INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

2003 EDITION

PROCESS INTEGRATION, DEVICES, AND STRUCTURES

INCLUDES
RF AND ANALOG/MIXED-SIGNAL TECHNOLOGIES FOR WIRELESS COMMUNICATIONS AND EMERGING RESEARCH DEVICES

TABLE OF CONTENTS

PROCESS INTEGRATION, DEVICES, AND STRUCTURES
INCLUDING RF AND ANALOG/MIXED-SIGNAL FOR WIRELESS COMMUNICATIONS AND EMERGING RESEARCH DEVICES

LIST OF FIGURES

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, passive electrical parameters, and reliability criteria. Nominal targets as well as statistical tolerances are discussed. The chapter deals with the following major topics: logic, memory, reliability, radio frequency (RF) and analog/mixed-signal (AMS) technologies for wireless communication, and emerging research devices. The section on RF and AMS technologies for wireless communication has been considerably expanded and focused on wireless communications for the 2003 ITRS. This focus was selected because wireless is becoming a major application area, and hence a major technology driver. The emerging research devices section, which was added to the PIDS chapter in the 2000 ITRS Update, has been considerably revamped, with expanded quantitative analysis and critical evaluation of the various novel approaches.

There are several key themes in the PIDS chapter of the 2003 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. These innovations are expected to be introduced at an accelerating pace, and hence understanding, modeling, and ensuring the reliability for all of them in a timely manner is expected to be a major issue for the industry. Another theme is that for non-volatile memory, several additional types, silicon-oxide-nitride-oxide-silicon (SONOS) and magnetic RAM (MRAM) are approaching implementation into mainstream production, and hence are evaluated in the memory section. A third theme is the wide variety of frequency range (1 GHz through 100 GHz) and supporting technologies (silicon MOSFET and bipolar, silicon-germanium HBT, and III-V FET and HBT) required for wireless communications systems. A fourth theme is the introduction and critical assessment of major new beyond-CMOS approaches to information and signal processing with potential to extend microelectronics scaling beyond the domain of CMOS. This is projected to occur toward the end of the Roadmap and beyond, and is primarily discussed in the Emerging Research Devices section. A final key theme is the need to include more and more types of embedded functions into the design of an IC, with simultaneous satisfaction of constraints of interconnection, power consumption, reliability, device performance, and cost. This is the fundamental issue associated with the System-on-a-Chip (SoC) application area, which is a major trend for future ICs. SoC issues are a larger crosscut problem that will be dealt with as a multidisciplinary activity in a separate chapter, but they will also be discussed in this chapter, particularly in the section on radio frequency and analog/mixed-signal technologies for wireless communications. An important consequence of the drive toward SoC is the need for multiple transistors and transistor types on a given chip to meet the needs of the different functions, as well as the need for embedded memory on logic and mixed-signal chips. In this chapter, however, the main focus will be on those transistors that drive the technology forward. In addition, alternative monolithic System-in-a-Package (SIP) integration may be preferred. In many applications, the SIP approach is especially suited to bring the specialized RF and AMS technologies together in a highly integrated, high-performance and low- cost unit, which will be addressed in the RF and AMS section in this chapter.

The aim of the Roadmap is to identify key technical requirements and challenges critical to sustaining the historical scaling of CMOS technology (i.e., per Moore's Law), and to stimulate the needed research and development to meet the key challenges. In particular, the objective of listing and discussing the potential solutions in this chapter is to provide example solutions addressing the key technical challenges. However, the solutions considered here are not comprehensive, and are not meant to limit research exploring new and different approaches.

LOGIC

A major portion of semiconductor device production is devoted to digital logic ICs as well as to memory ICs. In this section, both high-performance and low-power logic are included, and detailed technology requirements and potential solutions are considered for both types. Key considerations are performance, power, and density requirements and goals. The focus here is on the highest performance transistors, which pre-eminently drive the technology forward.

MEMORY

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

As mentioned above, NVM requirements and challenges are treated for several technologies, including Flash (NOR and NAND), Ferroelectric RAM (FeRAM), Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) and Magnetic RAM (MRAM). 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 nodes 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.

