RELIABILITY REQUIREMENTS

Reliability requirements are highly application dependent. For most customers, current overall chip reliability levels (including packaging reliability) need to be maintained over the next fifteen years in spite of the reliability risk inherent in massive technology changes. However, there are also niche markets that require reliability levels to improve. Applications that require higher reliability levels, harsher environments, and/or longer lifetimes are more difficult than the

mainstream office and mobile applications. Note that even with constant overall chip reliability levels, there must be continuous improvement in the reliability per transistor and the reliability per meter of interconnect because of scaling. Meeting reliability specifications is a critical customer requirement and failure to meet reliability requirements can be catastrophic.

These customer requirements flow down into requirements for manufacturers that include an in-depth knowledge of the physics of all relevant failure modes and the existence of powerful reliability engineering capabilities for design-for-reliability, building-in-reliability, reliability qualification, and defect screening. There are some significant gaps in these capabilities today. Furthermore, these gaps will become even larger with the introduction of new materials and new device structures. Inadequate reliability tools lead to unnecessary performance penalties and/or unnecessary risks. Finally, as trade-offs between reliability and performance become more difficult, excess reliability margins need to be eliminated.

Reliability qualification always involves some risk. There is a risk of qualifying a technology that does not, in fact, meet reliability requirements or a risk of rejecting a technology that does, in fact, meet requirements. At any point in time a qualification can be attempted on a new technology. However, the risk associated with that qualification can be large. The level of risk is directly related to the quality of the reliability physics and reliability engineering knowledge base and capabilities.

The color-coding of the Reliability technology requirements is meant to represent the reliability risk associated with incomplete knowledge and tools for new materials and devices. The progression from white to yellow to striped indicates a growing reliability risk. The requirements first turn to yellow (Manufacturing Solutions are Known) in 2008 indicating a relative smaller risk associated with scaling, increased power. The "wild card" in 2008 will be if manufacturers introduce low κ/metal gate transistor stacks, which will present a considerable reliability risk.

The requirements then turn to striped (Interim Solutions Known) in 2013. This date is approximate. It is meant to represent the point in time where novel devices or materials are introduced (e.g., optical interconnect or a non-CMOS transistor or memory). As mentioned above these changes present a considerable reliability risk and require a considerable lead time to develop the needed capabilities in reliability physics and reliability engineering. Since we do not know exactly what these disruptive technologies will be we have no way, in advance of knowing the reliability risk. When we know the exact technologies proposed we can provide a better assessment of the reliability risk. We choose to use striped as opposed to solid red to reflect that reliability qualification can always be attempted. However, the poorer the quality of our reliability knowledge, the greater the reliability risk.

Table PIDS7a Reliability 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*	*22*	*20*	*18*	*16*	*14*	*13*	*11*	*10*
Early failures (ppm) (First 4000 operating hours) [1]	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**
Long term reliability (FITS = failures in 1E9 hours) [2]	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**50–2000**
SRAM Soft error rate (FITs/MBit)	**1000-2000**	**1000-2000**	**1000-2000**	**1000-2000**	**1000-2000**	**1000-2000**	**1000-2000**	**1000-2000**	**1000-2000**
Relative failure rate per transistor (normalized to 2007 value) [3]	**1.00**	**0.83**	**0.71**	**0.66**	**0.57**	**0.51**	**0.46**	**0.40**	**0.37**
Relative failure rate per m of interconnect (normalized to 2007 value) [4]	**1.00**	**0.50**	**0.50**	**0.50**	**0.25**	**0.25**	**0.25**	**0.12**	**0.12**

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

Table PIDS7b Reliability 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*	*6.3*	*5.6*	*5.0*	*4.5*
Early failures (ppm) (First 4000 operating hours) [1]	50–2000	50–2000	50–2000	50–2000	50–2000	50–2000	50–2000
Long term reliability (FITS = failures in 1E9 hours) [2]	50–2000	50–2000	50–2000	50–2000	50–2000	50–2000	50–2000
SRAM Soft error rate (FITs/MBit)	1000-2000	1000-2000	1000-2000	1000-2000	1000-2000	1000-2000	1000-2000
Relative failure rate per transistor (normalized to 2007 value) [3]	0.31	0.29	0.26	0.23	0.20	0.18	0.16
Relative failure rate per m of interconnect (normalized to 2007 value) [4]	0.12	0.06	0.06	0.06	0.03	0.03	0.03

Notes for Table PIDS7a and b:
Reliability requirements vary with different applications. For many mainstream customers it will be sufficient to hold current reliability levels steady during this period of rapid technological change. However, other customers would like reliability levels to be improved. Degradation of current reliability levels is not acceptable. Reliability requirements are for the packaged device and include both chip and package related failure modes.

