INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

2005 EDITION

PROCESS INTEGRATION, DEVICES, AND STRUCTURES

TABLE OF CONTENTS

LIST OF FIGURES

MASTAR model

LIST OF TABLES

PROCESS INTEGRATION, DEVICES, AND STRUCTURES

SCOPE

The *Process Integration, Devices, and Structures (PIDS)* chapter deals with the full IC process flow and its overall integration, with the main IC devices and structures, and with the reliability tradeoffs associated with new options. Physical and electrical requirements and characteristics are included 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. The chapter deals with the following major topics: logic, memory (including both DRAM and non-volatile memory [NVM]), and reliability, each of which has its own section in the chapter. In addition, the chapter addresses key technical challenges facing the industry in this area, and it includes some of the best-known potential solutions to these challenges.

There are several key themes in the PIDS chapter of the 2005 International Technology Roadmap for Semiconductors (ITRS). One such theme is continued aggressive scaling of the MOSFETs for leading-edge logic technology in order to maintain historical trends of improved device performance. This aggressive 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. For NVM, there are serious issues with scaling that are dealt with at some length in the chapter. Numerous types of NVM are considered, including NAND and NOR flash, silicon-oxide-nitride-oxide-silicon (SONOS), ferroelectric RAM (FeRAM), and magnetic RAM (MRAM), which are all in mainstream production, although each has its own main application area. Phase change memory (PCM) has been added to the 2005 PIDS section because it is approaching mainstream production. For DRAM, the key issue is increasing scaling difficulties, especially with controlling leakage. For reliability, the key issue is ensuring the reliability of the numerous projected technological innovations in a timely manner.

Key aims of the ITRS include both 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.

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. 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

As mentioned above, NVM requirements and challenges are treated for several technologies, including Flash (NOR and NAND), FeRAM, SONOS, MRAM, and PCM. Read only memory (ROM) and one-time-programmable (OTP) technologies are excluded, since the discussion in this chapter is limited to NVM devices that can be written and read many times.

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 39a Process Integration Difficult Challenges—Near-term

Difficult Challenges ≥ 32 nm	*Summary of Issues*
1. Scaling of MOSFETs to the 32 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), stochastic doping variations, and difficulty in adequately controlling short channel effects. 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. 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.
2. Implementation of high-κ gate dielectric and metal gate electrode in a timely manner	High κ and metal gate electrode will be required beginning in ~2008. Timely implementation will involve dealing with numerous challenging issues, including appropriate tuning of metal gate work function, ensuring adequate channel mobility with high-κ, reducing the defects in high-κ to acceptable levels, ensuring reliability, and others.
3. Timely assurance for the reliability of multiple and rapid material, process, and structural changes	Multiple changes are projected over the next decade, such as.: Material: high-κ gate dielectric, metal gate electrodes by 2008 or so Process: elevated S/D (selective epi) and advanced annealing and doping techniques Structure: ultra-thin body (UTB) fully depleted (FD) SOI, followed by multiple-gate structures. 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 32 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. Also, reducing the cell area factor in a timely manner is quite challenging. (Cell area factor = a = cell area/F^2, where F=DRAM half pitch). 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 32 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. SONOS—ONO stack dimensions and material properties, including nitride layer trap distribution in space and energy MRAM—Magnetic material properties and dimensional control. Sensitivity to IC processing temperatures and conditions

Table 39b Process Integration Difficult Challenges—Long-term

Difficult Challenges<32 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 effectively scale MOSFETs to 11 nm gate length and below. 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-11 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 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. Toward the end of the Roadmap or beyond, identification, selection, and implementation of advanced, beyond-CMOS devices and architectures for advanced information processing	Will drive major changes in process, materials, device physics, design, etc. Performance, power dissipation, etc., of beyond-CMOS devices need to extend well beyond CMOS limits. Beyond-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 32 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 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, will be 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 a 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] Implementation of high-κ gate dielectric and metal gate electrode in a timely manner—According to projections in the PIDS tables, high-κ gate dielectric and metal gate electrode will be needed by 2008 to meet the scaling goals and at the same time to keep the gate leakage current within tolerable limits. There are numerous challenges associated with implementing such an advanced gate stack, including ensuring adequate channel carrier mobility with the high-κ dielectric, and reducing the defects, charge trapping, and instabilities in the high-κ to tolerable levels. Also, determining metal gate electrode materials with the appropriate work function for planar bulk CMOS devices (which require near band edge work functions) and ultra-thin body CMOS devices (which require near-midgap, tunable work functions), and integrating the various metal gate electrodes and high-κ gate dielectric into the CMOS process flow will be quite challenging. Finally, understanding and developing electrical characterization and reliability techniques for these new materials in a timely manner is a challenge.

[3] Timely assurance for the reliability of multiple and rapid material, process, and structural changes—In order to successfully scale MOSFETs and meet both device 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, etc., will need to be implemented in less than 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 MOSFETs with 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 32 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. Timely decreases in the cell area factor, a = (cell area)/F^2, where F is the Metal 1 half pitch, are critically important. 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. 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 32 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 predict 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 11 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 11 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 to 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] Toward the end of the Roadmap or beyond, identification, selection, and implementation of advanced, beyond-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 good metric for the intrinsic MOSFET delay, and hence $1/\tau$ is a good metric for the maximum intrinsic MOSFET switching frequency.

$1/\tau$ is used as the key transistor performance metric. To determine the projected parameter values in a table, a target is set for one of the key outputs, such as leakage current or $1/\tau$. 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://public.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 MOSFETs (such as FinFETs), are implemented in 2008 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. 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. 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 40a and b), the driver is the MOSFET 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 a critical 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 0.06 μA/μm in 2005, and it generally increases with succeeding years, which has important consequences for the chip power dissipation (to be discussed below).

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. These techniques have so far mainly been utilized for high-performance logic, but they may eventually be used for low-power logic as well. With scaling, it is expected that they will be more and more heavily exploited and perhaps more effective. In the next editions of the Roadmap, the Design and PIDS Working Groups will consider the impact of these and other architectural techniques on improving the chip performance for future technology generations, along with whether this architecture-based performance scaling suggests a slackening of 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 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 will be 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 10 pA/μm and is held (mainly) constant for LSTP, while it is 3 nA/μm for LOP in 2005, 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 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 41a and 41b for LSTP data. This issue 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 41c and 41d), 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. The scaling of $I_{sd,leak}$ and of τ is plotted in Figure 34 for high-performance, LOP, and LSTP logic. (τ is plotted rather than $1/\tau$ to enhance the clarity by reducing the clutter in the graph.) As expected, $I_{sd,leak}$ is highest and τ is lowest for high-performance logic, both parameters are intermediate for LOP logic, and $I_{sd,leak}$ is lowest and τ is highest for LSTP logic. For high-performance logic, the scaling of τ meets the 17% per year performance improvement target. Also note that $I_{sd,leak}$ for high-performance logic is consistently about four orders of magnitude larger than $I_{sd,leak}$ for LSTP logic.

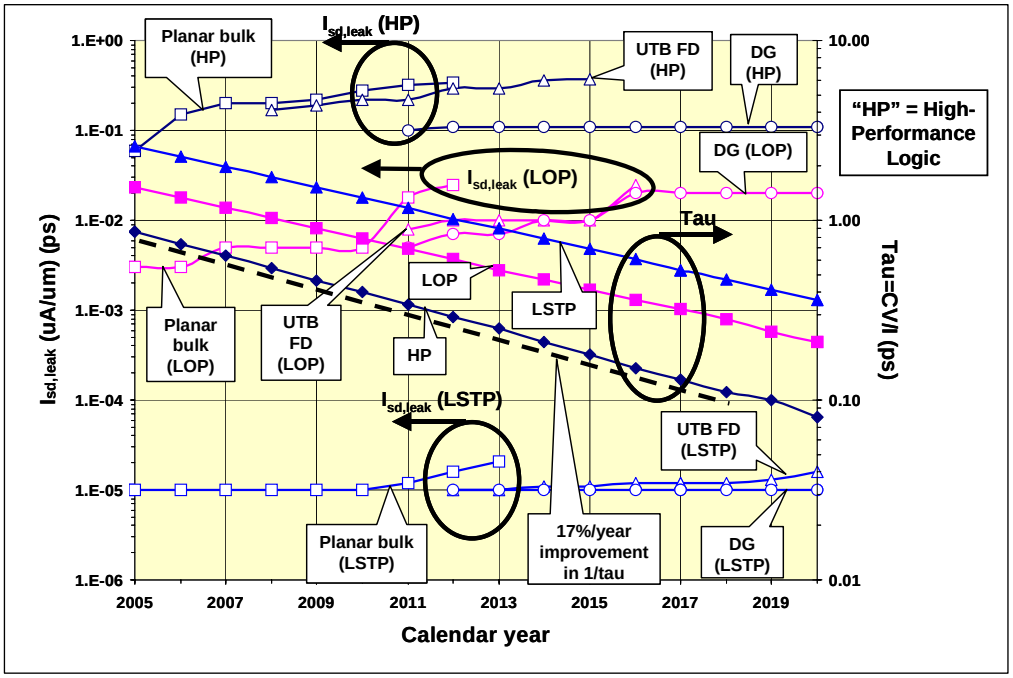


Figure 34 $\tau = CV/I$ and $I_{sd,leak}$ for All Logic Types.
(The dashed line represents the desired 17%/year transistor performance improvement.)

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 40a, 40b, 41a through 41d and to 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 from the tables. The results are shown in Figures 35 through 37, where "$J_{g,limit}$" is the gate leakage current density limit and "$J_{g,simulated}$" is the expected value of the gate leakage current density from the simulations. EOT is also plotted for reference. 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.