RF AND ANALOG/MIXED-SIGNAL TECHNOLOGIES FOR WIRELESS COMMUNICATIONS

The engine behind the rapid semiconductor market growth in wireless communication applications is the advance of radio frequency (RF) and analog/mixed-signal (AMS) ICs. RF and AMS ICs contain analog and/or radio frequency circuitry as well as digital circuitry to realize functionality of analog/RF signal amplification, analog-to-digital and digital-to-analog conversion, frequency synthesis and translation, etc. The technology requirements for meeting the demands of wireless communication systems are extremely varied, often conflicting with each other, and very different from the digital technology requirements. In comparison to the Mixed-Signal section in the 2001 ITRS, the scope and content of the RF and AMS section in the 2003 ITRS have been substantially expanded. For the first time, compound semiconductor (e.g., III-V) devices and technologies are included in ITRS. Depending on their application frequency, the RF and AMS technology requirements and potential solutions are addressed in four different areas: 1) analog and mixed-signal (DC -10 GHz), 2) RF transceivers (0.8–10 GHz), 3) power amplifiers and power management (0.8–10 GHz), and 4) millimeter wave (10–100 GHz). These frequencies refer to the operating frequencies of the radio systems and not to the device frequencies used in such systems. In addition to various active transistors of different structures in different materials, requirements and challenges in integrated passive components are discussed.

EMERGING RESEARCH DEVICES

The work on Emerging Research Devices aims at stimulating invention and research leading to feasibility demonstration of one or more Roadmap-extending concepts. It also aims at providing a balanced view of many of the exciting new approaches to information processing, articulating their potential contributions to extending the Roadmap balanced with brief discussion of their limiting challenges. In doing so it serves two purposes. The first is to provide a window into candidate approaches for advanced non-conventional or non-classical CMOS structures and memory techniques aimed at extending microelectronic technologies to the end of this Roadmap's timeframe. The second is to introduce and critique (without endorsement) new technological concepts for logic and architectures aimed at extending information and signal processing beyond the end of the current Roadmap. As such, the Emerging Research Devices activities form the bridge between bulk and non-classical CMOS and the realm of microelectronics beyond CMOS.

DIFFICULT CHALLENGES

Table 46a Process Integration Difficult Challenges—Near-term

Difficult Challenges ≥45 nm/Through 2010	*Summary of Issues*
1. High-performance applications: meeting performance and power dissipation requirements for highly scaled MOSFETs	Cost effectiveness, process control, and reliability of very thin oxy-nitride gate dielectrics, especially considering the high gate leakage Difficulty in controlling short-channel effects for highly scaled devices Negative impact of high channel doping needed for highly scaled devices. Also, the difficulty in controlling threshold voltage due to statistical fluctuations in the doping Need to reduce series S/D parasitic resistance Controlling static power dissipation in the face of rapidly increasing leakage: architecture and circuit design improvement and innovation will be needed.
2. Low-power applications: meeting performance and leakage requirements for highly scaled MOSFETs	Early availability of manufacturing-worthy high-κ gate dielectrics is necessary to meet stringent gate leakage and performance requirements. Slow scaling of V_{dd} for low standby power logic will make overall device scaling difficult. Rapid scaling of V_{dd} for low operating power logic will make overall device scaling difficult.
3. Assuring the reliability and implementing into manufacturing of multiple material, process, and structural changes in a relatively short period of time	Multiple material changes projected: high-κ gate dielectric, metal gate electrodes, strained Si, nickel silicide by 2008 or so Elevated S/D (selective epi) Ultra-thin body (UTB) SOI by 2008 or so, followed by multiple-gate structures. Near mid-gap metal gate electrodes will be desirable to set the threshold voltage for UTB SOI. Difficulty in ensuring reliability of all these new materials, processes, and structures in a timely manner
4. Implementation of DRAM, SRAM, and high-density nonvolatile memory (NVM) for scaled technologies	DRAM main issues—adequate storage capacitance for devices with reduced feature size, including difficulties in implementing high-κ storage dielectrics; access device design; holding the overall leakage to acceptably low levels; and deploying low sheet resistance materials for bit and word lines to ensure desired speed for scaled DRAMs SRAM—Difficulties with maintaining adequate noise margin and controlling key instabilities with scaling. Also, difficult lithography and etch issues with scaling NVM, flash—Scaling of tunnel dielectric and interpoly dielectric involves many complex tradeoffs. Dielectric material properties and dimensional control are key issues NVM, FeRAM—Ferroelectric material properties and dimensional control. Sensitivity to IC processing temperatures and conditions NVM, SONOS—ONO stack dimensions and material properties, including nitride layer trap distribution in space and energy NVM, MRAM—Magnetic material properties and dimensional control. Sensitivity to IC processing temperatures and conditions
5. High-performance and low-cost RF and analog/mixed-signal solutions	Signal isolation Optimizing RF/analog CMOS devices with scaled technologies: mismatch, 1/f noise, and leakage with high-κ gate dielectrics High density integrated passive element scaling and use of new materials: Q-factor value for inductors; matching and linearity for capacitors Reduced power supply voltages: degradation in SNR (signal-to-noise ratio) and signal distortion performance Reduced device breakdown voltage in scaled technologies High-frequency devices with increased operating voltage for base station applications Compound semiconductor substrates with good thermal dissipation and process equipment for fabrication at low cost See *section on RF and A/MS Technologies for Wireless Communications* for detailed discussion of these issues