A reliability qualification can always be attempted with available knowledge. The better the knowledge the less risk in the qualification and vice versa. Yellow coloring indicates some risk. Striped indicates a greater risk (due to changed and possible new failure modes). Finally, red indicates an unspecified solution (e.g., what technology will be used for post-Cu) for which the reliability risk cannot be assessed until details about the solution are provided.

[1] Failures during the first 4000 hours of operation (~1 year's use at 50% duty cycle). Early failures are associated with defects.

[2] Long term reliability rate applies for the specified lifetime of the IC.

[3] While the overall IC failure rate does not change with time, as the number of transistors per chip increases [from ORTC], the relative failure rate per transistor must decrease

[4] As the length of interconnect per chip increases [from Interconnect Technology Requirements tables], the failure rate per m of interconnect must decrease. Even more important for reliability is the increase in the number of vias.

RELIABILITY POTENTIAL SOLUTIONS

The most effective way to meet requirements is to have complete built-in-reliability and design-for-reliability solutions available at the start of the development of each new technology generation. This would enable finding the optimum reliability/performance/power choice and would enable designing a manufacturing process that can consistently have high reliability yields. Unfortunately, there are serious gaps in these capabilities today and these gaps are likely to grow even larger in the future. The penalty will be an increasing risk of reliability problems and a reduced ability to push performance, cost and time-to-market.

Meeting requirements requires an in-depth understanding of the physics of each failure mechanism and the development of powerful and practical reliability engineering tools. Historically, it has taken many years (typically a decade) before the start of production for a new technology generation to develop these capabilities (R&D is conducted on characterizing failure modes, deriving validated, predictive models and developing design for reliability and reliability TCAD tools.) The ability to qualify technologies has improved, but there still are significant gaps.

However, there is a limit to how fast reliability capabilities can be developed, especially for major technology discontinuities such as alternate gate insulators or non-traditional devices. An eleventh-hour "sprint" to try and qualify major technology shifts will be highly problematical without an existing and adequate reliability knowledge base.

The Reliability Potential Solutions shown in Figure PIDS10 cover the major technical discontinuities over the lifetime of the Roadmap. (There is a wide variety of changes not listed in this figure that also could impact reliability.) Because these are major discontinuities with serious reliability issues it takes several years to conduct the R&D to identify and model the failure modes (black bars), turn these results into practical reliability engineering capabilities (blue bars), and, finally to perform the qualification of a new technology (white bars). Even when new materials or devices enter production, there still is a need to improve the reliability models and the reliability engineering capabilities continually. Of course, less profound changes can be characterized in much less time. At present, the actual development of these potential solutions lags behind the needed milestones shown in Figure PIDS10.

For reliability capabilities to catch up requires a substantial increase in reliability research-development-application and cleverness in acquiring the needed capabilities in much less than the historic time scales. Work is needed on rapid characterization techniques, validated models, and design tools for each failure mechanism. The impact of new materials

like Cu, low κ and alternate gate dielectrics needs particular attention. Breakthroughs may be needed to develop design for reliability tools that can provide a high fidelity simulation of a large fraction of an IC in a reasonable time. As mentioned above, increased reliability resources also will be needed to handle the introduction of a large number of major technology changes in a brief period of time.