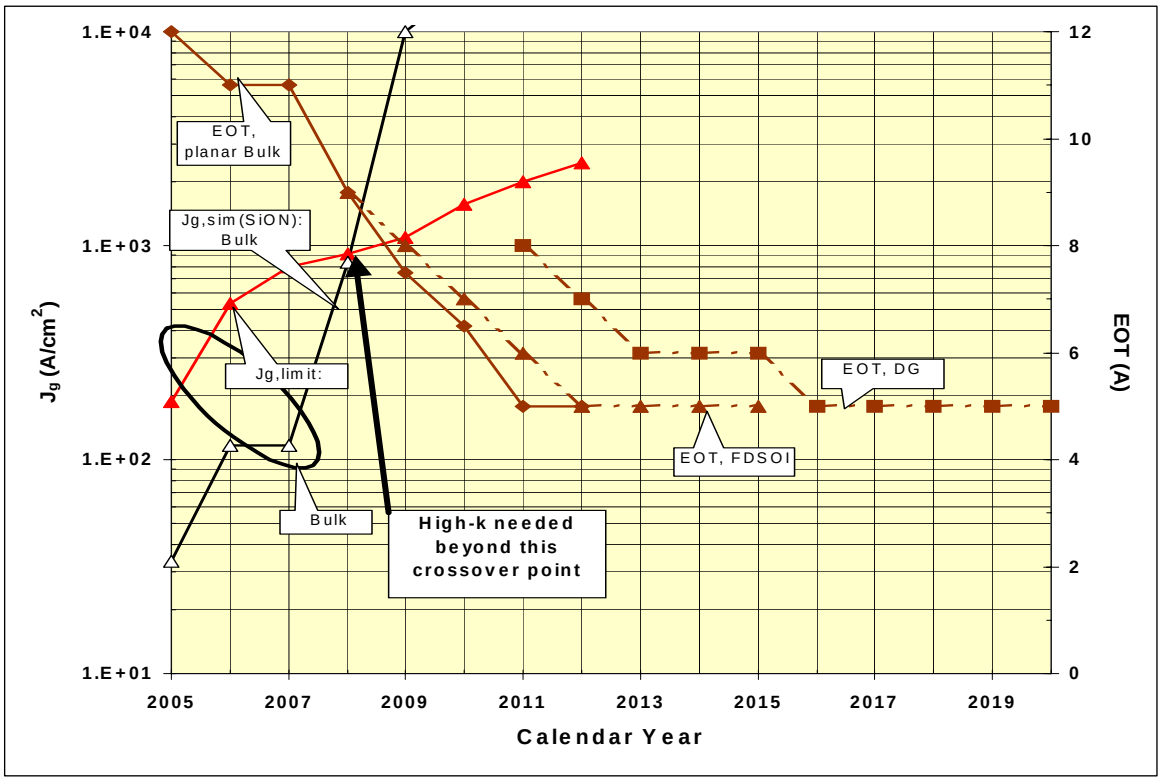


Figure 35 $J_{g,limit}$ versus $J_{g,simulated}$ for High-Performance Logic

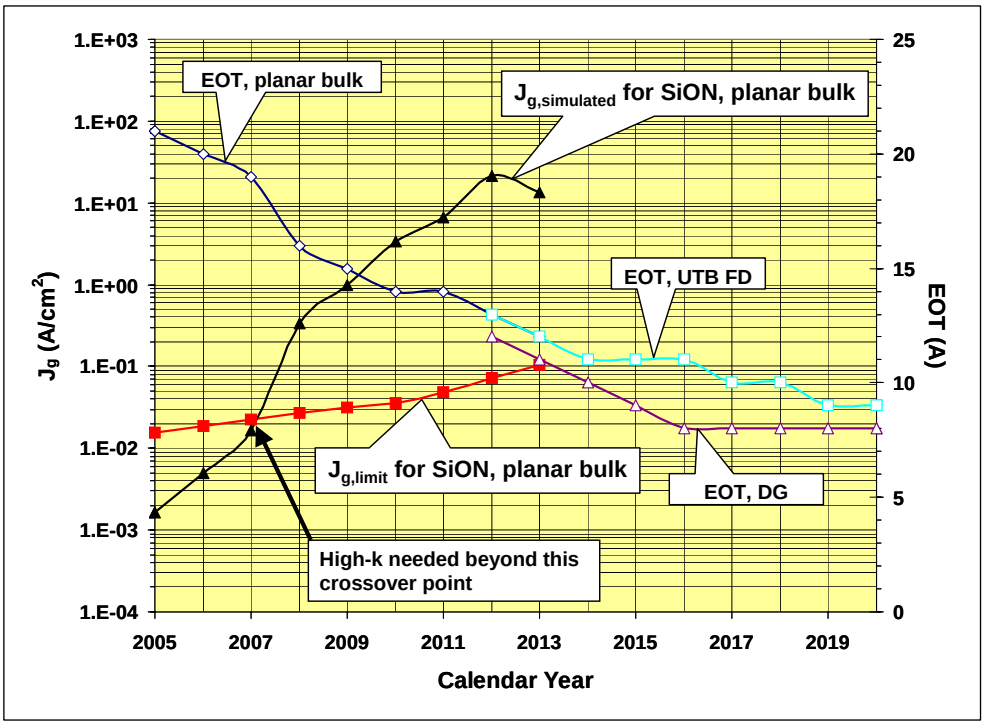


Figure 36 $J_{g,limit}$ versus $J_{g,simulated}$ for Low Standby Power

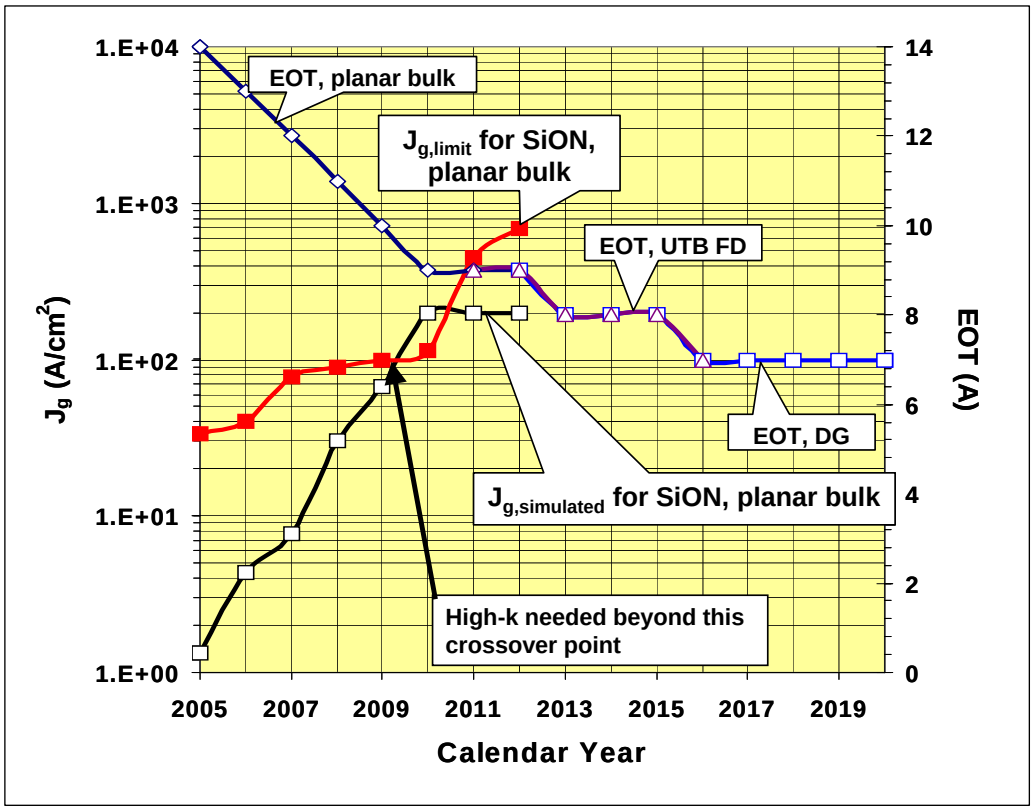


Figure 37 $J_{g,limit}$ versus $J_{g,simulated}$ for Low Operating Power

Table 40a High-Performance Logic Technology Requirements—Near-term

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	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*
DRAM ½ Pitch (nm) (contacted)	*80*	*70*	*65*	*57*	*50*	*45*	*40*	*36*	*32*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*90*	*78*	*68*	*59*	*52*	*45*	*40*	*36*	*32*
MPU Physical Gate Length (nm)	*32*	*28*	*25*	*22*	*20*	*18*	*16*	*14*	*13*
L_g: Physical L_{gate} for High Performance logic (nm) [1]	**32**	**28**	**25**	**22**	**20**	**18**	**16**	**14**	**13**
EOT: Equivalent Oxide Thickness [2]									
Extended planar bulk (Å)	**12**	**11**	**11**	**9**	**7.5**	**6.5**	**5**	**5**	
UTB FD (Å)				**9**	**8**	**7**	**6**	**5**	**5**
DG (Å)							**8**	**7**	**6**
Gate Poly Depletion and Inversion-Layer Thickness [3]									
Extended Planar Bulk (Å)	**7.3**	**7.4**	**7.4**	**2.9**	**2.8**	**2.7**	**2.5**	**2.5**	
UTB FD (Å)				**4**	**4**	**4**	**4**	**4**	**4**
DG (Å)							**4**	**4**	**4**
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]									
Extended Planar Bulk (Å)	**19.3**	**18.4**	**18.4**	**11.9**	**10.3**	**9.2**	**7.5**	**7.5**	
UTB FD (Å)				**13**	**12**	**11**	**10**	**9**	**9**
DG (Å)							**12**	**11**	**10**
$J_{g,limit}$: Maximum gate leakage current density [5]									
Extended Planar Bulk (A/cm^2)	**1.88E+02**	**5.36E+02**	**8.00E+02**	**9.09E+02**	**1.10E+03**	**1.56E+03**	**2.00E+03**	**2.43E+03**	
FDSOI (A/cm^2)				**7.73E+02**	**9.50E+02**	**1.22E+03**	**1.38E+03**	**2.07E+03**	**2.23E+03**
DG (A/cm^2)							**6.25E+02**	**7.86E+02**	**8.46E+02**
V_{dd}: Power Supply Voltage (V) [6]	**1.1**	**1.1**	**1.1**	**1**	**1**	**1**	**1**	**0.9**	**0.9**
$V_{t,sat}$: Saturation Threshold Voltage [7]									
Extended Planar Bulk (mV)	**195**	**168**	**165**	**160**	**159**	**151**	**146**	**148**	
UTB FD (mV)				**169**	**168**	**167**	**170**	**166**	**167**
DG (mV)							**181**	**184**	**185**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]									
Extended Planar Bulk (µA/µm)	**0.06**	**0.15**	**0.2**	**0.2**	**0.22**	**0.28**	**0.32**	**0.34**	
UTB FD (µA/µm)				**0.17**	**0.19**	**0.22**	**0.22**	**0.29**	**0.29**
DG (µA/µm)							**0.1**	**0.11**	**0.11**
$I_{d,sat}$: effective NMOS Drive Current [9]									
Extended Planar Bulk (µA/µm)	**1020**	**1130**	**1200**	**1570**	**1810**	**2050**	**2490**	**2300**	
UTB FD (µA/µm)				**1486**	**1625**	**1815**	**2015**	**2037**	**2198**
DG (µA/µm)							**1899**	**1932**	**2220**
Mobility Enhancement Factor for $I_{d,sat}$ [10]									
Extended Planar Bulk	**1.09**	**1.09**	**1.08**	**1.09**	**1.10**	**1.10**	**1.12**	**1.11**	
UTB FD				**1.06**	**1.06**	**1.06**	**1.06**	**1.05**	**1.05**
DG							**1.05**	**1.04**	**1.05**
Effective Ballistic Enhancement Factor [11]									
Extended Planar Bulk	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	
UTB FD				**1**	**1**	**1**	**1**	**1**	**1.1**
DG							**1.17**	**1.25**	**1.31**

Table 40a High-Performance Logic Technology Requirements—Near-term (continued)

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	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*
DRAM ½ Pitch (nm) (contacted)	*80*	*70*	*65*	*57*	*50*	*45*	*40*	*36*	*32*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*90*	*78*	*68*	*59*	*52*	*45*	*40*	*36*	*32*
MPU Physical Gate Length (nm)	*32*	*28*	*25*	*22*	*20*	*18*	*16*	*14*	*13*
R_{sd}: *Effective Parasitic series source/drain resistance [12]*									
Planar Bulk (Ω-µm)	180	170	140	140	120	105	80	70	
UTB FD (Ω-µm)				155	140	125	110	90	75
DG (Ω-µm)							110	100	90
$C_{g,ideal}$: *Ideal NMOS Device Gate Capacitance [13]*									
Extended Planar Bulk (F/µm)	5.73E-16	5.25E-16	4.69E-16	6.37E-16	6.72E-16	6.78E-16	7.39E-16	6.41E-16	
UTB FD (F/µm)				5.84E-16	5.75E-16	5.65E-16	5.52E-16	5.37E-16	4.98E-16
DG (F/µm)							4.60E-16	4.39E-16	4.48E-16
$C_{g,total}$: *Total gate capacitance for calculation of CV/I [14]*									
Extended Planar Bulk (F/µm)	8.13E-16	7.65E-16	6.99E-16	8.47E-16	8.42E-16	8.28E-16	8.59E-16	7.51E-16	
UTB FD (F/µm)				8.04E-16	7.55E-16	7.35E-16	6.92E-16	6.67E-16	6.18E-16
DG (F/µm)							6.50E-16	6.29E-16	6.28E-16
$\tau = CV/I$: *NMOSFET intrinsic delay (ps) [15]*	0.870	0.740	0.640	0.540	0.460	0.400	0.340	0.290	0.250
$1/\tau$: *NMOSFET intrinsic switching speed (GHz) [16]*	1149	1351	1563	1852	2174	2500	2941	3448	4000

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known

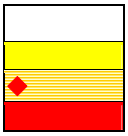


Table 40b High-Performance Logic Technology Requirements—Long-term

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	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
DRAM ½ Pitch (nm) (contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU Physical Gate Length (nm)	*11*	*10*	*9*	*8*	*7*	*6*	*6*
L_g: Physical L_{gate} for High Performance logic (nm) [1]	**11**	**10**	**9**	**8**	**7**	**6**	**5**
EOT: Equivalent Oxide Thickness [2]							
Extended planar bulk (Å)							
UTB FD (Å)	**5**	**5**					
DG (Å)	**6**	**6**	**5**	**5**	**5**	**5**	**5**
Gate Poly Depletion & Inversion-Layer 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)	**9**	**9**					
DG (Å)	**10**	**10**	**9**	**9**	**9**	**9**	**9**
$J_{g,limit}$: Maximum gate leakage current density [5]							
Extended Planar Bulk (A/cm^2)							
FDSOI (A/cm^2)	**3.27E+03**	**3.70E+03**					
DG (A/cm^2)	**1.00E+03**	**1.10E+03**	**1.22E+03**	**1.38E+03**	**1.57E+03**	**1.83E+03**	**2.20E+03**
V_{dd}: Power Supply Voltage (V) [6]	**0.9**	**0.8**	**0.8**	**0.7**	**0.7**	**0.7**	**0.7**
$V_{t,sat}$: Saturation Threshold Voltage [7]							
Extended Planar Bulk (mV)							
UTB FD (mV)	**164**	**166**					
DG (mV)	**190**	**192**	**195**	**200**	**201**	**205**	**208**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**0.36**	**0.37**					
DG (µA/µm)	**0.11**	**0.11**	**0.11**	**0.11**	**0.11**	**0.11**	**0.11**
$I_{d,sat}$: effective NMOS Drive Current [9]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**2290**	**2188**					
DG (µA/µm)	**2354**	**2275**	**2713**	**2533**	**2740**	**2744**	**2981**
Mobility Enhancement Factor for $I_{d,sat}$ [10]							
Extended Planar Bulk							
UTB FD	**1.04**	**1.04**					
DG	**1.04**	**1.04**	**1.04**	**1.04**	**1.03**	**1.03**	**1.03**
Effective Ballistic Enhancement Factor [11]							
Extended Planar Bulk							
UTB FD	**1.15**	**1.28**					
DG	**1.37**	**1.53**	**1.67**	**1.87**	**1.99**	**1.97**	**2.11**

Table 40b High-Performance Logic Technology Requirements—Long-term (continued)

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).

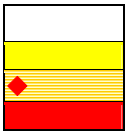
Year of Production	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
DRAM ½ Pitch (nm) (contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU Physical Gate Length (nm)	*11*	*10*	*9*	*8*	*7*	*6*	*6*
R_{sd}: Effective Parasitic series source/drain resistance [12]							
Planar Bulk (Ω-μm)							
UTB FD (Ω-μm)	75	75					
DG (Ω-μm)	85	80	75	70	65	60	55
Cg,ideal: Ideal NMOS Device Gate Capacitance [13]							
Extended Planar Bulk (F/μm)							
UTB FD (F/μm)	4.22E-16	3.83E-16					
DG (F/μm)	3.80E-16	3.45E-16	3.45E-16	3.07E-16	2.68E-16	2.30E-16	1.92E-16
$C_{g,total}$: Total gate capacitance for calculation of CV/I [14]							
Extended Planar Bulk (F/μm)							
UTB FD (F/μm)	5.42E-16	5.03E-16					
DG (F/μm)	5.59E-16	5.25E-16	5.25E-16	4.87E-16	4.48E-16	4.10E-16	3.62E-16
$\tau = CV/I$: NMOSFET intrinsic delay (ps) [15]	0.210	0.180	0.150	0.130	0.110	0.100	0.080
$1/\tau$: NMOSFET intrinsic switching speed (GHz) [16]	4762	5556	6667	7692	9091	10000	12500

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known



Notes for Tables 40a and b:

As described in the text, MASTAR, 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 this 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 CV/I. 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 2020. The impact of the challenges in scaling planar bulk are clear from this table, since for planar bulk, the Source/Drain subthreshold leakage current, $I_{sd,leak}$, increases sharply for the latter years, from 0.22 μa/μm in 2009 to 0.34 μa/μm in 2012, and $I_{sd,leak}$ is always higher for planar bulk than for UTB FD and DG. Furthermore, both EOT and the effective parasitic series resistance, R_{sd}, are scaled more rapidly (to meet the performance target) from 2008 through 2012 for planar bulk than for UTB FD or DG. 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 or DG. In a similar vein, for DG, $I_{sd,leak}$ is lower than for UTB FD, while EOT and R_{sd} are 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 2020.

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. For the planar bulk CMOS option, another goal was to delay the projected need for such major innovations as metal gate electrode, high-κ gate dielectric, and novel doping and annealing techniques to reduce the value of the parasitic series source/drain resistance. 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 ± 12% × 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] 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.) Note that the rate of scaling of EOT is quite 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. Red coloring for 2008 and beyond reflects the projected implementation of high-κ gate dielectric. The color is red because it is felt that the solutions for EOT below 1.0 nm are not understood. Measurement of EOT is complicated, and is usually done via sophisticated MOS capacitor-voltage (CV) measurements on MOS capacitors or via optical measurements.

[3] 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 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. The red color reflects the current lack of a well-known solution for metal gate electrodes with well-controlled work functions. For planar bulk CMOS, the work function needs to be near the silicon conduction band for NMOS and near the silicon valence band 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.

[4] EOT_{elec} is the sum of EOT and electrical thickness adjustment (see Notes [2] and [3] above). For MOSFETs in inversion, ideal gate capacitance per unit area (see Note [13]) 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 CV/I intrinsic delay (see Note [16]). Red/yellow coloring follows that of EOT and Electrical Thickness Adjustment.

[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 $I_{sd,leak}$, the nominal subthreshold leakage current per micron device width (see Note [8] below). Specifically, $J_{g,limit}$= [Initial Factor] x [$I_{sd,leak}$ / (physical gate length)] × [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 yellow color is associated with the very high substrate doping approaching or exceeding 5E18 cm^{-3} (from MASTAR) 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 yellow in 2008 because of the challenges of controlling the very thin silicon body thickness (T_{si}) required to control $V_{t,sat}$ and short channel effects. The color becomes red in 2009 when the required body thickness becomes less than ~7 nm. For DG devices, the color is red right from the beginning because there are very many 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 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). 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 12 below), the equivalent electrical oxide thickness in inversion (see Note 4), the required mobility/transconductance improvement factor (see Note 10), and the ballistic enhancement factor (see Note [11]).

[10] 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 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} = [enhanced mobility]/[reference mobility], where [enhanced mobility] is the actual mobility including the enhancement, and [reference mobility] is the mobility in the absence of enhancement. Generally, $I_{d,ratio}$ is significantly less than μ_{ratio} due to short channel effects and velocity saturation. Following the literature, the value of μ_{ratio} is limited to a maximum of 1.8[4]. Mobility enhancement was implemented in product in 2004[5] 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[6], different process induced strain approaches such as strained thin overlayers of SiN and selective epitaxial SiGe in the S/D, hybrid orientations, etc.[5,7,8]), and as we continue to scale MOSFETs, it is unclear what the optimal approach(es) will be and how to integrate them into the process flow. Consequently, the row is colored yellow in 2009, when L_g=20 nm and the scaling becomes difficult enough that the doping approaches 5E18 cm^{-3} according to the MASTAR modeling. For both FD SOI and DG, the row starts out yellow (in 2008 for FD and 2011 for DG) because we don't at this point understand manufacturable solutions to mobility enhancement for these device types.