Table 46b Process Integration Difficult Challenges—Long-term

Difficult Challenges < 45 nm/Beyond 2010	*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, ultra-thin body [UTB] MOSFETs) with lightly doped body will be needed to effectively scale MOSFETs to well under 20 nm gate length (L_g). Most likely, advanced material solutions such as strained Si (enhanced mobility) channels, elevated source/drain, high-κ gate dielectric, metal gate electrode, etc., will be utilized along with the advanced non-classical CMOS Particularly for the highly scaled UTB MOSFETs required towards the end of the Roadmap, with body thickness well under 10 nm, electrical performance and the impact of quantum effects are not well understood To attain adequate drive current for the highly scaled MOSFETs, quasi-ballistic operation with enhanced carrier saturation velocity appears to be needed See *Emerging Research Devices section* for more detail.
7. Dealing with atomic-level fluctuations and statistical process variations in sub-20 nm MOSFETs	Fundamental issues of atomic-level statistical fluctuations for sub-20 nm MOSFETs are not completely understood, including the impact of quantum effects.
8. Identifying, selecting, and implementing new memory structures	Highly scaled, dense, fast, non-volatile memory will become highly desirable Increasing difficulty is expected in scaling DRAMs, especially scaling down the dielectric equivalent oxide thickness, attaining the very low leakage currents that will be required, and reducing the cell area factor 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.
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 into a CMOS platform. Integration of the two may be difficult, especially for mixed signal. See *Emerging Research Devices* sections for more discussion and detail.

DESCRIPTION OF PROCESS INTEGRATION, DEVICES, AND STRUCTURES DIFFICULT CHALLENGES

[1] High-performance applications: meeting performance and power dissipation requirements for highly scaled MOSFETs—A key issue here is extending oxy-nitride gate dielectrics to an equivalent oxide thickness (EOT) of 1.0 nm and perhaps below. Statistical control of the thickness and ensuring reliability for such thin oxy-nitrides is expected to be a major challenge, and the high gate leakage for such oxy-nitrides is a major issue. With rapid transistor scaling, it will become increasingly difficult to meet the maximum allowable parasitic series source/drain resistance requirement. Also, with MOSFET scaling, it will become very difficult to control short-channel effects, and undesirably high channel doping is required to do so. Furthermore, the statistical variation of the channel doping and the resultant threshold voltage will become undesirably large. Finally, in the face of increased chip complexity and high leakage current with the succeeding years, chip static power dissipation is expected to become particularly difficult to control. Innovations in circuit design and architecture will be needed to design chips with both the desired performance and power dissipation.

[2] Low-power applications: meeting performance and leakage requirements for highly scaled MOSFETs—One key issue is that high-κ gate dielectric is projected to be required by about 2006 in order to meet stringent gate leakage current specifications. For LSTP, since V_{dd} hardly scales, the lateral electric field will become uncomfortably large with scaling, resulting in difficulties in controlling short channel effects and possibly reliability problems. For LOP, since V_{dd} scales relatively rapidly, it will become difficult to meet the device performance requirements.

[3] Assuring the reliability and implementing into manufacturing of multiple material, process, and structural changes in a relatively short period of time—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, nickel silicide, etc., will need to be implemented in less than a decade. The industry will be hard pressed to implement all these major changes in a timely manner. Eventually, it is expected 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, and this will be a major challenge for the industry. Furthermore, ensuring the reliability of all these innovations in a timely manner is expected to be particularly difficult.

[4] Implementation of DRAM, SRAM, and high density nonvolatile memory (NVM) for scaled technologies—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 leakage, in order to preserve adequate retention time. The requirement of low leakage currents causes problems in obtaining the desired access transistor performance. 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). Solving these SRAM challenges is critical to system performance, since SRAM is typically used for fast, on-chip memory. 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. 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.

[5] High-performance and low-cost RF and analog/mixed-signal solutions—Signal isolation between different function blocks (digital/analog/RF) on chip is a particularly difficult challenge in integrated system solutions for scaled technologies. Scaling of active devices, reduced power supply voltage, and use of new materials in scaled technologies present significant challenges in device matching, noise, linearity, signal-to-noise ratio (SNR), and device breakdown. Also, obtaining high density, high Q, well-matched, linear passive devices on chip will be a challenge. The demand for high-frequency and/or high operating voltage devices in emerging wireless communication applications requires continued device innovation and low- cost manufacturing infrastructure for compound semiconductor ICs. Finally, the difficulty and cost of integrating analog/RF (including eventually micro-electromechanical systems, MEMS, and possibly compound semiconductors such as GaAs and InP) and high-performance digital functions on a chip or a module are expected to increase over time.