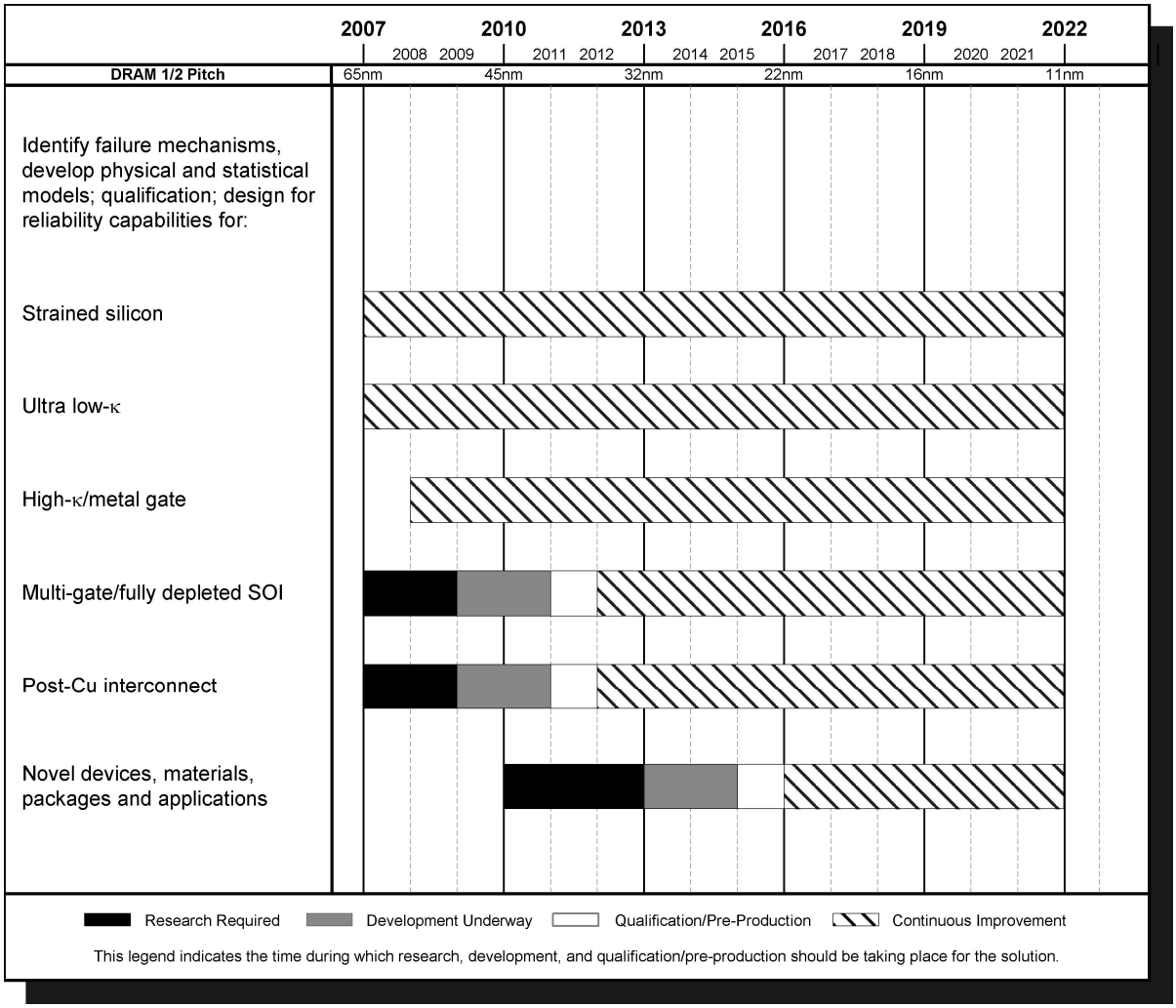

Figure PIDS10 Reliability Potential Solutions

Notes for Figure PIDS10:
[1] Strained Silicon has entered volume production. More aggressive use of strained silicon is expected in future technologies. Need to assess its impact on transistor failure mechanisms (e.g., hot carrier and NBTI) continually.
[2] Low-κ interlevel dielectrics have entered production. Successive technology generations will introduce lower and lower κ materials that may modify existing failure mechanisms and could introduce new failure mechanisms.
[3] Driven by PIDS Logic requirement to introduced high-κ gate insulator in 2008.
[4] Driven by PIDS Logic requirement to introduced metal gate(s) in 2008
[5] Driven by PIDS Logic requirement to introduce fully depleted SOI in 2008
[6] Driven by PIDS Logic requirement to introduce double (or triple) gate transistors in 2011.
[7] The timing of the need for a Post-Cu interconnect solution is not clear. We have assumed in this table that it will be introduced in 2013. If it is earlier or later than these boxes will need to be correspondingly shifted. The key message is that we need approximately a 6-year lead time for reliability.
[8] The timing of the need novel, non-CMOS devices is not clear. We have assumed in this table that it will be introduced in 2016. If it is earlier or later, then these boxes will need to be correspondingly shifted. The key message is that we need approximately a 6 year lead time for reliability.

CROSS TWG ISSUES

MODELING AND SIMULATION

Modeling and simulation needs to be enhanced to deal with the key innovations requested by the *PIDS* section, including enhanced mobility, high-κ gate dielectrics, metal gate electrodes, non-classical CMOS (ultra-thin body fully depleted SOI and multiple-gate MOSFETs), and quasi-ballistic transport leading to enhanced saturation current. These innovations will collectively drive major changes in process, materials, physics, design, etc. Other long-term issues requiring enhanced modeling and simulation include atomic-level fluctuations, statistical process variations, new interconnect schemes, and mixed-signal device technology. With the shrinking of feature sizes, new process steps, architectures and materials reliability issues at the device, interconnect and circuit level will become even more important and will need support from modeling and simulation to achieve the development speed required. Especially for devices that use SOI material, existing models (e.g., for dopant diffusion and activation, carrier transport or for stress) must be extended to cope with interface effects, which become increasingly important compared with bulk properties. These issues are in the *Modeling and Simulation* chapter of this ITRS, especially included in the subchapters on "Front-End Process Modeling," "Device Modeling" and "Interconnects and Integrated Passives Modeling." Finally, non-classical CMOS devices require the development of appropriate compact models to support their introduction.