[11] Effective Ballistic Enhancement Factor is a multiplying factor for $I_{d,sat}$, reflecting quasi-ballistic enhanced transport in highly scaled, ultra-thin body MOSFETs, both UTB FD SOI and DG MOSFETs. Planar bulk CMOS does not have ballistic enhancement because of the 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 initial yellow coloring reflects that quasi-ballistic enhancement is expected (and predicted by MASTAR) for undoped, very scaled UTB FD and DG MOSFETs. The later red coloring reflects the lack of known manufacturable enhanced transport solutions for transistors with gate length approaching 10 nm.

[12] R_{sd} is the maximum allowable parasitic series source plus drain resistance for a MOSFET of one micron 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.

[13] $C_{g,ideal}$ is the ideal gate capacitance per micron device width, in inversion. $C_{g,ideal}$ =[εox /(EOT_{elec})] × L_g, where εox is the dielectric constant of thermal silicon dioxide, EOTelec 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 EOTelec (see Note [4]).

[14] $C_{g,total}$ 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 $C_{g,ideal}$.

[15] τ is the intrinsic transistor delay for NMOS devices at 25°C. τ = ($C_{g,total}$ × 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 good metric for the intrinsic switching delay of the device, while 1/τ is a good 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 [14]).

[16] 1/τ is the NMOS intrinsic switching speed. Maintenance of the historical 17% per year device performance improvement scaling trend is the key scaling goal for high-performance logic. Red and yellow coloring follows that of τ.

Table 41a Low Standby Power Technology Requirements—Near-term

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 MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in Production	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*
DRAM ½ Pitch (nm) (contacted)	*80*	*70*	*65*	*57*	*50*	*45*	*40*	*36*	*32*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*90*	*78*	*68*	*59*	*52*	*45*	*40*	*36*	*32*
MPU Physical Gate Length (nm)	*32*	*28*	*25*	*22*	*20*	*18*	*16*	*14*	*13*
L_g: Physical gate length for LSTP [1]									
Extended Planar Bulk and DG (nm)	**65**	**53**	**45**	**37**	**32**	**28**	**25**	**22**	**20**
UTB FD (nm)								**22**	**20**
EOT: Equivalent Oxide Thickness [2]									
Extended planar bulk (Å)	**21**	**20**	**19**	**16**	**15**	**14**	**14**	**13**	**12**
UTB FD (Å)								**12**	**11**
DG (Å)								**13**	**12**
Gate Poly Depletion and Inversion-Layer Thickness [3]									
Extended planar bulk (Å)	**6.3**	**6.3**	**6.3**	**3.3**	**3.2**	**3.1**	**3.2**	**3.1**	**3.1**
UTB FD (Å)								**4**	**4**
DG (Å)								**4**	**4**
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]									
Extended planar bulk (Å)	**27.3**	**26.3**	**25.3**	**19.3**	**18.2**	**17.1**	**17.2**	**16.1**	**15.1**
UTB FD (Å)								**16**	**15**
DG (Å)								**17**	**16**
$J_{g,limit}$: Maximum gate leakage current density [5]									
Extended Planar Bulk (A/cm^2)	**1.5E-02**	**1.9E-02**	**2.2E-02**	**2.7E-02**	**3.1E-02**	**3.6E-02**	**4.8E-02**	**7.3E-02**	**1.1E-01**
UTB FD (A/cm^2)								**4.5E-02**	**5.0E-02**
DG (A/cm^2)								**4.5E-02**	**5.0E-02**
V_{dd}: Power Supply Voltage (V) [6]	**1.2**	**1.2**	**1.2**	**1.1**	**1.1**	**1.1**	**1**	**1**	**1**
$V_{t,sat}$: Saturation Threshold Voltage [7]									
Extended Planar Bulk (mV)	**482**	**515**	**524**	**501**	**501**	**502**	**502**	**491**	**483**
UTB FD (mV)								**483**	**486**
DG (mV)								**441**	**435**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]									
Extended Planar Bulk (µA/µm)	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.2E-05**	**1.6E-05**	**2.1E-05**
UTB FD (µA/µm)								**1.0E-05**	**1.0E-05**
DG (µA/µm)								**1.0E-05**	**1.0E-05**
$I_{d,sat}$: effective NMOS Drive Current [9]									
Extended Planar Bulk (µA/µm)	**497**	**500**	**519**	**573**	**612**	**666**	**580**	**625**	**684**
UTB FD (µA/µm)								**678**	**719**
DG (µA/µm)								**673**	**747**

Table 41a Low Standby Power Technology Requirements—Near-term (continued)

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 MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in Production	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*
DRAM ½ Pitch (nm) (contacted)	*80*	*70*	*65*	*57*	*50*	*45*	*40*	*36*	*32*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*90*	*78*	*68*	*59*	*52*	*45*	*40*	*36*	*32*
MPU Physical Gate Length (nm)	*32*	*28*	*25*	*22*	*20*	*18*	*16*	*14*	*13*
Mobility Enhancement Factor for $I_{d,sat}$ [10]									
Extended Planar Bulk	1.11	1.11	1.1	1.1	1.11	1.15	1.17	1.16	1.16
UTB FD								1.04	1.05
DG								1	1.04
Effective Ballistic Enhancement Factor [11]									
Extended Planar Bulk	1	1	1	1	1	1	1	1	1
UTB FD								1	1
DG								1	1
R_{sd}: Effective Parasitic series source/drain resistance [12]									
Extended Planar Bulk (Ω-µm)	180	180	180	180	180	180	170	170	160
UTB FD (Ω-µm)								180	180
DG (Ω-µm)								180	180
$C_{g,ideal}$: Ideal NMOS Device Gate Capacitance [13]									
Extended Planar Bulk (F/µm)	8.21E-16	6.96E-16	6.14E-16	6.62E-16	6.06E-16	5.64E-16	5.01E-16	4.70E-16	4.58E-16
UTB FD (F/µm)								4.74E-16	4.60E-16
DG (F/µm)								4.46E-16	4.31E-16
$C_{g,total}$: Total gate capacitance for calculation of CV/I [14]									
Extended Planar Bulk (F/µm)	1.06E-15	9.36E-16	8.54E-16	9.02E-16	8.46E-16	8.04E-16	6.81E-16	6.40E-16	6.18E-16
UTB FD (F/µm)								6.94E-16	6.50E-16
DG (F/µm)								6.86E-16	6.71E-16
τ = CV/I: NMOSFET intrinsic delay (ps) [15]	2.56	2.25	1.97	1.73	1.52	1.33	1.17	1.02	0.90
1/τ: NMOSFET intrinsic switching speed (GHz) [16]	391	444	508	578	658	752	855	980	1111

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

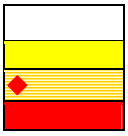


Table 41b Low Standby Power Technology Requirements—Long-term

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 MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in Production	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
DRAM ½ Pitch (nm) (contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU Physical Gate Length (nm)	*11*	*10*	*9*	*8*	*7*	*6*	*6*
L_g: Physical gate length for LSTP [1]							
Extended Planar Bulk and DG (nm)	**18**	**16**	**14**	**13**	**12**	**11**	**10**
UTB FD (nm)	**18**	**17**	**16**	**15**	**14**	**13**	**12**
EOT: Equivalent Oxide Thickness [2]							
Extended planar bulk (Å)							
UTB FD (Å)	**10**	**9**	**8**	**8**	**8**	**8**	**8**
DG (Å)	**11**	**11**	**11**	**10**	**10**	**9**	**9**
Gate Poly Depletion and Inversion-Layer Thickness [3]							
Extended planar bulk (Å)							
UTB FD (Å)	**4**	**4**	**4**	**4**	**4**	**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)	**14**	**13**	**12**	**12**	**12**	**12**	**12**
DG (Å)	**15**	**15**	**15**	**14**	**14**	**13**	**13**
$J_{g,limit}$: Maximum gate leakage current density [5]							
Extended Planar Bulk (A/cm^2)							
UTB FD (A/cm^2)	**6.1E-02**	**6.5E-02**	**7.5E-02**	**8.0E-02**	**8.6E-02**	**1.0E-01**	**1.3E-01**
DG (A/cm^2)	**5.6E-02**	**6.3E-02**	**7.1E-02**	**7.7E-02**	**8.3E-02**	**9.1E-02**	**1.0E-01**
V_{dd}: Power Supply Voltage (V) [6]	**1**	**1**	**1**	**1**	**1**	**1**	**1**
$V_{t,sat}$: Saturation Threshold Voltage [7]							
Extended Planar Bulk (mV)							
UTB FD (mV)	**486**	**489**	**487**	**487**	**492**	**488**	**486**
DG (mV)	**432**	**434**	**436**	**438**	**440**	**443**	**443**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**1.1E-05**	**1.1E-05**	**1.2E-05**	**1.2E-05**	**1.2E-05**	**1.30E-05**	**1.60E-05**
DG (µA/µm)	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**	**1.0E-05**
$I_{d,sat}$: effective NMOS Drive Current [9]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**773**	**882**	**1016**	**1108**	**1188**	**1289**	**1392**
DG (µA/µm)	**825**	**863**	**908**	**1011**	**1090**	**1192**	**1283**
Mobility Enhancement Factor for $I_{d,sat}$ [10]							
Extended Planar Bulk							
UTB FD	**1.06**	**1.06**	**1.05**	**1.05**	**1.05**	**1.04**	**1.04**
DG	**1.06**	**1.06**	**1.05**	**1.05**	**1.05**	**1.05**	**1.04**
Effective Ballistic Enhancement Factor [11]							
Extended Planar Bulk							
UTB FD	**1**	**1.08**	**1.15**	**1.24**	**1.32**	**1.4**	**1.48**
DG	**1**	**1**	**1.1**	**1.16**	**1.24**	**1.28**	**1.36**

Table 41b Low Standby Power Technology Requirements—Long-term (continued)

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 MOSFETs have reached the limits of practical scaling (see the text and the table notes for further discussion).

Year in Production	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
DRAM ½ Pitch (nm) (contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU Physical Gate Length (nm)	*11*	*10*	*9*	*8*	*7*	*6*	*6*
R_{sd}: *Effective Parasitic series source/drain resistance [12]*							
Extended Planar Bulk (Ω-µm)							
UTB FD (Ω-µm)	175	170	160	155	150	145	140
DG (Ω-µm)	180	175	170	165	160	155	150
$C_{g,ideal}$: *Ideal NMOS Device Gate Capacitance [13]*							
Extended Planar Bulk (F/µm)							
UTB FD (F/µm)	4.44E-16	4.51E-16	4.60E-16	4.31E-16	4.02E-16	3.74E-16	3.45E-16
DG (F/µm)	4.14E-16	3.68E-16	3.22E-16	3.20E-16	2.96E-16	2.92E-16	2.65E-16
$C_{g,total}$: *Total gate capacitance for calculation of CV/I [14]*							
Extended Planar Bulk (F/µm)							
UTB FD (F/µm)	6.14E-16	6.11E-16	6.20E-16	5.91E-16	5.63E-16	5.34E-16	5.05E-16
DG (F/µm)	6.54E-16	5.98E-16	5.52E-16	5.40E-16	5.16E-16	4.92E-16	4.65E-16
$\tau = CV/I$: *NMOSFET intrinsic delay (ps) [15]*	0.79	0.69	0.61	0.53	0.47	0.41	0.36
$1/\tau$: *NMOSFET intrinsic switching speed (GHz) [16]*	1266	1449	1639	1887	2128	2439	2778

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known

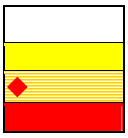


Table 41c Low Operating Power Technology Requirements—Near-term

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	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*
DRAM ½ Pitch (nm) (contacted)	*80*	*70*	*65*	*57*	*50*	*45*	*40*	*36*	*32*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*90*	*78*	*68*	*59*	*52*	*45*	*40*	*36*	*32*
MPU Physical Gate Length (nm)	*32*	*28*	*25*	*22*	*20*	*18*	*16*	*14*	*13*
L_g: Physical gate length for LOP (nm) [1]	**45**	**37**	**32**	**28**	**25**	**22**	**20**	**18**	**16**
EOT: Equivalent Oxide Thickness [2]									
Extended planar bulk (Å)	**14**	**13**	**12**	**11**	**10**	**9**	**9**	**9**	
UTB FD (Å)							**9**	**9**	**8**
DG (Å)							**9**	**9**	**8**
Gate Poly Depletion and Inversion-Layer Thickness [3]									
Extended planar bulk (Å)	**6.5**	**6.5**	**6.4**	**3.3**	**3.2**	**3.2**	**3.3**	**3.2**	
UTB FD (Å)							**4**	**4**	**4**
DG (Å)							**4**	**4**	**4**
EOT_{elec}: Electrical Equivalent Oxide Thickness in inversion [4]									
Extended planar bulk (Å)	**20.5**	**19.5**	**18.4**	**14.3**	**13.2**	**12.2**	**12.3**	**12.2**	
UTB FD (Å)							**13**	**13**	**12**
DG (Å)							**13**	**13**	**12**
$J_{g,limit}$: Maximum gate leakage current density [5]									
Extended Planar Bulk (A/cm^2)	**3.3E+01**	**4.1E+01**	**7.8E+01**	**8.9E+01**	**1.0E+02**	**1.1E+02**	**4.5E+02**	**6.9E+02**	
UTB FD (A/cm^2)							**2.0E+02**	**2.8E+02**	**3.1E+02**
DG (A/cm^2)							**1.3E+02**	**1.9E+02**	**2.2E+02**
V_{dd}: Power Supply Voltage (V) [6]	**0.9**	**0.9**	**0.8**	**0.8**	**0.8**	**0.7**	**0.7**	**0.7**	**0.6**
$V_{t,sat}$: Saturation Threshold Voltage [7]									
Extended Planar Bulk (mV)	**288**	**303**	**285**	**274**	**275**	**226**	**233**	**231**	
UTB FD (mV)							**273**	**268**	**272**
DG (mV)							**261**	**255**	**257**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]									
Extended Planar Bulk (µA/µm)	**3.0E-03**	**3.0E-03**	**5.0E-03**	**5.0E-03**	**5.0E-03**	**5.0E-03**	**1.8E-02**	**2.5E-02**	
UTB FD (µA/µm)							**8.0E-03**	**1.0E-02**	**1.0E-02**
DG (µA/µm)							**5.0E-03**	**7.0E-03**	**7.0E-03**
$I_{d,sat}$: effective NMOS Drive Current [9]									
Extended Planar Bulk (µA/µm)	**589**	**607**	**573**	**712**	**775**	**749**	**749**	**774**	
UTB FD (µA/µm)							**740**	**765**	**718**
DG (µA/µm)							**783**	**822**	**789**
Mobility Enhancement Factor for $I_{d,sat}$ [10]									
Extended Planar Bulk	**1.12**	**1.11**	**1.11**	**1.12**	**1.12**	**1.11**	**1.11**	**1.11**	
UTB FD							**1.07**	**1.06**	**1.06**
DG							**1.06**	**1.06**	**1.06**
Effective Ballistic Enhancement Factor [11]									
Extended Planar Bulk	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	
UTB FD							**1**	**1**	**1.26**
DG							**1.12**	**1.14**	**1.37**

Table 41c Low Operating Power Technology Requirements—Near-term (continued)

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	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*
DRAM ½ Pitch (nm) (contacted)	*80*	*70*	*65*	*57*	*50*	*45*	*40*	*36*	*32*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*90*	*78*	*68*	*59*	*52*	*45*	*40*	*36*	*32*
MPU Physical Gate Length (nm)	*32*	*28*	*25*	*22*	*20*	*18*	*16*	*14*	*13*
R_{sd}: *Effective Parasitic series source/drain resistance [12]*									
Planar Bulk (Ω-µm)	180	180	180	180	180	180	170	165	
UTB FD (Ω-µm)							145	140	135
DG (Ω-µm)							160	155	150
$C_{g,ideal}$: *Ideal NMOS Device Gate Capacitance [13]*									
Extended Planar Bulk (F/µm)	7.57E-16	6.55E-16	6.00E-16	6.76E-16	6.53E-16	6.20E-16	5.63E-16	5.09E-16	
UTB FD (F/µm)							5.31E-16	4.78E-16	4.25E-16
DG (F/µm)							5.31E-16	4.78E-16	4.60E-16
$C_{g,total}$: *Total gate capacitance for calculation of CV/I [14]*									
Extended Planar Bulk (F/µm)	9.97E-16	8.95E-16	8.40E-16	9.16E-16	8.73E-16	8.40E-16	7.43E-16	6.79E-16	
UTB FD (F/µm)							7.31E-16	6.68E-16	6.40E-16
DG (F/µm)							7.71E-16	7.18E-16	7.00E-16
$\tau = CV/I$: *NMOSFET intrinsic delay (ps) [15]*	1.52	1.33	1.17	1.03	0.90	0.79	0.69	0.61	0.53
$1/\tau$: *NMOSFET intrinsic switching speed (GHz) [16]*	658	752	855	971	1111	1266	1449	1639	1887

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known

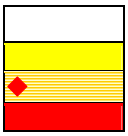


Table 41d Low Operating Power Technology Requirements—Long-term

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).