[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 well under 20 nm, advanced non-classical structures such as 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. Finally, for these advanced, highly scaled MOSFETs, quasi-ballistic operation, in which the saturation velocity is enhanced, and hence the drive current is enhanced, is expected to become necessary to meet performance requirements. Implementation of these advanced, non-classical MOSFETs and understanding and controlling the quasi-ballistic operation is expected to be a major challenge for the industry.

[7] Dealing with atomic-level fluctuations and statistical process variations in sub-20 nm MOSFETs—For very short channel planar bulk devices, the total number of dopants in the depletion region is relatively small, and hence the statistical fluctuations are relatively large, limiting threshold voltage control. In addition, dimensional control of such short channel devices will become increasing difficult, further increasing the statistical process variations in MOSFET electrical parameters. For ultra-thin body, fully depleted, non-classical MOSFET devices, the statistical variation of the body thickness and the gate length are projected to be key issues.

[8] Identifying, selecting, and implementing new memory structures—In the long term, increasing difficulty is expected in scaling DRAMs, and 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 cannot be reduced by scaling, 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 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 power. LSTP chips are typically for lower performance 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) is 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 three years, reflecting the ultra-low leakage current required.

For generating the entries in the logic technology requirements tables, an approach was used in which simplified models were created and embedded in a spreadsheet. These models capture the essentials of the impact of such key input parameters as the power supply voltage, (V_{dd}), equivalent oxide thickness (EOT), gate length, etc., on the important transistor electrical output characteristics such as leakage current, saturation drive current, etc. 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 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. The spreadsheet capabilities are used to iteratively vary the input parameters until the target is met, and the final set of values for the input parameters is entered into the table. (Refer to the *detailed spreadsheets used to generate the Logic technology requirements tables*.) 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. An additional goal in generating the particular scaling scenarios in the tables was to delay as long as possible the projected need for major technology innovations such as high-κ gate dielectric, metal gate electrode, elevated source/drain, non-classical CMOS transistor structures, etc. (see Potential Solutions figures and text) while at the same time scaling each input parameter as regularly as possible. 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 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 and/or the channel doping, would have to be scaled faster 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, but there likely will be considerable variance in the actual paths that the various companies will take at each node.

For the high-performance logic tables, 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. 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 0.03 μA/μm in 2003, and it increases with succeeding years owing to the reduction of the saturation threshold voltage, V_t, with succeeding years. (At 0.1μA/μm for the long-term years, the subthreshold leakage current is especially high then.) Since $I_{d,sat}$ is strongly dependent on the overdrive, $(V_{dd}–V_t)$, and V_{dd} is decreasing with device scaling, V_t must be reduced along with V_{dd} to keep $I_{d,sat}$ up to the specified values. But $(1/I_{sd,leak})$ is exponentially dependent on V_t, and hence $I_{sd,leak}$ increases with succeeding years, as noted above. This increase in $I_{sd,leak}$ causes the static power dissipation per micron device width to generally increase with device scaling, despite the reduction in V_{dd} with scaling, and this has important consequences for the chip power dissipation (to be discussed below).

For high-performance chips, the rapid increase in subthreshold leakage current with scaling 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, and 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 or in circuits that are constantly switching, 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: in 2003, $I_{sd,leak}$ is 10 pA/μm for LSTP and 1 nA/μm for LOP, and in both cases the target 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 the historic 17% per year rate for both LOP and LSTP. One key issue for LSTP logic is the slow scaling of V_{dd}. Refer to Tables 48c and 48d for LSTP data. This issue is a result of the relatively slow scaling of V_t required to meet the very low subthreshold leakage current targets. V_{dd} must follow V_t in scaling slowly for two reasons: to obtain reasonable device performance the overdrive, $(V_{dd}-V_t)$, must remain relatively large, and for adequate circuit switching noise margins, V_{dd} must be no smaller than $2.3 \times V_t$. 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. Another issue is with the lateral electric field, which is approximately (V_{dd}/[gate length]). Due to the slow scaling of V_{dd}, the lateral field increases sharply with device scaling, and this sharp increase may result in difficulty in controlling short channel effects and possibly in reliability problems in the long-term years. In contrast to LSTP logic, V_{dd} scales relatively quickly for LOP logic (see technology requirements tables for LOP, Tables 48a and 48b), where, as mentioned above, the focus is on minimizing the operating power (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 and there is adequate circuit noise margin. The scaling of $I_{sd,leak}$ and of the MOSFET performance

metric, 1/τ, is plotted in *Figure 22* for high-performance, LOP, and LSTP logic. As expected, both $I_{sd,leak}$ and 1/τ are highest for high-performance logic, intermediate for LOP logic, and lowest for LSTP logic.