INTER-FOCUS ITWG DISCUSSION

EMERGING RESEARCH DEVICES

The Emerging Research Devices (ERD) chapter describes and evaluates potential technology, including devices, architectures, and materials, beyond the current standard silicon CMOS technology. As such, it is concerned with the potential successor(s) to the CMOS described in the PIDS chapter. Toward or beyond the end of this Roadmap, when CMOS scaling will likely become ineffective and/or prohibitively costly, some version(s) of ERD technology will presumably be needed if the industry is to continue to enjoy rapid improvements in performance, lower power dissipation and cost per function, and higher functional density. Hence, the PIDS potential solutions tables include ERD solutions late in the Roadmap time period, and refer to the ERD chapter for detail about them.

FRONT END PROCESSES

There is strong linkage between the Front End Processes (FEP) and the PIDS chapters. Key areas of joint concern regarding planar bulk MOSFETs include the replacement of silicon oxy-nitride gate dielectric and polysilicon gate electrodes with high-κ dielectric and metal gate electrodes. Also, the challenge of keeping the parasitic series source/drain resistance within tolerable limits with scaling, and the difficult tradeoffs, including very high channel doping, required to set the threshold voltage and to control short-channel effects (SCEs) as scaling proceeds beyond about 2008. For ultra-thin body fully depleted SOI and multiple-gate MOSFETs, which are expected to be introduced beginning in 2008, some key issues are similar to those for planar bulk, such as high-κ gate dielectric and metal gate electrode and keeping the parasitic resistance within tolerable limits, but channel doping is not an issue, since these devices are essentially undoped. However, there are new issues, such as control of the very thin silicon body required for these devices, and designing and fabricating these devices for optimal operation. For DRAMs, key areas of joint concern include implementation of Metal Insulator Metal (MIM) storage capacitors with high-κ dielectric to scale the equivalent oxide thickness aggressively, as well as keeping the leakage of the access transistor ultra-low as the DRAM is scaled. For non-volatile memory, a key issue of joint concern involves the difficult tradeoffs in scaling the interpoly and the tunneling dielectrics in Flash memory.

REFERENCES

[1] T.Skotnicki, et al., "A new punchthrough current model based on the voltage-doping transformation," IEEE Transactions on Electron Devices, vol. 35, no. 7, pp. 1076–1086, June 1988.

[2] T. Skotnicki et al., "A new analog/digital CAD model for sub-half micron MOSFETs," Technical Digest of IEEE International Electron Devices Meeting, pp. 165–168, December 1994.

[3] T. Skotnicki and F. Boeuf, "CMOS Technology Roadmap – Approaching Up-hill Specials," in Proceedings of the 9th Intl. Symp. On Silicon Materials Science and Technology, Editors H.R. Huff, L. Fabry, S. Kishino, pp. 720–734, ECS Volume 2002-2.

[4] M. H. Nia et al., IEDM Technical Digest, p. 121, Dec. 2006.

[5] S. Takagi et al., "Channel Structure Design, Fabrication and Carrier Transport Properties of Strained-Si/SiGe-On-Insulator (Strained-SOI) MOSFETs," Technical Digest of IEEE International Electron Devices Meeting, pp. 57–60, December 2003.

[6]T. Ghani et al., "A 90 nm High Volume Manufacturing Logic Technology Featuring Novel 45nm Gate length Strained Silicon CMOS transistors," Technical Digest of IEEE International Electron Devices Meeting, pp. 978–980, December 2003.

[7] K. Rim et al., "Characteristics and Device Design of Sub-100 nm Strained Si N- and PMOSFETs," Symposium on VLSI Technology, pp. 98–99, June 2002.

[8] C. D. Sheraw et al., "Dual Stress Liner Enhancement in Hybrid Orientation Technology," Symposium on VLSI Technology, pp. 12–13, June 2005.

[9] B. Doris et al., "A Simplified Hybrid Orientation Technology (SHOT) for High Performance CMOS," Symposium on VLSI Technology, pp. 86–87, June 2004.