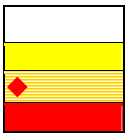
Year in Production	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
DRAM ½ Pitch (nm) (contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU Physical Gate Length (nm)	*11*	*10*	*9*	*8*	*7*	*6*	*6*
L_g: Physical gate length for LOP (nm) [1]	**14**	**13**	**11**	**10**	**9**	**8**	**7**
EOT: Equivalent Oxide Thickness [2]							
Extended planar bulk (Å)							
UTB FD (Å)	**8**	**8**	**7**				
DG (Å)	**8**	**8**	**7**	**7**	**7**	**7**	**7**
Gate Poly Depletion and Inversion-Layer Thickness [3]							
Extended planar bulk (Å)							
UTB FD (Å)	**4**	**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)	**12**	**12**	**11**				
DG (Å)	**12**	**12**	**11**	**11**	**11**	**11**	**11**
$J_{g,limit}$: Maximum gate leakage current density [5]							
Extended Planar Bulk (A/cm^2)							
UTB FD (A/cm^2)	**3.6E+02**	**3.8E+02**	**1.1E+03**				
DG (A/cm^2)	**3.6E+02**	**3.8E+02**	**9.1E+02**	**1.0E+03**	**1.1E+03**	**1.3E+03**	**1.4E+03**
V_{dd}: Power Supply Voltage (V) [6]	**0.6**	**0.6**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
$V_{t,sat}$: Saturation Threshold Voltage [7]							
Extended Planar Bulk (mV)							
UTB FD (mV)	**275**	**277**	**254**				
DG (mV)	**250**	**251**	**238**	**239**	**242**	**243**	**246**
$I_{sd,leak}$: Source/Drain Subthreshold Off-State Leakage Current [8]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**1.0E-02**	**1.0E-02**	**2.5E-02**				
DG (µA/µm)	**1.0E-02**	**1.0E-02**	**2.0E-02**	**2.0E-02**	**2.0E-02**	**2.0E-02**	**2.0E-02**
$I_{d,sat}$: effective NMOS Drive Current [9]							
Extended Planar Bulk (µA/µm)							
UTB FD (µA/µm)	**738**	**796**	**695**				
DG (µA/µm)	**829**	**892**	**760**	**820**	**873**	**929**	**931**
Mobility Enhancement Factor for $I_{d,sat}$ [10]							
Extended Planar Bulk							
UTB FD	**1.05**	**1.05**	**1.04**				
DG	**1.06**	**1.06**	**1.06**	**1.06**	**1.05**	**1.05**	**1.05**
Effective Ballistic Enhancement Factor [11]							
Extended Planar Bulk							
UTB FD	**1.28**	**1.37**	**1.39**				
DG	**1.38**	**1.47**	**1.59**	**1.7**	**1.8**	**1.9**	**1.92**

Table 41d Low Operating Power Technology Requirements—Long-term (continued)

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	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
DRAM ½ Pitch (nm) (contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU Physical Gate Length (nm)	*11*	*10*	*9*	*8*	*7*	*6*	*6*
R_{sd}: Effective Parasitic series source/drain resistance [12]							
Planar Bulk (Ω-μm)							
UTB FD (Ω-μm)	130	125	90				
DG (Ω-μm)	145	140	130	125	120	115	115
$C_{g,ideal}$: Ideal NMOS Device Gate Capacitance [13]							
Extended Planar Bulk (F/μm)							
UTB FD (F/μm)	4.02E-16	3.74E-16	3.45E-16				
DG (F/μm)	4.02E-16	3.74E-16	3.45E-16	3.14E-16	2.82E-16	2.51E-16	2.20E-16
$C_{g,total}$: Total gate capacitance for calculation of CV/I [14]							
Extended Planar Bulk (F/μm)							
UTB FD (F/μm)	5.83E-16	5.44E-16	5.05E-16				
DG (F/μm)	6.43E-16	6.14E-16	5.55E-16	5.24E-16	4.82E-16	4.41E-16	4.00E-16
$\tau = CV/I$: NMOSFET intrinsic delay (ps) [15]	0.47	0.41	0.36	0.32	0.28	0.24	0.21
$1/\tau$: NMOSFET intrinsic switching speed (GHz) [16]	2128	2439	2778	3125	3571	4167	4762

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



Notes for Tables 41a through 41d (LSTP and LOP):

As described in the text, MASTAR, 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 this 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 CV/I. 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 2020. 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 so 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 UTB FD devices, late in the ITRS, L_g scaling lags slightly behind that for DG MOSFETs because of the difficulty in scaling UTB FD MOSFETs 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 ± 12% × 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] 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. Red coloring from 2008 on reflects the projected implementation of high-κ gate dielectric, due to the inability of silicon oxy-nitride gate dielectric to meet the gate leakage current density limits for those years. The red color reflects the current lack of a well-known solution for high-κ gate dielectric with metal gate electrode, which is also projected for 2008 (see Note 3). Measurement of EOT is complicated, and is usually done via sophisticated MOS capacitor-voltage (CV) measurements on MOS capacitors or via optical measurements.

[3] 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. The red color for metal gate electrodes reflects the current lack of a well-known solution for metal gate electrodes with well-controlled and tunable work functions. For planar bulk CMOS, the work function needs to be near the silicon conduction band for NMOS and near the silicon valence band 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.

[4] EOT_{elec} is the sum of EOT and electrical thickness adjustment (see Notes [2] and [3] above). For MOSFETs in inversion, ideal gate capacitance per unit area (see Note [14]) is ε_{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 CV/I intrinsic delay (see Note [16]). Red/yellow coloring follows that of EOT and Electrical Thickness Adjustment.

[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 $I_{sd,leak}$, the nominal subthreshold leakage current per micron device width (see Note [8] below). Specifically, $J_{g,limit}$= [$I_{sd,leak}$ / (physical gate length)] × [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 the very high substrate doping approaching or exceeding 5E18 cm^{-3} (from MASTAR) 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.

[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.

[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 approximate 17% per year device performance scaling (see Note 16 below). 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 12 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 11).

[10] 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 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} = [enhanced mobility]/[reference mobility], where [enhanced mobility] is the actual mobility including the enhancement, and [reference mobility] is the mobility in the absence of enhancement. Generally, $I_{d,ratio}$ is significantly less than μ_{ratio} due to short channel effects and velocity saturation. Following the literature, the value of μ_{ratio} is limited to a maximum of 1.8[4]. Mobility enhancement was implemented in product in 2004[5] 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[6], different process induced strain approaches such as strained thin overlayers of SiN and selective epitaxial SiGe in the S/D, hybrid orientations, etc.[5,7,8]), and as we continue to scale MOSFETs, it is unclear what the optimal approach(es) will be and how to integrate them into the process flow. Consequently, for both LSTP and LOP the row is colored yellow when L_g=20 nm and the scaling becomes difficult enough that the doping approaches 5E18 cm^{-3} according to the MASTAR modeling. This occurs in 2013 for LSTP and in 2011 for LOP. For both FD SOI and DG, the row starts out yellow because we don't at this point understand manufacturable solutions to mobility enhancement for these device types.

[11] Effective Ballistic Enhancement Factor is a multiplying factor for $I_{d,sat}$, reflecting quasi-ballistic enhanced transport in highly scaled, ultra-thin body MOSFETs, both UTB FD SOI and DG MOSFETs. Planar bulk CMOS does not have ballistic enhancement because of the 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 initial yellow coloring reflects that quasi-ballistic enhancement is expected (and predicted by MASTAR) for undoped, very scaled UTB FD and DG MOSFETs. The later red coloring reflects the lack of known manufacturable enhanced transport solutions for transistors with gate length approaching 10nm.

[12] R_{sd} is the maximum allowable parasitic series source plus drain resistance for a MOSFET of one micron 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.

[13] $C_{g,ideal}$ 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, EOTelec 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 EOTelec (see Note [4]).

[14] $C_{g,total}$ 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 $C_{g,ideal}$.

[15] 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 good metric for the intrinsic switching delay of the device, while $1/\tau$ is a good metric for the intrinsic switching speed of the device. Red/yellow coloring follows that of both saturation drive current (see Note [9]) and Cg,total (see Note [14]).

[16] $1/\tau$ 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. 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 listed in Figure 38 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 six years: enhanced mobility (already implemented), high-κ gate dielectric, metal gate electrodes, and ultra-thin body, fully depleted SOI and multiple-gate MOSFETs with quasiballistic enhanced transport.

The first potential solution, enhanced mobility, is needed to enhance the saturation current drive to meet transistor performance targets. It was implemented in 2004 for high-performance logic, and it is projected to be implemented in 2005 for low-power 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 eventually, use of SiGe or Ge channels. The potential solutions table 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 optimally integrate mobility enhancement with the overall process flow, and eventually to utilize mobility enhancement for advanced MOSFETs such as UTB SOI and multiple-gate MOSFETs.

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 device gate stack (the combination of the gate dielectric and the corresponding electrode) need to be addressed. As the physical gate length is scaled, ideally the gate dielectric

equivalent oxide thickness is scaled correspondingly to control short-channel effects and to increase the saturation current drive. However, continued thinning of the currently conventional gate dielectric, silicon oxy-nitride, results in a significant increase in gate leakage current due to an approximately exponential increase in the direct tunneling current. 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 section on logic technology requirements, including Tables 40a and b and 41a through 41d, and Figures 35–37, 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 for PMOSFETs and near the silicon conduction band for NMOSFETs. Hence, different metals will probably be needed for the PMOSFET and NMOSFET.

As scaling proceeds to the 65 nm technology generation in 2007 (physical gate length = 25 nm) and beyond, it is expected to 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 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 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 is 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 2008 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 operate in a quasi-ballistic mode, with enhanced transport due to reduced scattering, and hence enhanced saturation current drive. 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.

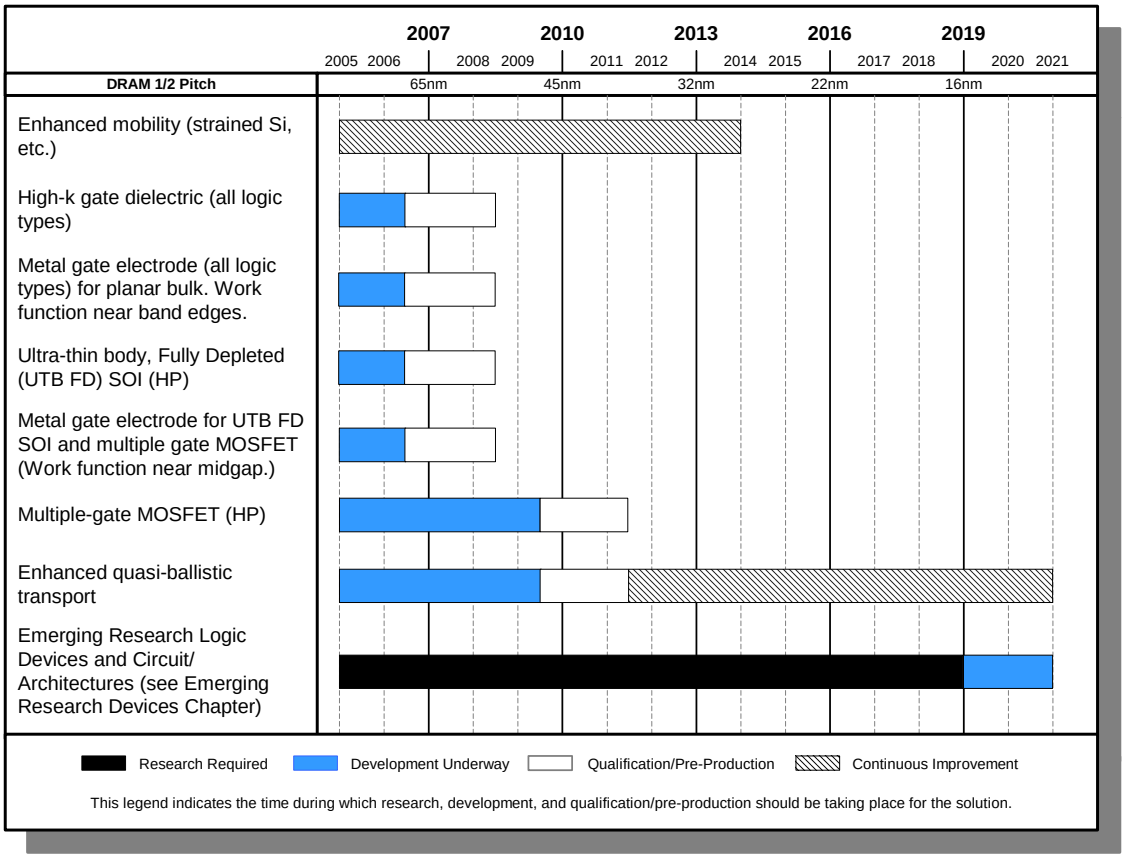


Figure 38 Logic Potential Solutions

MEMORY TECHNOLOGY REQUIREMENTS AND POTENTIAL SOLUTIONS

DRAM

Technical requirements for DRAMs become more difficult with scaling (see Table 42a and b). The process associated with 193 nm argon fluoride (ArF) immersion lithography technology is the key for 70 nm or smaller half pitch (hp) 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. 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. As a result, there are reliability concerns due to high electric field across the gate insulator. 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 39. Because of the gate insulator reliability concerns, the electric field in the dielectric is held approximately constant with scaling. Process requirements for DRAMs such as front end isolation, low resistance materials for the word lines, self-aligned and high aspect ratio etches, and planarization 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 (k) will be needed. Several manufacturers have introduced MIS (Metal Insulator Semiconductor) capacitors using Ta_2O_5 and Al_2O_3 (k~ 10–25) for DRAMs with 80 nm ½ pitch in 2005. Eventually, beyond 2006, MIM (Metal Insulator Metal) structures and dielectric materials with even higher k values than Ta_2O_5 and Al_2O_3 will likely be required. Finally, it is expected that very high κ values of 50 and greater will be needed later in the Roadmap (See Figure 41, 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. With the scaling of the EOT of the cell node capacitor insulator, the electric field across this insulator becomes higher, resulting in insulator reliability concerns. The scaling of DRAM storage node cell dielectric, DRAM storage node capacitor voltage, and electric field of the capacitor dielectric is plotted in Figure 40. As shown in the figure, the electric field in the capacitor dielectric is expected to increase sharply with scaling.