A critical issue is the gate leakage current, and whether oxy-nitride gate dielectric can meet the gate leakage current density limit as the oxy-nitride becomes increasingly thin with scaling (Refer to Tables 47a, 47b, 48a through 48d and to Notes [2] and [5]). 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, 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. The results are shown in *Figures 23 through 25*, respectively, 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 simulation data. EOT is also plotted for reference. For the LSTP and LOP transistors, the two Jg curves cross over each other just before 2006, and hence, for 2006 and beyond, the leakage current limit cannot be met using oxy-nitride because of direct tunneling. For LSTP in particular, the $J_{g,simulated}$ curve separates rapidly from the $J_{g,limit}$ curve after 2006, 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 2006. From Figure 23, note that EOT is the same for 2004 and 2005; this was done to slow the increase of gate leakage for those years, and to extend the ability of oxy-nitride to be utilized for the gate dielectric. For high-performance logic, from Figure 25, oxy-nitride gate dielectric will be unable to meet the gate current density limit in 2007 and beyond. The leading potential solution is to utilize high-κ gate dielectric in place of oxy-nitride. For more detail, refer to the Logic Potential Solutions section.

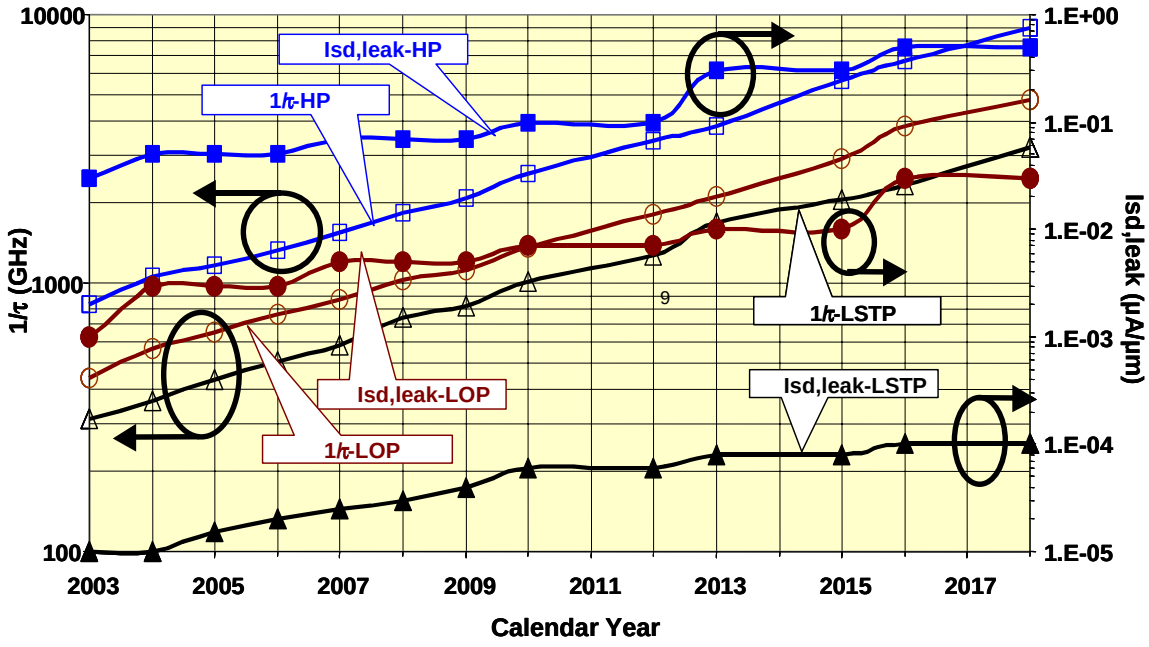


Figure 22 1/τ and $I_{sd,leak}$ Scaling for High-performance (HP), Low-operating Power, and Low-standby Power Logic