[10] S. Takagi et al., "Channel Structure Design, Fabrication and Carrier Transport Properties of Strained-Si/SiGe-On-Insulator (Strained-SOI) MOSFETs," Technical Digest of IEEE International Electron Devices Meeting, pp. 57–60, December 2003.

[11]T. Ghani et al., "A 90 nm High Volume Manufacturing Logic Technology Featuring Novel 45nm Gate length Strained Silicon CMOS transistors," Technical Digest of IEEE International Electron Devices Meeting, pp. 978–980, December 2003.

[12] K. Rim et al., "Characteristics and Device Design of Sub-100 nm Strained Si N- and PMOSFETs," Symposium on VLSI Technology, pp. 98–99, June 2002.

[13] C. D. Sheraw et al., "Dual Stress Liner Enhancement in Hybrid Orientation Technology," Symposium on VLSI Technology, pp. 12–13, June 2005.

[14] B. Doris et al., "A Simplified Hybrid Orientation Technology (SHOT) for High Performance CMOS," Symposium on VLSI Technology, pp. 86–87, June 2004.

[15] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer, and D. Finzi, "NROM: A Novel Localized Trapping, 2 bit Nonvolatile Memory Cell," IEEE Electron Device Lett., **21**, pp. 543–545, Nov. (2000).

[16] H. T. Lue, S. Y. Wang, E. K. Lai, Y. H. Shih, S. C. Lai, L. W. Yang, K. C. Chen, J. Ku, K. Y. Hsieh, R. Liu, and C. Y. Lu, "BE-SONOS: A Bandgap Engineered SONOS with Excellent Performance and Reliability," in *Tech. Digest 2005 International Electron Devices Meeting*, pp. 547-550, 2005.

[17] Y. Shin, J. Choi, C. Kang, C. Lee, K.T. Park, J.S. Lee, J. Sel, V. Kim, B. Choi, J. Sim, D. Kim, H.J. Cho and K. Kim, "A Novel NAND-type MONOS Memory using 63nm Process Technology for Multi-Gigabit Flash EEPROMs," *Tech. Digest 2005 International Electron Devices Meeting*, pp. 337-340, 2005.

[18] S-M. Jung, J. Jang, W. Cho, H. Cho, J. Jeong, Y. Chang, J. Kim, Y. Rah, Y. Son, J. Park, M-S. Song, K-H. Kim, J-S. Lim and K. Kim, "Three Dimensionally Stacked NAND Flash Memory Technology Using Stacking Single Crystal Si Layers on ILD and TANOS Structure for Beyond 30nm Node," *Tech. Digest 2006 International Electron Devices Meeting*, pp. 37-40, 2006.

[19] E. K. Lai, H. T. Lue, Y. H. Hsiao, J. Y. Hsieh, C. P. Lu, S. Y. Wang, L. W. Yang, T. H. Yang, K. C. Chen, J. Gong, K. Y. Hsieh, R. Liu and C. Y. Lu, "A Multi-Layer Stackable Thin-Film Transistor (TFT) NAND-Type Flash Memory," *Tech. Digest 2006 International Electron Devices Meeting*, pp. 41-44, 2006.

[20] H. Tanaka, M. Kido, K. Yahashi, M. Oomura, R. Katsumata, M. Kito, Y. Fukuzumi, M. Sato, Y. Nagata, Y. Matsuoka, Y. Iwata, H. Aochi and A. Nitayama, "Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory," *Digest of Technical Papers, 2007 Symposium on VLSI Technology*, pp. 14-15, 2007.

[21] Y. K. Hong, D. J. Jung, S. K. Kang, H. S. Kim, J. Y. Jung, H. K. Koh, J. H. Park, D. Y. Choi, S. E. Kim, W. S. Ann, Y. M. Kang, H. H. Kim, J.-H. Kim, W. U. Jung, E. S. Lee, S. Y. Lee, H. S. Jeong and K. Kim, "130 nm-technology, 0.25 μm^2, 1T1C FRAM Cell for SoC (System-on-a-Chip)-friendly Applications," *Digest of Technical Papers, 2007 Symposium on VLSI Technology*, pp. 230-231, 2007.

[22] K. Miura, T. Kawahara, R. Takemura, J. Hayakawa, S. Ikeda, H. Takahashi, H. Matsuoka and H. Ohno, "A novel SPRAM (SPin-transfer torque RAM) with a synthetic ferromagnetic free layer for higher immunity to read disturbance and reducing write-current dispersion," *Digest of Technical Papers, 2007 Symposium on VLSI Technology*, pp. 234-235, 2007.