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 is critically important, along with the overall technology scaling. One company has started production of DRAMs with an *a* of 6, but the other companies are staying with an *a* of 8 for now. The year of production start of DRAM with "*a*" of 6 by the other companies is estimated to be after 2007. 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. The 2005 version of the DRAM Table 42a and b doesn't include $a = 4$ because a $4F^2$ memory storage cell structure is not considered feasible.

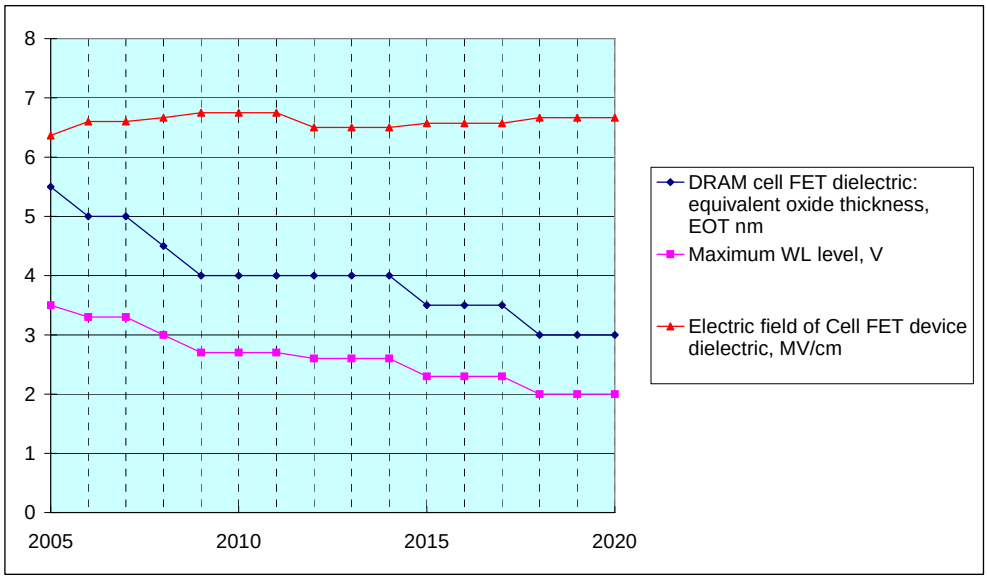


Figure 39 Cell FET Devices

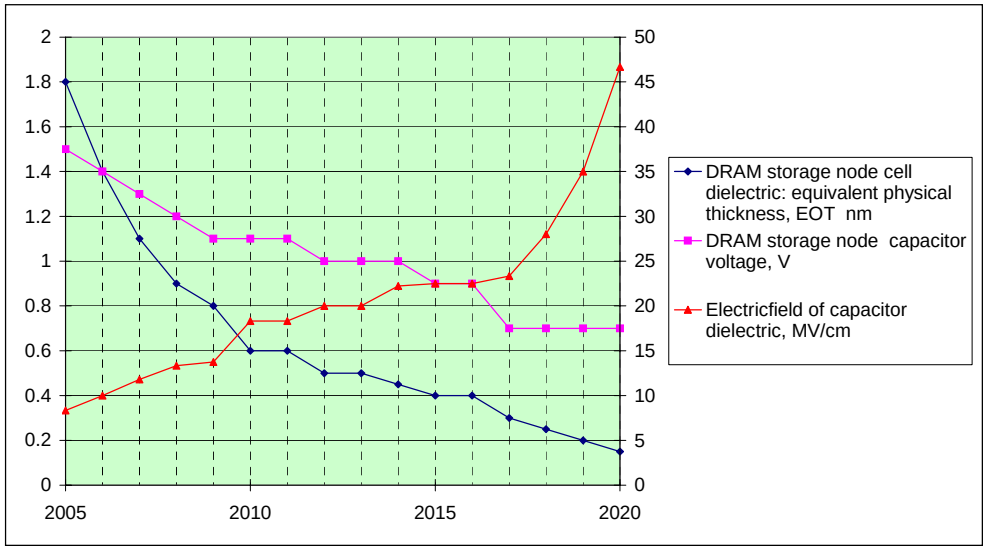


Figure 40 Storage Node Capacitor

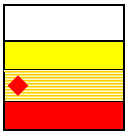
Table 42a DRAM Technology Requirements—Near-term

Year in Production	2005	2006	2007	2008	2009	2010	2011	2012	2013
DRAM ½ Pitch (nm) [1]	80	70	65	57	50	45	40	35	32
DRAM cell size (μm^2) [2]	0.0514	0.0408	0.0324	0.0193	0.0153	0.0122	0.0096	0.0077	0.0061
DRAM storage node cell capacitor dielectric: equivalent oxide thickness EOT (nm) [3]	1.8	1.4	1.1	0.9	0.8	0.6	0.6	0.5	0.5
DRAM storage node cell capacitor voltage (V) [4]	1.5	1.4	1.3	1.2	1.1	1.1	1.1	1	1
Electric field of capacitor dielectric, (MV/cm) [5]	8	10	12	13	14	18	18	20	20
DRAM cell FET dielectric: equivalent oxide thickness, EOT (nm) [6]	5.5	5	5	4.5	4	4	4	4	4
Maximum Wordline (WL) level (V) [7]	3.5	3.3	3.3	3	2.7	2.7	2.7	2.6	2.6
Electric field of cell FET device dielectric (MV/cm) [8]	6.4	6.6	6.6	6.7	6.8	6.8	6.8	6.5	6.5
Cell Size Factor: a [9]	8	8	8	6	6	6	6	6	6
Array Area Efficiency [10]	0.63	0.63	0.63	0.56	0.56	0.56	0.56	0.56	0.56
Minimum DRAM retention time (ms) [11]	64	64	64	64	64	64	64	64	64
DRAM soft error rate (fits) [12]	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table 42b DRAM Technology Requirements—Long-term

Year in Production	2014	2015	2016	2017	2018	2019	2020
DRAM ½ Pitch (nm) [1]	28	25	22	20	18	16	14
DRAM cell size (μm^2) [2]	0.0048	0.0038	0.0030	0.0024	0.0019	0.0015	0.0012
DRAM storage node cell dielectric: equivalent physical thickness EOT (nm) [3]	0.45	0.4	0.4	0.3	0.25	0.2	0.15
DRAM storage node capacitor voltage (V) [4]	1	0.9	0.9	0.7	0.7	0.7	0.7
Electric field of capacitor dielectric, (MV/cm) [5]	22	23	23	23	28	35	47
DRAM cell FET dielectric: equivalent oxide thickness, EOT (nm) [6]	4	3.5	3.5	3.5	3	3	3
Maximum Wordline (WL) level (V)1 [7]	2.6	2.3	2.3	2.3	2	2	2
Electric field of cell FET device dielectric (MV/cm) [8]	6.5	6.6	6.6	6.6	6.7	6.7	6.7
Cell Size Factor: a [9]	6	6	6	6	6	6	6
Array Area Efficiency [10]	0.56	0.56	0.56	0.56	0.56	0.56	0.56
Minimum DRAM retention time (ms) [11]	64	64	64	64	64	64	64
DRAM soft error rate (fits) [12]	1000	1000	1000	1000	1000	1000	1000

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



Notes for Tables 42a 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 2004 ITRS due to no further speed up in the pace of DRAM half pitch scaling during 2004.
[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 8 for DRAM ½ pitch of 80- 65 nm and 6 for 57 nm and beyond,

[3] Storage node cell dielectric EOT is defined as (dielectric physical thickness / [k/3.9]), where k 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 the 65 nm technology generation in 2007, the dielectric material is based on Al_2O_3 or Ta_2O_5 with MIS structure, and hence the color is white. Beyond2007, breakthroughs such as MIM structure and higher k material are needed, so the color is yellow. Finally, for the 45 nm technology generation 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. Trench capacitors have other requirements for the cell dielectric material.

[4] The DRAM storage node capacitor voltage is driven by two opposing needs. In conjunction with the storage node capacitance, which is inversely dependent on EOT (see note [3]), this voltage should be large enough that the stored charge is tolerably large. On the other hand, the voltage must be low enough that the resulting electric field in the dielectric (see Note [5]) is within acceptable limits.

[5] The electric field in the capacitor dielectric is (DRAM storage node capacitor voltage / DRAM storage node dielectric equivalent oxide thickness, EOT). Due to the sharp increase in the field with scaling, the color turns yellow in 2008, when the electric field is 13 MV/cm, and red in 2010, when the field becomes 18 MV/cm.

[6] DRAM cell FET dielectric EOT is defined as (dielectric physical thickness / [k/3.9]), where k 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 7) and the need to keep the electric field in the dielectric within tolerable limits (see Note 8)

[7] Maximum wordline 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.

[8] The electric field in the cell FET device dielectric is (maximum wordline level / DRAM cell FET dielectric equivalent oxide thickness, EOT).

[9] Cell size factor = a = (DRAM cell size/F^2), where F is the DRAM ½ pitch. The required values of a are 8 for DRAM ½ pitch of 80–65 nm and 6 for 57 nm and beyond. In contrast to previous versions, the 2005 version of the DRAM table doesn't have a = 4 because a $4F^2$ cell structure is considered to be unrealistic.

[10] 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 after 2007, the array area efficiency is decreased to 0.56, assuming the same relative peripheral circuit area.

[11] 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.

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

DRAM POTENTIAL SOLUTIONS

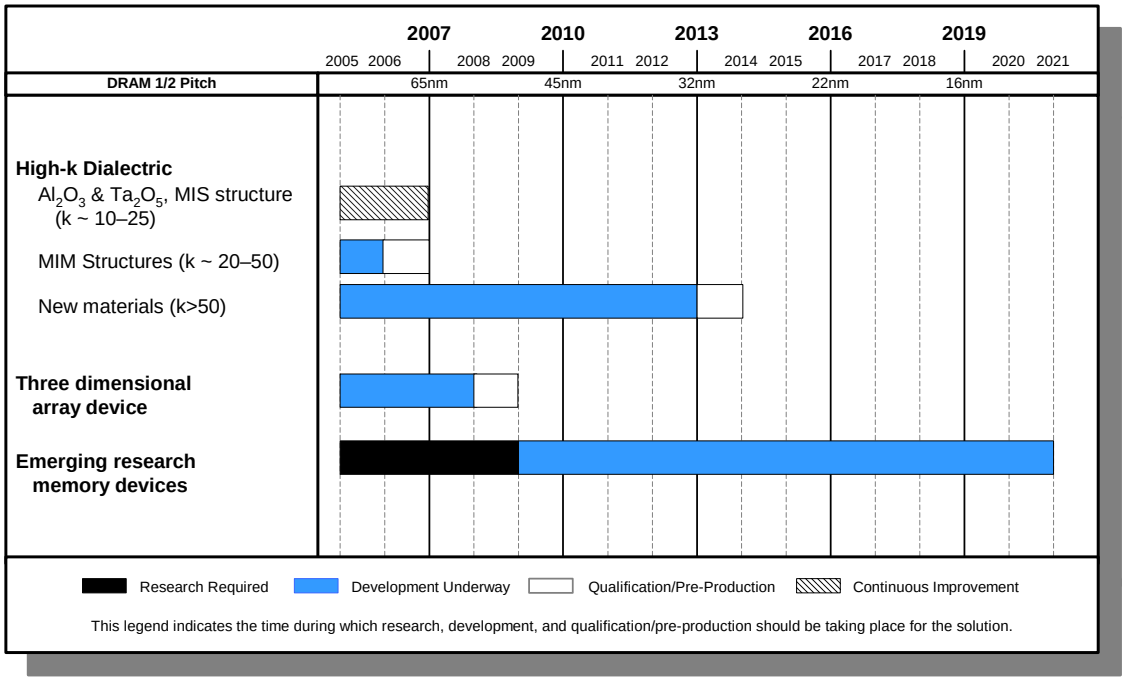


Figure 41 DRAM Potential Solutions

NON-VOLATILE MEMORY TECHNOLOGY REQUIREMENTS

Nonvolatile memory technologies are in a general sense a combination of a CMOS structure and a memory element structure. The progression through sequential technology generations is a bit more complicated than the basic CMOS scaling problem because the requirements of the memory element impose additional constraints on process integration and structure design. Requirements are presented in Tables 43a and 43b for Flash (NOR and NAND), FeRAM, SONOS, MRAM and phase change RAM (PCRAM) technologies.

Historically the NVM devices in a given time period have not been referenced to the then-current CMOS feature size, F (whose characteristic feature size is the DRAM ½ pitch), but have lagged behind by one or more years. The tables identify both the current CMOS feature size and the feature size actually used to form the NVM cells (i.e., the NVM technology "F" in nanometers, where F is the polysilicon half pitch). Depending on the particular NVM technology, the time lag from the CMOS to the NVM feature size is expected to reduce and eventually vanish. For floating gate NAND and NOR technologies the time lag not only completely disappears but also in some cases (NAND Flash) the NVM is more aggressive in driving some design rules than logic and DRAM counterparts in the same year.

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, the expected area factor "a" is given (cell size in terms of the number of F^2 units required), and then a typical cell size in micrometers squared is computed. Second, the tabulation presents a number of parameters important to each specific technology such as gate lengths, write-erase voltage maximums, 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.

The technical challenges for each technology differ depending on the nature of the memory element and the degree of compatibility with an underlying CMOS technology. Flash devices transfer charge to and from a floating gate that is isolated by surrounding dielectric materials. Ferroelectric memory operates by switching and sensing the polarization state of a ferroelectric capacitor. Silicon-oxide-nitride-oxide-silicon memory transfers charge to and from traps in a silicon nitride layer. Magnetic RAM switches the direction of magnetic spin in a layer of stacked magnetic materials that form a magnetic tunnel junction (MTJ) and senses the resultant resistance of the junction. Phase change RAM detects the

resistance change of chalcogenide glass when it is switched between amorphous and crystalline states. All nonvolatile memory technologies, however, face steep scaling challenges as volatility increases when the number of stored charges and the volume of state switching decreases.