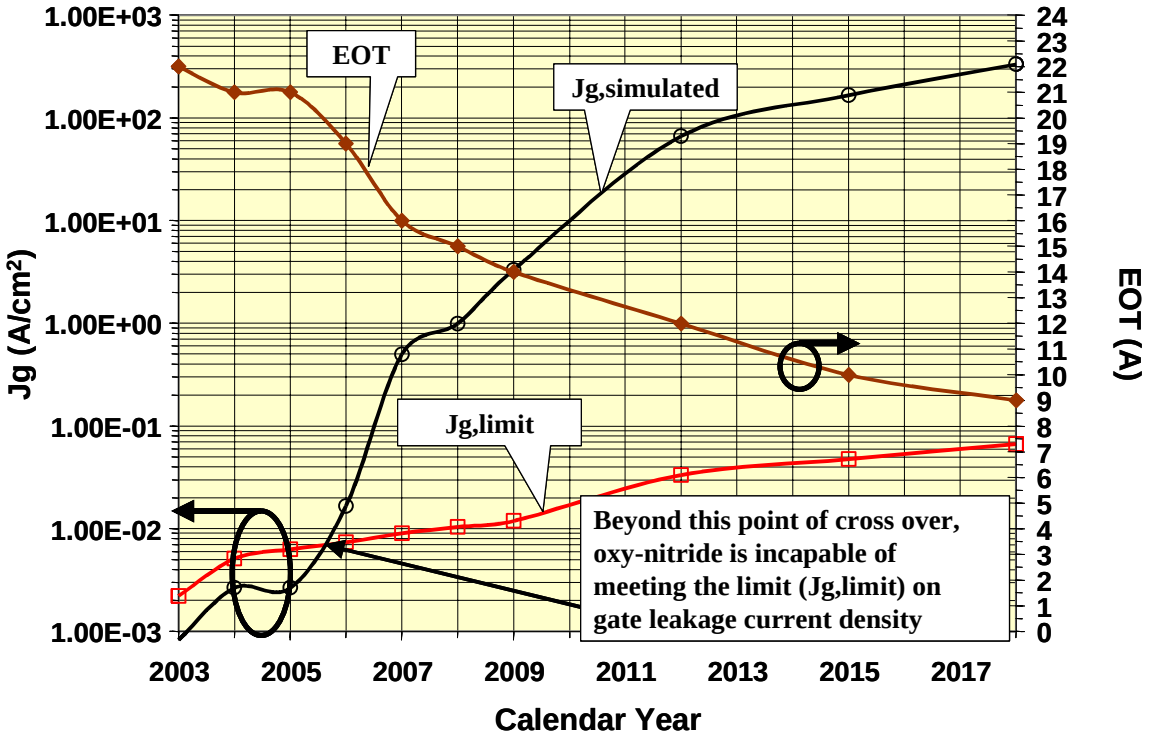


Figure 23 LSTP Logic Scaling-up of Gate Leakage Current Density Limit and of Simulated Gate Leakage due to Direct Tunneling

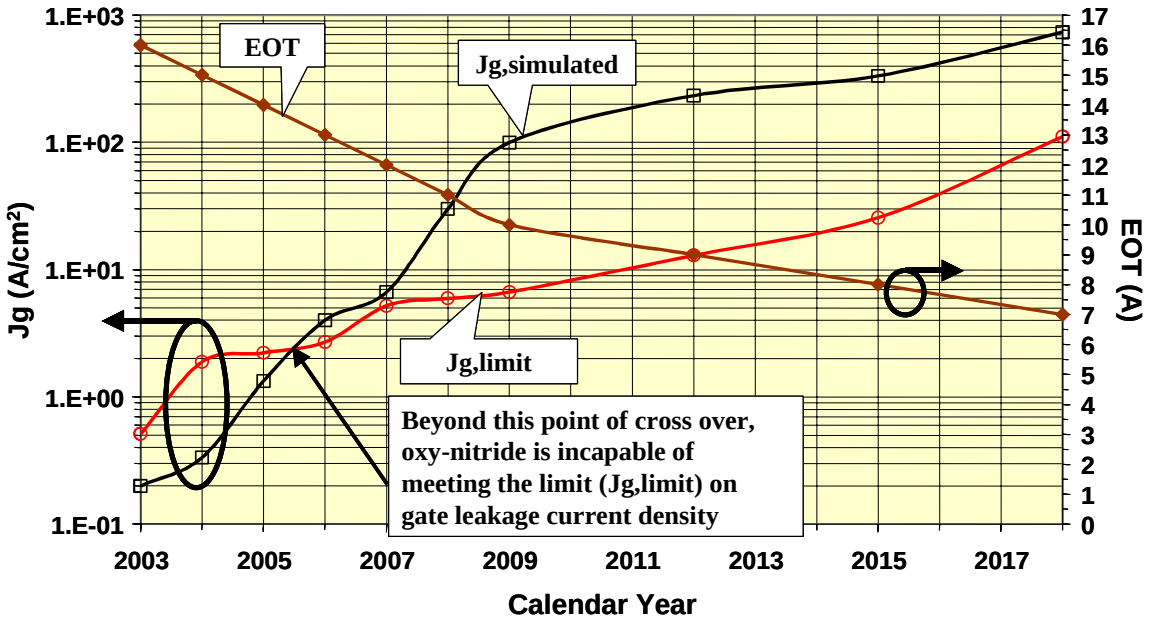


Figure 24 LOP Logic Scaling-up of Gate Leakage Current Density Limit and of Simulated Gate Leakage due to Direct Tunneling