Table 43a Non-Volatile Memory Technology Requirements—Near-term

Year of Production	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*
DRAM ½ Pitch (nm) (contacted)	*80*	*70*	*65*	*57*	*50*	*45*	*40*	*36*	*32*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*90*	*78*	*68*	*59*	*52*	*45*	*40*	*36*	*32*
MPU Physical Gate Length (nm)	*32*	*28*	*25*	*22*	*20*	*18*	*16*	*14*	*13*
Flash technology NOR/NAND – F (nm) [1]	80/76	70/64	65/57	57/51	50/45	45/40	40/36	35/32	32/28
Flash NOR cell size – area factor a in multiples of F^2 [2], [3], [4], [5]	9–11	9–11	9–11	9–12	10–12	9–12	9–12	10–12	10–12
Flash NAND cell size – area factor a in multiples of F^2 SLC/MLC [6]	4.0/2.0	4.0/2.0	4.0/2.0	4.0/2.0	4.0/2.0	4.0/1.0	4.0/1.0	4.0/1.0	4.0/1.0
Flash NOR typical cell size (μm^2) [7], [8]	0.064	0.049	0.042	0.034	0.028	0.021	0.017	0.013	0.011
Flash NOR L_g-stack (physical – µm) [8], [9]	0.14	0.135	0.13	0.12	0.12	0.11	0.11	0.1	0.1
Flash NOR highest W/E voltage (V) [10], [11]	7-9	7-9	7-9	7-9	7-9	6-8	6-8	6-8	6-8
Flash NAND highest W/E voltage (V) [12]	17–19	17–19	15–17	15–17	15–17	15–17	15-17	15–17	15–17
Flash NOR I_{read} (µA) [13]	29–37	28–36	27–35	26–34	25–33	27–33	27–33	26–32	25–31
Flash coupling ratio [14]	0.65–0.75	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
Flash NOR tunnel oxide thickness EOT (nm) [15]	8–9	8–9	8–9	8–9	8–9	8	8	8	8
Flash NAND tunnel oxide thickness EOT (nm) [16]	7–8	7-8	6–7	6–7	6–7	6–7	6–7	6–7	6–7
Flash NOR interpoly dielectric thickness EOT (nm) [17]	13-15	13-15	13-15	13-15	13-15	10-12	10-12	10-12	10-12
Flash NAND interpoly dielectric thickness (nm) [18]	13–15	13–15	10–13	10–13	10–13	10–13	10–13	10–13	9–10
Flash endurance (erase/write cycles) [19]	1.00E+05	1.00E+05	1.00E+05	1.00E+05	1.00E+05	1.00E+06	1.00E+06	1.00E+06	1.00E+06
Flash nonvolatile data retention (years) [20]	10–20	10–20	10–20	10–20	10–20	10–20	10–20	10–20	20
Flash maximum number of bits per cell (MLC) [21]	2	2	2	2	2	4	4	4	4
FeRAM technology – F (nm) [22]	130	110	100	90	80	65	57	50	45
FeRAM cell size – area factor a in multiples of F^2 [23]	34	34	30	30	30	24	24	24	20
FeRAM cell size (μm^2) [24]	0.575	0.411	0.300	0.243	0.192	0.101	0.078	0.060	0.041
FeRAM cell structure [25]	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C
FeRAM capacitor structure [26]	stack	stack	stack	stack	stack	3D	3D	3D	3D
FeRAM capacitor footprint (μm^2) [27]	0.32	0.23	0.158	0.128	0.101	0.049	0.038	0.029	0.018
FeRAM capacitor active area (μm^2) [28]	0.32	0.23	0.158	0.128	0.101	0.076	0.069	0.064	0.059
FeRAM cap active area/footprint ratio [29]	1	1	1	1	1	1.55	1.85	2.2	3.31

Table 43a Non-Volatile Memory Technology Requirements—Near-term (continued)

Year of Production	2005	2006	2007	2008	2009	2010	2011	2012	2013
DRAM ½ Pitch (nm) (contacted)	80	70	65	57	50	45	40	36	32
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	90	78	68	59	52	45	40	36	32
MPU Physical Gate Length (nm)	32	28	25	22	20	18	16	14	13
Ferro capacitor voltage (V) [30]	1.5	1.5	1.2	1.2	1.2	1	1	1	0.7
FeRAM minimum switching charge density ($\mu C/cm^2$) [31]	11.4	14.2	19	22	26	30	30	30	30
FeRAM endurance (read/write cycles) [32]	1.0E+13	1.0E+14	1.0E+15	>1.0E16	>1.0E16	>1.0E16	>1.0E16	>1.0E16	>1.0E16
FeRAM nonvolatile data retention (years) [33]	10	10	10	10	10	10	10	10	10
SONOS/NROM technology – F (nm) [34]	100	90	70	65	55	50	45	40	35
SONOS/NROM cell size – area factor a in multiples of F^2 [35]	5.5	5.5	6	6	6	6	6	6	6.5
SONOS/NROM typical cell size (μm^2) [36]	0.055	0.045	0.029	0.025	0.018	0.015	0.012	0.01	0.008
SONOS/NROM maximum number of bits per cell ((physical 2-bit/cell) x MLC) [37]	2	2	2	2	2	4	4	4	4
SONOS/NROM area per bit (μm^2) [38]	0.028	0.022	0.015	0.013	0.009	0.0038	0.003	0.0024	0.002
SONOS L_g-stack (physical – µm) [39]	0.17	0.17	0.16	0.16	0.16	0.16	0.16	0.16	0.15
SONOS highest W/E voltage (V) [40]	5.0–6.0	5.0–6.0	5.0–5.5	5.0–5.5	5.0–5.5	5.0–5.5	5.0–5.5	5.0–5.5	5.0–5.5
SONOS/NROM I_{read} (µA) [41]	31–41	29–39	27–37	25–35	25–35	25–35	25–35	24–34	23–33
SONOS/NROM tunnel oxide thickness (nm) [42]	4.5	4	3.5	3.5	3.5	3.5	3.5	3.5	3
SONOS/NROM nitride dielectric thickness (nm) [43]	5	4.5	4	4	4	4	4	4	4
SONOS/NROM blocking (top) oxide or dielectric thickness (nm) [44]	4.5	4.5	4	6	6	6	6	6	6
SONOS/NROM endurance (erase/write cycles) [45]	1.00E+07	1.00E+07	1.00E+07	1.00E+07	1.00E+07	1.00E+08	1.00E+08	1.00E+08	1.00E+08
SONOS/NROM nonvolatile data retention (years) [46]	10–20	10–20	10–20	10–20	10–20	10–20	10–20	10–20	10–20
MRAM technology F (nm) [47]	180	90	90	65	65	45	45	45	32
MRAM cell size area factor a in multiples of F^2 [48]	25	23	20	22	19	20	18	18	19
MRAM typical cell size (μm^2) [49]	0.81	0.19	0.16	0.09	0.08	0.041	0.036	0.036	0.019
MRAM switching field (Oe) [50]	35	35	35	35	35	35	35	35	35
MRAM write energy (pJ/bit) [51]	150	100	70	35	35	25	25	25	20
MRAM active area per cell (μm^2) [52]	0.11	0.05	0.05	0.025	0.025	0.013	0.013	0.013	0.009
MRAM resistance-area product (Kohm-(μm^2) [53]	4	2	2	1.1	1	0.8	0.8	0.8	0.6
MRAM magnetoresistance ratio (%) [54]	40	70	70	70	70	70	70	70	70
MRAM nonvolatile data retention (years) [55]	>10	>10	>10	>10	>10	>10	>10	>10	>10
MRAM write endurance (read/write cycles) [56]	>3e16	>3e16	>3e16	>3e16	>3e16	>3e16	>3e16	>3e16	>3e16
MRAM endurance – tunnel junction reliability (years at bias) [57]	>10	>10	>10	>10	>10	>10	>10	>10	>10

Table 43a Non-Volatile Memory Technology Requirements—Near-term (continued)

Year of Production	2005	2006	2007	2008	2009	2010	2011	2012	2013
DRAM ½ Pitch (nm) (contacted)	80	70	65	57	50	45	40	36	32
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	90	78	68	59	52	45	40	36	32
MPU Physical Gate Length (nm)	32	28	25	22	20	18	16	14	13
PCRAM technology F (nm) [58]	90	70	65	57	50	45	40	35	32
PCRAM cell size area factor a in multiples of F^2 (BJT access device) [59]	7.2	7.0	6.4	5.6	5.8	5.8	5.8	6.1	5.8
PCRAM cell size area factor a in multiples of F^2 (nMOSFET access device) [60]	17.0	14.9	12.8	11.8	11.6	11.0	10.5	10.1	9.5
PCRAM typical cell size (μm^2) (BJT access device) [61]	0.059	0.034	0.027	0.018	0.015	0.012	0.0092	0.0074	0.0059
PCRAM typical cell size (μm^2) (nMOSFET access device) [62]	0.14	0.073	0.054	0.038	0.029	0.022	0.017	0.012	0.0097
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]	0.059	0.034	0.014	0.009	0.008	0.003	0.0023	0.0018	0.0015
PCRAM typical cell area per bit size (μm^2) (nMOSFET access device) [65]	0.14	0.073	0.027	0.019	0.015	0.006	0.004	0.003	0.0025
PCRAM storage element CD (nm) [66]	32	25	23	21	18	16	14	13	12
PCRAM phase change volume (nm^3) [67]	17,157	8,181	6,371	4,849	3,054	2,145	1,437	1,150	905
PCRAM reset current (μA) [68]	270	191	170	150	121	102	85	77	68
PCRAM set resistance (Kohm) [69]	2.5	3.4	3.7	4.3	5.0	5.7	6.5	7.6	8.5
PCRAM BJT current density (A/cm^2) [70]	4.3E+6	5.0E+6	5.1E+6	5.9E+6	6.2E+6	6.5E+6	6.8E+6	8.0E+6	8.5E+6
PCRAM BJT emitter area (μm^2) [71]	0.0064	0.0039	0.0033	0.0026	0.0020	0.0016	0.0013	0.00096	0.00080
PCRAM nMOSFET current density for reset (μA/μm) [72]	643	689	802	896	842	853	849	924	987
PCRAM nMOSFET device width (μm) [73]	0.42	0.28	0.21	0.17	0.14	0.12	0.10	0.083	0.069
PCRAM nonvolatile data retention (years) [74]	>10	>10	>10	>10	>10	>10	>10	>10	>10
PCRAM write endurance (read/write cycles) [75]	1.0E+12	1.0E+12	1.0E+12	1.0E+12	1.0E+12	1.0E+13	1.0E+13	1.0E+13	1.0E+14

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

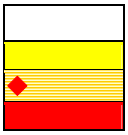


Table 43b Non-Volatile Memory Technology Requirements—Long-term

Year of Production	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
DRAM ½ Pitch (nm) (contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	*28*	*25*	*22*	*20*	*18*	*16*	*14*
MPU Physical Gate Length (nm)	*11*	*10*	*9*	*8*	*7*	*6*	*6*
Flash technology NOR/NAND – F (nm) [1]	28/25	25/23	22/20	20/18	18/16	16/14	14/13
Flash NOR cell size – area factor a in multiples of F^2 [2], [3], [4], [5]	10–12	10–13	10-–13	11–14	11–14	12–14	12–14
Flash NAND cell size – area factor a in multiples of F^2 SLC/MLC [6]	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
Flash NOR typical cell size (μm^2) [7], [8]	0.0086	0.0073	0.0057	0.005	0.004	0.0034	0.0026
Flash NOR L_g-stack (physical – µm) [8], [9]	0.09	0.09	0.08	0.08	0.07	0.07	0.06
Flash NOR highest W/E voltage (V) [10], [11]	6–8	6–8	6–8	6–8	6–8	6–8	6–8
Flash NAND highest W/E voltage (V) [12]	15–17	15–17	15–17	15-17	15–17	15–17	15–17
Flash NOR I_{read} (µA) [13]	24–30	23–29	22–28	21–27	20–26	19–25	18–24
Flash coupling ratio [14]	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
Flash NOR tunnel oxide thickness EOT (nm) [15]	7 – 8	7 – 8	7 – 8	7 – 8	7 – 8	7 – 8	7 – 8
Flash NAND tunnel oxide thickness EOT (nm) [16]	6–7	6–7	6–7	6–7	6–7	6–7	6–7
Flash NOR interpoly dielectric thickness EOT (nm) [17]	8–10	8–10	8–10	8–10	7–9	6–8	6–8
Flash NAND interpoly dielectric thickness (nm) [18]	9–10	9–10	9–10	9–10	9–10	9–10	9–10
Flash endurance (erase/write cycles) [19]	1.00E+06	1.00E+06	1.00E+07	1.00E+07	1.00E+07	1.00E+07	1.00E+07
Flash nonvolatile data retention (years) [20]	20	20	20	20	20	20	20
Flash maximum number of bits per cell (MLC) [21]	4	4	4	4	4	4	4
FeRAM technology – F (nm) [22]	40	35	32	28	25	22	20
FeRAM cell size – area factor a in multiples of F^2 [23]	20	20	16	16	16	14	14
FeRAM cell size (μm^2) [24]	0.032	0.025	0.016	0.013	0.010	0.007	0.006
FeRAM cell structure [25]	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C
FeRAM capacitor structure [26]	3D	3D	3D	3D	3D	3D	3D
FeRAM capacitor footprint (μm^2) [27]	0.014	0.011	0.0064	0.0049	0.0039	0.0024	0.002
FeRAM capacitor active area (μm^2) [28]	0.055	0.05	0.047	0.043	0.04	0.037	0.035
FeRAM cap active area/footprint ratio [29]	3.88	4.63	7.38	8.81	10.25	15.12	17.17
Ferro capacitor voltage (V) [30]	0.7	0.7	0.7	0.7	0.7	0.7	0.7
FeRAM minimum switching charge density ($\mu C/cm^2$) [31]	30	30	30	30	30	30	30
FeRAM endurance (read/write cycles) [32]	>1.0E16	>1.0E16	>1.0E16	>1.0E16	>1.0E16	>1.0E16	>1.0E16
FeRAM nonvolatile data retention (years) [33]	10	10	10	10	10	11	12
SONOS/NROM technology – F (nm) [34]	32	28	25	23	20	19	18
SONOS/NROM cell size – area factor a in multiples of F^2 [35]	6.5	6.5	7	7	7	7	7
SONOS/NROM typical cell size (μm^2) [36]	0.007	0.005	0.004	0.0037	0.003	0.0025	0.002
SONOS/NROM maximum number of bits per cell ((physical 2-bit/cell) x MLC) [37]	4	4	4	4	4	4	4
SONOS/NROM area per bit (μm^2) [38]	0.0018	0.0013	0.0011	0.0009	0.0007	0.0006	0.0005
SONOS L_g-stack (physical – µm) [39]	0.15	0.15	0.14	0.14	0.14	0.13	0.13
SONOS highest W/E voltage (V) [40]	5.0–5.5	5.0–5.5	4.5–5.0	4.5–5.0	4.0–4.5	4.0–4.5	4.0–4.5
SONOS/NROM I_{read} (µA) [41]	23–33	22–32	21–31	21–31	20–30	20–30	20–30

Table 43b Non-Volatile Memory Technology Requirements—Long-term (continued)