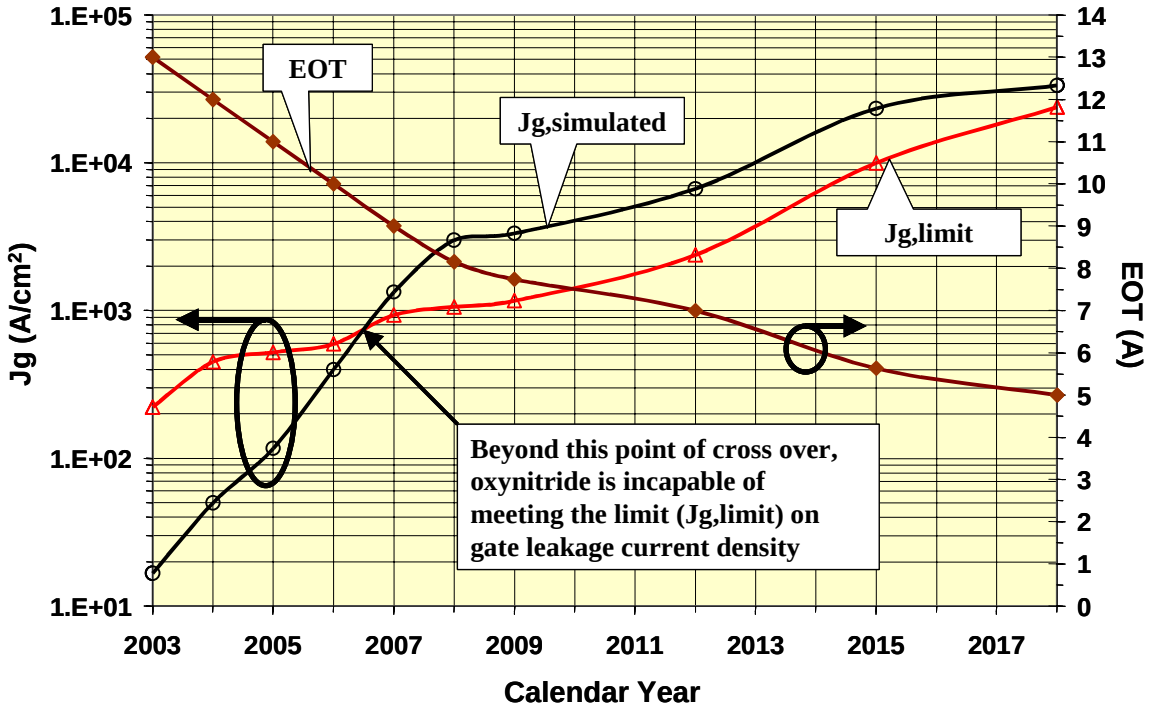


Figure 25 High-performance Logic Scaling-up of Gate Leakage Current Density Limit and of Simulated Gate Leakage due to Direct Tunneling