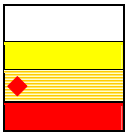
Year of Production	2014	2015	2016	2017	2018	2019	2020
DRAM ½ Pitch (nm) (contacted)	28	25	22	20	18	16	14
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	28	25	22	20	18	16	14
MPU Physical Gate Length (nm)	11	10	9	8	7	6	6
SONOS/NROM tunnel oxide thickness (nm) [42]	3	3	2.5	2.5	2.5	2	2
SONOS/NROM nitride dielectric thickness (nm) [43]	4	4	4	4	4	4	4
SONOS/NROM blocking (top) oxide or dielectric thickness (nm) [44]	6	6	5	5	5	5	5
SONOS/NROM endurance (erase/write cycles) [45]	1.00E+08	1.00E+08	1.00E+09	1.00E+09	1.00E+09	1.00E+09	1.00E+09
SONOS/NROM nonvolatile data retention (years) [46]	10–20	10–20	10–20	10–20	10–20	10–20	10–20
MRAM technology F (nm) [47]	32	32	22	22	22	16	16
MRAM cell size area factor a in multiples of F^2 [48]	17	17	18	16	16	17	16
MRAM typical cell size (μm^2) [49]	0.017	0.017	0.009	0.0077	0.0077	0.0044	0.0041
MRAM switching field (Oe) [50]	35	35	35	35	35	35	35
MRAM write energy (pJ/bit) [51]	20	20	20	20	20	20	20
MRAM active area per cell (μm^2) [52]	0.009	0.009	0.007	0.007	0.007	0.005	0.005
MRAM resistance-area product (Kohm-(μm^2) [53]	0.6	0.6	0.6	0.6	0.6	0.6	0.6
MRAM magnetoresistance ratio (%) [54]	70	70	70	70	70	70	70
MRAM nonvolatile data retention (years) [55]	>10	>10	>10	>10	>10	>10	>10
MRAM write endurance (read/write cycles) [56]	>3e16	>3e16	>3e16	>3e16	>3e16	>3e16	>3e16
MRAM endurance – tunnel junction reliability (years at bias) [57]	>10	>10	>10	>10	>10	>10	>10
PCRAM technology F (nm) [58]	28	25	22	20	18	16	14
PCRAM cell size area factor a in multiples of F^2 (BJT access device) [59]	5.9	5.9	5.0	4.8	4.7	4.7	4.7
PCRAM cell size area factor a in multiples of F^2 (nMOSFET access device) [60]	8.7	8.2	7.4	6.8	6.3	5.8	5.4
PCRAM typical cell size (μm^2) (BJT access device) [61]	0.0046	0.0037	0.0024	0.0019	0.0015	0.0012	0.0009
PCRAM typical cell size (μm^2) (nMOSFET access device) [62]	0.0068	0.0051	0.0036	0.0027	0.0020	0.0015	0.0011
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]	0.0012	0.0009	0.0006	0.0005	0.0004	0.0003	0.0002
PCRAM typical cell area per bit size (μm^2) (nMOSFET access device) [65]	0.0017	0.0013	0.0009	0.0007	0.0005	0.0004	0.0003
PCRAM storage element CD (nm) [66]	10	9	7.9	7.2	6.5	5.8	5.0
PCRAM phase change volume (nm^3) [67]	524	382	268	180	113	102	65
PCRAM reset current (μA) [68]	53	46	39	32	26	21	16
PCRAM set resistance (Kohm) [69]	9.9	11.3	13.2	14.7	16.7	18.7	21.7
PCRAM BJT current density (A/cm^2) [70]	8.6E+6	9.3E+6	1.0E+7	1.0E+7	1.0E+7	1.0E+7	1.0E+7
PCRAM BJT emitter area (μm^2) [71]	0.00062	0.00049	0.00038	0.00031	0.00026	0.00020	0.00015
PCRAM nMOSFET current density for reset (μA/μm) [72]	997	1,056	1,202	1,270	1,310	1,320	1,340
PCRAM nMOSFET device width (μm) [73]	0.053	0.043	0.032	0.025	0.020	0.016	0.012
PCRAM nonvolatile data retention (years) [74]	>10	>10	>10	>10	>10	>10	>10
PCRAM write endurance (read/write cycles) [75]	1.0E+14	1.0E+14	1.0E+15	1.0E+15	1.0E+15	1.0E+15	1.0E+15

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known



Notes for Table 43a and b:

[1] In the past Flash devices tended to lag behind the current CMOS technology's feature size, F, but that delay no longer exists. This entry provides the F value for designs in the indicated time period.

[2] The area factor "a" = cell area/F^2, so this entry presents the expected range for Flash NOR cell area in multiples of the implementation technology's F^2. Note the slowly increasing trend that reflects the difficulty of scaling the gate length when the tunnel oxide thickness is fixed.

[3] High-κ interpoly dielectric is projected at the 45 nm technology generation and beyond, and gate coupling ratio of >0.7 can be achieved which helps to maintain the cell size.[9] This helps to slow down the increase in the area factor.

[4] Although virtual ground array may significantly decrease the cell size in the near term [10]this effect has not been included in the current table.

[5] 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 deployment of high κ in interpoly may help to reduce the Lg somewhat.

[6] The area factor "a" = cell area/F^2, so this entry presents the Flash NAND cell area in multiples of F^2 of the implementation technology. Flash NAND enjoys a small cell size because much of the cell structure is shared among a group of cells. (SLC = single level cell, MLC = multilevel cell; see Note 21 below for more detail.)

[7] A typical Flash NOR cell size in micrometers squared is estimated using the midrange area factor "a."

[8] Both the cell size and the gate length for NOR Flash have been more aggressively scaled recently.[11, 12, 13]

[9] This is the physical length of the control gate of Flash NOR devices.

[10, 12] This is the highest voltage relative to ground seen in the cell array. It is not usually an external supply.

[11] The introduction of high-κ interpoly dielectric will help to reduce the erase voltage.

*[13] The current reduces with scaling at a rate higher than W/(L*Cox) to reduce the voltage overdrive factor.*

[14] The coupling ratio is the (control gate to floating gate capacitance)/(total floating gate to source, drain and substrate capacitance).

[15, 16] Tunnel oxides must be thick enough to assure retention but thin enough to allow ease of erase/write. This difficult trade off problem hinders scaling. Tunnel oxides less than 7 nm seem to pose fundamental problems for retention reliability.

[17, 18] Interpoly dielectric must be thick enough to assure retention but thin enough to assure an almost constant coupling ratio. Charge retention when the dielectric is scaled downward is the major issue. High-κ interpoly will help reducing the interpoly EOT and maintain constant coupling ratio without losing retention.

[19] 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 product. It is expected that emerging technology will allow both tradeoffs of endurance for retention as well as increases in the specified minimum endurance capability as device design options.

[20] 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.

[21] Cell read out distinguishes between four levels of charge storage to provide two storage bits. Progression to 16 levels is anticipated but maintaining reasonable V_t, read speed and array efficiency beyond 2-bit/cell are challenging. (MLC multilevel cell).

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

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

[24] FeRAM cell size is presented in terms of micrometers squared. It is the product "a"xF^2.

[25] FeRAM cell structures have migrated to one transistor, one capacitor (1T1C) formats.[14,15] Other alternative configurations are under investigation such as Chain-FeRAM.[16, 17]

[26] 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.

[27] 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 or less (3D capacitor) are assumed.

[28] 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.

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

[30] 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.[18]

[31] 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}.

[32] 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. Test time is a serious concern. Note that operation at 100 MHz for 10 years would accumulate 1E16 cycles.

[33] This is the data retention requirement while the device is disconnected from power. It is usually specified at 85 ºC.

[34] SONOS/NROM devices have recently been introduced into the commercial market and will tend to lag the feature size of the current CMOS technology by one year. This entry provides the F value for designs in the indicated time period.

[35] The area factor "a" = cell area/F^2. This entry depicts the expected SONOS/NROM NOR cell area in multiples of the implementation technology's F^2. SONOS/NROM device stores two physical bits of data per device. This area factor "a" is per cell, not per bit.

[36] The expected "typical" SONOS/NROM NOR cell size is presented in terms of micrometers squared. Again, this cell size is per cell, not per physical or MLC bit.

[37] MBC signifies "multiple bit storage," while MLC signifies "multiple level storage." The SONOS/NROM cell stores charge in two distinct locations – in the nitride over the source and drain junctions. Thus, in the simplest case there are two distinct bits within each cell; however, each charge location may be partitioned into multiple levels (MLC), thereby, increasing the bit storage per cell.

[38] The expected SONOS/NROM NOR area per bit is presented in terms of micrometers squared. The stored bit includes both physical 2-bit/device and MLC.

[39] This is the physical length of the gate of SONOS/NROM devices in micrometers as there is only a single gate, similar to a MOSFET.

[40] This is the highest voltage relative to ground seen in the cell array. It is not usually an external supply.

*[41] Reduction rate is higher than (W/L)*Cox) to reduce the voltage overdrive factor.*

[42] Tunnel oxides must be thick enough to assure retention but thin enough to allow ease of erase/write at low voltage. This offers a challenge to scaling.

[43] The nitride dielectric provides the charge storage medium and its thickness is a compromise between program/erase voltages, erase/write window, retention, process control and endurance. This offers a challenge to scaling.

[44] The blocking (top) oxide thickness isolates the charge storage region (nitride) from the gate electrode. Its thickness is a compromise between program/erase voltages and retention. This offers a challenge to scaling. With the advent of high-κ dielectrics, such as aluminum oxide, and advanced deposition techniques, such as atomic layer deposition (ALD), the thickness of the blocking (top) insulator may be increased to prevent gate injection, while maintaining program speed and long-term retention. This technology will probably mature in the year 2008 and beyond. High-κ dielectrics also ease the requirements on scaling the nitride and tunnel oxide since the electric fields may be preserved in the latter.

[45] E/W endurance (erase/write cycles) requirements vary with the specifics of an application, but 1E5 cycles has been accepted as the historical minimum acceptable level for a useful product. It is expected that emerging technology will allow both tradeoffs of endurance for retention as well as increases in the specified minimum endurance capability as device design options.

[46] SONOS/NROM retention follows a stretched exponential curve and saturates and becomes time independent afterwards. Thus the charge loss affects the V_t programming window but is not a long-term reliability issue. As long as the programming window is sufficiently designed retention by itself is not a concern. However, charge loss mechanisms and retention models are still being improved and further modifications are possible in the future[19,20].

[47] MRAM devices are expected to lag the feature size of the CMOS current technology until 2010. This entry provides the F value for designs in the indicated time period.

[48] The area factor "a" = cell area/F^2. This entry is the expected MRAM cell area in multiples of the implementation technology's F^2.

[49] The expected "typical" MRAM cell size is presented in micrometers squared.

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

*[51] 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).*

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

*[53] 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.*

[54] 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.

[55] MRAM devices are required to retain data while unpowered. This entry states the retention requirement in years.

[56] This entry is the required number of read/write cycles that an MRAM device must be able to endure without degradation that impacts the ability of the device to pass all operating specifications.

[57] An MRAM device is required to meet this minimum life requirement when the magnetic material stack is continuously under bias.

[58] PCRAM devices are expected to follow the feature size of the current CMOS technology. This entry provides the F value for designs in the indicated time period.

[59] The area factor "a" = cell area/F^2. This entry is the expected PCRAM cell area in multiples of the implementation technology's 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's 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's feature size, 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.[21, 22]

[75] This entry is the expected PCRAM W/E cycling endurance. Recent published data indicate cycling endurance from 1E+9 to 1E+13.[23, 24]

NON-VOLATILE MEMORY POTENTIAL SOLUTIONS

Flash devices achieve non-volatility by storing and sensing the charge on a floating gate. Charge storage and charge removal require current flow through the dielectric materials that surround the floating gate. This implies high electric field stress. The conventional memory transistor vertical stack consists of a refractory salicide control gate, an interpoly dielectric, a polysilicon floating gate, a tunnel dielectric, and the silicon substrate. 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 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. Choices of current injection methods, voltage levels and waveforms during erase and program, dielectric materials, and cell geometry constitute a part of the trade-off space for the technology. The complexity of balancing these numerous concerns while struggling to achieve smaller cell sizes and retaining compatibility with a scaled CMOS technology is the major challenge for the technology. Simple adjustments of the cell dimensions are inadequate to continue scaling, and appropriate modifications of circuit architecture, waveforms, algorithms, etc. need to be included as part of the development effort. The potential solutions chart, Figure 42, points out that the process of exploring combinations of possible tradeoffs is a continuous task that does not involve fundamental research, and the indicated timescale represents only the near term activity. This type of procedure will be repeated for successive technology generations.

The tunnel oxide thickness for the floating gate device poses the largest scaling challenge, and there is currently no recognized solution. This inhibits the scaling of physical gate length. The use of high-κ dielectric in the blocking dielectric will be helpful to reduce the total EOT while maintaining or even increasing the gate-coupling ratio. This helps to reduce the Flash cell in both the word line and the bit line directions, and allows the cell size to continue to scale, although at a slower pace than F^2. 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. Neither of these potential solutions provides a perfect path for scaling, but both provide more breathing space to extend the useful years of the conventional floating gate devices. Non-planar and multi-gate devices may further relieve the punchthrough limitation and allow further scaling. Another inevitable problem is the floating gate interference issue when the spacing between neighboring floating gates approaches the floating gate height. The use of nitride trapping layer instead of polysilicon (see SONOS/NROM) will defuse this problem, but is not without its own issues.

FeRAM devices achieve nonvolatile memory by switching and sensing the polarization state of a ferroelectric capacitor. It is a technical challenge to find ferroelectric 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. The ferroelectric material must be physically and chemically isolated from the underlying CMOS. The ferroelectric materials, buffer materials, and process conditions are still being refined. In addition, in order to achieve density goals the basic geometry of the cell must be modified while maintaining the desired isolation. Recent progress in electrode materials shows promise for thinning down the ferroelectric capacitor and extends the viability of 2D stacked capacitor through much of the near term years. Beyond this, the need for 3D capacitor still poses steep challenges. More details for FeRAM operation, requirements and materials development are described in the *Front End Process* chapter.

SONOS/NROM devices store and remove charge in traps in a nitride layer to achieve nonvolatile memory. As the acronym indicates, the memory transistor cross section (top down) consists of a polysilicon gate, a blocking oxide, a silicon nitride layer, a tunnel oxide and the silicon channel. SONOS is a well established mature technology that is entering a new phase because of the advent of the NROM concept where charge is stored in localized regions of the nitride layer adjacent to the source and drain junctions. Having two local regions allows a single memory transistor to store two bits of information. As this class of device is scaled, the short term challenge will be to control and optimize the properties of the ONO stack, including layer thicknesses and trap energy and space distributions. In the long term, scaling of the gate length will reduce the separation between the localized charge storage regions and the interference between the two bits will reduce the programming window (second bit effect). Nevertheless, devices with physical gate length ~ 60 nm have been demonstrated. The potential solutions table points out that the stack refinement effort now underway can serve the needs of the 65 nm technology generation in 2007, but further innovations are needed beyond the 45 nm technology generation in 2010.

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 was stored. Control of the MTJ dimensions and material properties is the major challenge. 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. The potential solutions table provides a snapshot of the current materials research activity that must converge to a manufacturable solution for 65 nm.

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: (A) RESET, for which the chalcogenide glass is momentarily melted by a short electric pulse and re-solidifies into amorphous solid with high resistivity, and (B) SET, for which a lower amplitude but longer pulse (10 ns–100 ns) anneals the amorphous phase into low resistance crystalline state. The major challenges for PCRAM are the high current (0.5 mA) required to reset the phase change element, and the relatively long set time. Since the volume of phase change material decreases rapidly with time, with scaling both challenges become easier to deal with.

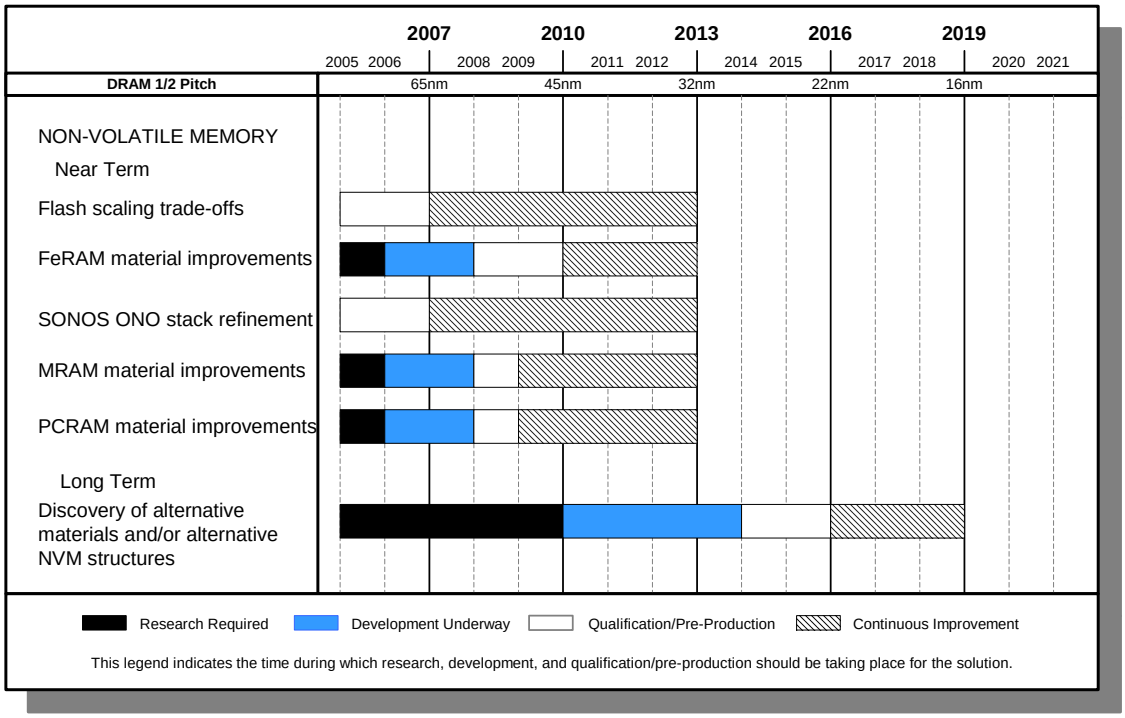


Figure 42 Non-volatile Memory Potential Solutions

RELIABILITY

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, 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 interlevel dielectrics.

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.

Furthermore, 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.

Details of the broad range of reliability challenges can be found in the accompanying document, "*Critical Reliability Challenges for the International Technology Roadmap for Semiconductors (ITRS)*" published by the Reliability Technical Advisory Board (RTAB) at SEMATECH. The report identifies reliability concerns and research needs in 13 areas—High κ, Metal Gate, Cu/low-κ, SOI, Novel Devices, Microsystems, Flash Memories, Soft Errors, ESD, Latchup, Packaging, Design for Reliability and Defect Screening.

TOP RELIABILITY CHALLENGES

Table 44 indicates the RTAB consensus on the top five near-term 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.

It now seems likely that the move from silicon dioxide to a new high-κ gate insulator will be accompanied by a simultaneous move from polysilicon to metal gate electrodes. This has led to listing high κ and metal gate as a single challenge in this edition of the roadmap (they were separate challenges in previous editions). 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, one needs only to realize that 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.

There are additional reliability challenges associated with advanced packaging for higher performance, higher power integrated circuits. Increasing power, increasing pin count, 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.

Design for reliability tools are needed so that reliability can be assured proactively during the technology development and design. Furthermore, reliability testing and defect screening are becoming more challenging in advanced, higher power generating technologies. Again, the reader is referred to the above-mentioned SEMATECH document for detailed explanations of all these issues.

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.

Table 44 Reliability Difficult Challenges

Difficult Challenges ≥ 32 nm	*Summary of Issues*
High-κ gate dielectrics with metal gate electrodes	Dielectric breakdown characteristics (hard and soft breakdown) Transistor stability (charge trapping, work function stability, metal ion drift or diffusion) Impact of implantation Metal gate thermomechanical issues (coefficient of thermal expansion mismatch)
Copper/Low-κ interconnects	Stress migration of Cu vias and lines Cu via and line electromigration performance Impact of degradation of properties with lowering k (strength, adhesion, thermal conductivity, coefficient of thermal expansion) Time Dependent Dielectric Breakdown of the Cu/low-κ system Impact of packaging
Packaging	Impact of increasing Coefficient of Thermal Expansion (CTE) mismatch between low-κ, silicon and organic packages Increasing use of multi-chip packages and heterogeneous integration (e.g., CMOS plus MEMs or Sensor) Electromigration in package traces, vias and bumps Impact of assembly and packaging on on-chip failure mechanisms (cracking, stack delamination) Ability of bumps to withstand thermal and mechanical stresses while providing sufficient current carrying capability
Design and test for reliability	Simulation tools for concurrent optimization of circuit performance and reliability Tools to simulate electromigration, thermal-mechanical stress and process induced charging Soft error detection and correction at chip and system level, including random logic faults Screens for resistive and capacitively coupled interconnect defects Alternative screens for decreasing burn-in effectiveness
Negative bias temperature instability	Degradation of p channel current Dependence on scaling and nitrogen in gate insulator Impact on burn-in
Difficult Challenges< 32 nm	*Summary of Issues*
Reliability of novel devices, structures, materials and applications	Need to identify and model failure modes, develop acceleration techniques and qualify Post-Cu interconnect solutions (e.g., optical, robust thermal solution, superconductors) Non-CMOS transistors and memory elements New packaging paradigms Novel applications

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 tradeoffs 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 to red indicates a growing reliability risk. The requirements first turn to yellow (Manufacturing Solutions are Known) in 2007 indicating a relative smaller risk associated with scaling, increased power and the introduction into manufacturing of strained silicon substrates.

The requirements first turn to striped (Interim Solutions Known) in 2008 due to the increased reliability risk resulting from introduction of high-κ gate insulators, metal gates and fully depleted SOI (driven by the needs of high performance logic). The reliability knowledge base and tools will not be as well developed as for oxide insulators and polysilicon gates. As an interim solution, the techniques that are currently used to qualify oxy-nitride/poly gate transistors can be used to qualify transistors with high-κ insulators and metal gates. However, there is a real risk associated with new failure modes and with changes in the known failure modes in the move to high κ. For all these reasons, it is not known to what requirements (failure rate, time, operating temperature) these high-κ insulators will be capable of being qualified to. However, reliability qualification, *per se*, is not considered a show-stopper (such as a red box).

Red is now reserved for the reliability requirements nearer to the end of the Roadmap. Red is only used where a reliability solution is not known because the technology choices have not been specified (e.g., what specifically will be used for post Cu interconnect or what specific novel devices will be required?). When the details of these technology changes are defined better a better assessment of the reliability risk will be possible.

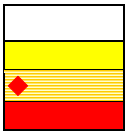
Table 45a Reliability Technology Requirements—Near-term

Year of Production	2005	2006	2007	2008	2009	2010	2011	2012	2013	
DRAM ½ Pitch (nm) (contacted)	80	70	65	57	50	45	40	36	32	
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	90	78	68	59	52	45	40	36	32	
MPU Physical Gate Length (nm)	32	28	25	22	20	18	16	14	13	
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]	10-100	10-100	50–2000	10–100	10–100	10–100	10–100	10–100	10–100	
Soft error rate (FITs)	1000	1000	1000	1000	1000	1000	1000	1000	1000	
Relative failure rate per transistor (normalized to2005 value) [3]	1.00	0.79	0.63	0.50	0.40	0.32	0.25	0.20	0.16	Number of transistors
Relative failure rate per m of interconnect (normalized to2005 value) [4]	1.00	0.84	0.71	0.59	0.51	0.47	0.41	0.37	0.33	Customer needs; J11 length of interconnect

Table 45b Reliability Technology Requirements—Long-term

Year of Production	2014	2015	2016	2017	2018	2019	2020	
DRAM ½ Pitch (nm) (contacted)	28	25	22	20	18	16	14	
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	28	25	22	20	18	16	14	
MPU Physical Gate Length (nm)	11	10	9	8	7	6	6	
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]	10–100	10–100	10–100	10–100	10–100	10–100	10–100	
Soft error rate (FITs)	1000	1000	1000	1000	1000	1000	1000	
Relative failure rate per transistor (normalized to2005 value) [3]	0.13	0.10	0.08	0.06	0.05	0.04	0.03	Number of transistors
Relative failure rate per m of interconnect (normalized to2005 value) [4]	0.29	0.27	0.22	0.20	0.18	0.16	0.14	Customer needs; J11 length of interconnect

Please note that in the above Long-term table, the "Relative failure rate per transistor" value for 2019–2020 is not entered, because there is not projection in the ORTC tables for the number of transistors per chip in those years.



Notes for Table 45a 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 technologygeneration. 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 43 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 continually improve the reliability models and the reliability engineering capabilities. 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 43.

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.

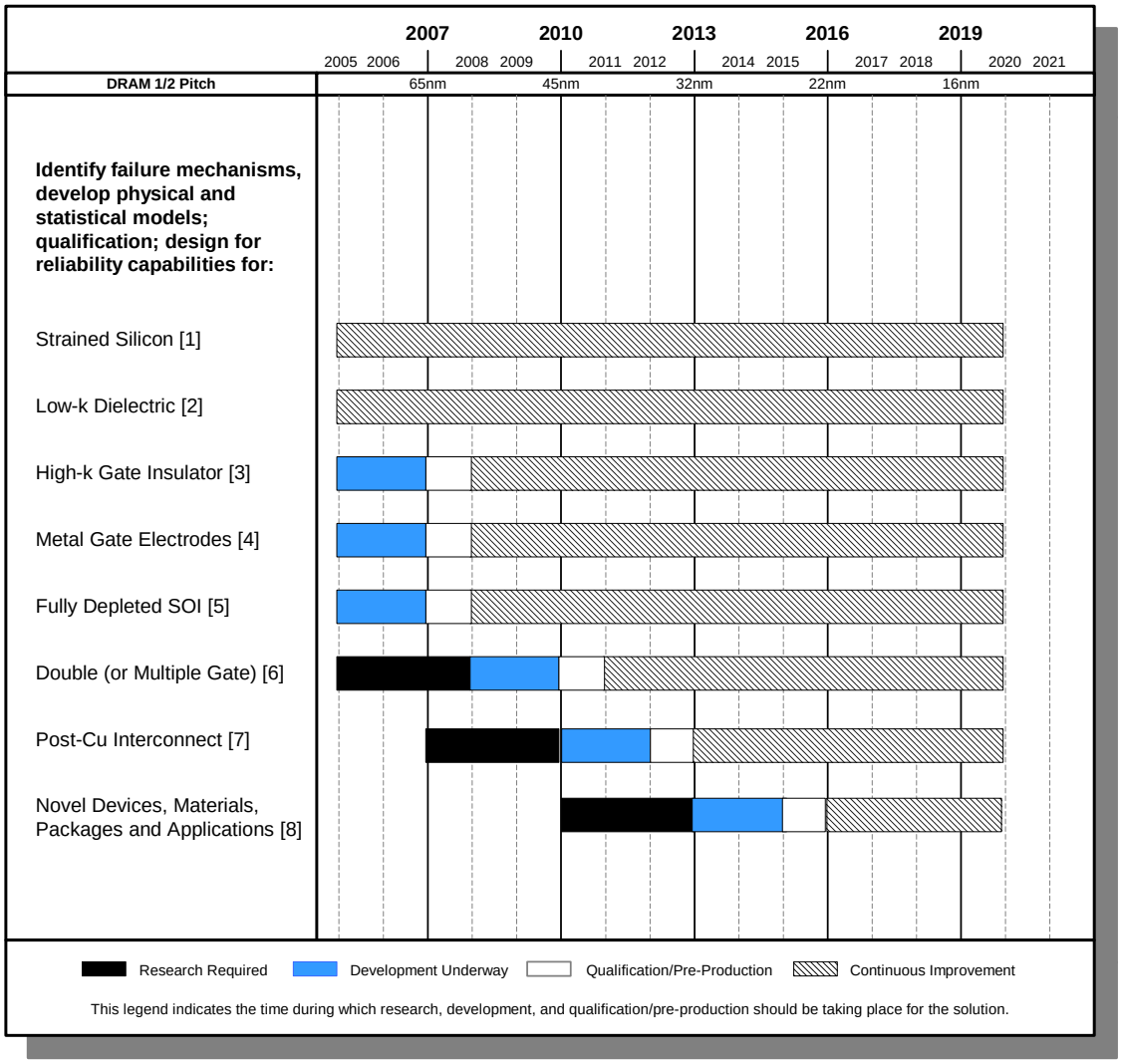


Figure 43 Reliability Potential Solutions

Notes for Figure 43 Reliability Potential Solutions:
[1] Strained Silicon has entered volume production. More aggressive use of strained silicon is expected in future technologies. Need to continually assess its impact on transistor failure mechanisms (e.g., hot carrier and NBTI)
[2] Low-κ interlevel dielectrics have entered production. Successive technology generations will introduce lower and lower k 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-k 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 aggressively scale the equivalent oxide thickness, 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

LOGIC TECHNOLOGY REQUIREMENTS

[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] 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.

[5]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.

[6] 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.

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

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

NON VOLATILE MEMORY TECHNOLOGY REQUIREMENTS

[9] E. S. Cho, et al., "Hf-silicate Inter-Poly Dielectric Technology for sub 70 nm Body Tied FinFET Flash Memory", *Symposium on VLSI Technology*, pp. 208-209, June 2005.

[10] R. Koval, et al. "Flash ETOX Virtual Ground Architecture: A Future Scaling Direction", *Symposium on VLSI Technology*, pp. 204–205, June 2005.

[11] "A 90nm 512Mb 166 MHz Multilevel Cell Flash Memory with 1.5MBytes/s Programming", *International Solid State Circuits Conference,* pp. 54–55, February 2005.

[12] "Highly Manufacturable 90 nm NOR Flash Technology with 0.081 μm^2 Cell Size", *Symposium on VLSI Technology*, pp. 91–92, June 2003.

[13] "A 70 nm NOR Flash Technology with 0.049 μm^2 Cell Size", *Symposium on VLSI Technology*, pp. 238–239, June 2004.

[14] J.H. Park, et al., "Fully Logic Compatible (1.6V Vcc, 2 Additional FRAM Masks) Highly Reliable Sub $10F^2$ Embedded FRAM with Advanced Direct Via Technology and Robust 100 nm thick MOCVD PZT Technology", *Technical Digest of IEEE International Electron Devices Meeting,* pp. 591–594, December 2004.

[15] Y. M. Kang et al., "Sub-1.2V Operational, 0.15 $\mu m/12F^2$ Cell FRAM Technologies for Next Generation SoC Applications", *Symposium on VLSI Technology*, pp. 102–103, June 2005.

[16] H. Kanaya et al., "A 0.602 μm^2 Nestled Chain Cell Structure Formed by One Mask Etching Process for 64 Mbit FeRAM", *Symposium on VLSI Technology*, pp. 150–151, June 2004.

[17] N. Nagel et al., ""New Highly Scalable 3 Dimensional Chain FeRAM Cell with Vertical Capacitor", *Symposium on VLSI Technology*, pp. 146–147, June 2004.

[18] D. C. Yoo et al., "Highly Reliable 50 nm-thick PZT Capacitor and Low Voltage FRAM Device Using $Ir/SrRuO_2$/MOCVD PZT Capacitor Technology", *Symposium on VLSI Technology*, pp. 100–101, June 2005.

[19] M. Janai and B. Eitan, "The Kinetics of Degradation of Data Retention of Post-cycled NROM Non-Volatile Memory Products", *IEEE International Reliability Physics Symposium Proceedings*, pp. 175–180, 2005.

[20] T. Wang et al., "Reliability Models of Data Retention and Read-disturb in 2-bit Nitride Storage Flash Memory Cells", *Technical Digest of IEEE International Electron Devices Meeting*, pp. 169–172, December 2003.

[21] S. J. Ahn et al., "Highly Manufacturable High Density Phase Change Memory of 64 Mb and Beyond", *Technical Digest of IEEE International Electron Devices Meeting*, pp. 907–910, December 2004

[22] A. L. Lacaita et al., "Electrothermal and Phase Change Dynamics in Chalcogenide-Based Materials", *Technical Digest of IEEE International Electron Devices Meeting*, pp. 911–914, December 2004.

[23] S.J. Ahn et al., "Highly Manufacturable High Density Phase Change Memory of 64 Mb and Beyond", *Technical Digest of IEEE International Electron Devices Meeting*, pp. 907–910, December 2004.

[24] S. Lai et al., "Current Status of Phase Change Memory and Its Future", *Technical Digest of IEEE International Electron Devices Meeting*, pp. 255–258, December 2003.