Table 47a High-performance Logic Technology Requirements—Near-term

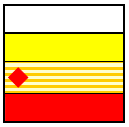
Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*120*	*107*	*95*	*85*	*76*	*67*	*60*
MPU/ASIC ½ Pitch (nm)	*107*	*90*	*80*	*70*	*65*	*57*	*50*
MPU Printed Gate Length (nm)	*65*	*53*	*45*	*40*	*35*	*32*	*28*
MPU Physical Gate Length (nm)	*45*	*37*	*32*	*28*	*25*	*22*	*20*
Physical gate length high-performance (HP) (nm) [1]	**45**	**37**	**32**	**28**	**25**	**22**	**20**
EOT: equivalent oxide thickness (physical) for high-performance (nm) [2]	**1.3**	**1.2**	**1.1**	**1.0**	**0.9**	**0.8**	**0.8**
Electrical thickness adjustment for gate depletion and inversion layer effects (nm) [3]	**0.8**	**0.8**	**0.7**	**0.7**	**0.4**	**0.4**	**0.4**
Equivalent electrical oxide thickness in inversion (nm) [4]	**2.1**	**2.0**	**1.8**	**1.7**	**1.3**	**1.2**	**1.2**
Nominal gate leakage current density limit (at 25°C) (A/cm^2) [5]	**2.2E+02**	**4.5E+02**	**5.2E+02**	**6.0E+02**	**9.3E+02**	**1.1E+03**	**1.2E+03**
Nominal power supply voltage (V_{dd}) (V) [6]	**1.2**	**1.2**	**1.1**	**1.1**	**1.1**	**1.0**	**1.0**
Saturation threshold voltage (V) [7]	**0.21**	**0.20**	**0.20**	**0.21**	**0.18**	**0.17**	**0.16**
Nominal high-performance NMOS sub-threshold leakage current, $I_{sd,leak}$ (at 25°C) (µA/µm) [8]	**0.03**	**0.05**	**0.05**	**0.05**	**0.07**	**0.07**	**0.07**
Nominal high-performance NMOS saturation drive current, $I_{d,sat}$ (at V_{dd}, at 25°C) (mA/mm) [9]	**980**	**1110**	**1090**	**1170**	**1510**	**1530**	**1590**
Required "mobility/transconductance improvement" factor [10]	**1.0**	**1.3**	**1.3**	**1.4**	**2.0**	**2.0**	**2.0**
Sub-threshold slope adjustment factor (full depletion/multiple-gate effects) (0–1) [11]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.8**	**0.7**
Effective saturation carrier velocity enhancement factor (due to quasi-ballistic transport) [12]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
Parasitic source/drain series resistance (R_{sd}) (Ohm-µm) [13]	**180**	**180**	**180**	**171**	**162**	**153**	**144**
Ideal NMOS device gate capacitance (F/µm) [14]	**7.40E-16**	**6.39E-16**	**6.14E-16**	**5.69E-16**	**6.64E-16**	**6.33E-16**	**5.76E-16**
Parasitic fringe/overlap capacitance (F/µm) [15]	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.30E-16**	**2.20E-16**	**2.00E-16**	**1.90E-16**
High-performance NMOS intrinsic delay, $\tau = C_{gate} * V_{dd} / I_{d,sat}$ (ps) [16]	**1.20**	**0.95**	**0.86**	**0.75**	**0.64**	**0.54**	**0.48**
Relative NMOS intrinsic switching speed, $1/\tau$, normalized to 2003 [17]	**1.00**	**1.26**	**1.39**	**1.60**	**1.86**	**2.20**	**2.49**
Nominal logic gate delay (NAND Gate) (ps) [18]	**30.24**	**23.94**	**21.72**	**18.92**	**16.23**	**13.72**	**12.13**
NMOSFET power-delay product (J/µm) [19]	**1.41E-15**	**1.27E-15**	**1.03E-15**	**9.66E-16**	**1.07E-15**	**8.33E-16**	**7.66E-16**
NMOSFET static power dissipation due to drain and gate leakage (W/µm) [20]	**3.96E-07**	**6.60E-07**	**6.05E-07**	**6.05E-07**	**8.47E-07**	**7.70E-07**	**7.70E-07**

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

Table 47b High-performance Logic Technology Requirements—Long-term

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*54*	*42*	*38*	*30*	*27*	*21*
MPU/ASIC ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU Printed Gate Length (nm)	*25*	*20*	*18*	*14*	*13*	*10*
MPU Physical Gate Length (nm)	*18*	*14*	*13*	*10*	*9*	*7*
Physical gate length high-performance (HP) (nm) [1]	**18**	**14**	**13**	**10**	**9**	**7**
EOT: equivalent oxide thickness (physical) for high-performance (nm) [2]	**0.7**	**0.7**	**0.6**	**0.6**	**0.5**	**0.5**
Electrical thickness adjustment for gate depletion and inversion layer effects (nm) [3]	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**
Equivalent electrical oxide thickness in inversion (nm) [4]	**1.1**	**1.1**	**1.0**	**1.0**	**0.9**	**0.9**
Nominal gate leakage current density limit (at 25°C) (A/cm^2) [5]	**1.9E+03**	**2.4E+03**	**7.7E+03**	**1.0E+04**	**1.9E+04**	**2.4E+04**
Nominal power supply voltage (V_{dd}) (V) [6]	**1.0**	**0.9**	**0.9**	**0.8**	**0.8**	**0.7**
Saturation threshold voltage (V) [7]	**0.15**	**0.14**	**0.11**	**0.12**	**0.10**	**0.11**
Nominal high-performance NMOS sub-threshold leakage current, $I_{sd,leak}$ (at 25°C) (mA/µm) [8]	**0.1**	**0.1**	**0.3**	**0.3**	**0.5**	**0.5**
Nominal high-performance NMOS saturation drive current, $I_{d,sat}$ (at V_{dd}, at 25°C) (mA/µm) [9]	**1900**	**1790**	**2050**	**2110**	**2400**	**2190**
Required "mobility/transconductance improvement" factor [10]	**2.0**	**2.0**	**2.0**	**2.0**	**2.0**	**2.0**
Sub-threshold slope adjustment factor (Full depletion/multiple-gate effects) (0–1) [11]	**0.6**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Effective saturation carrier velocity enhancement factor (due to quasi-ballistic transport) [12]	**1.1**	**1.1**	**1.1**	**1.3**	**1.3**	**1.3**
Parasitic source/drain series resistance (R_{sd}) (Ohm-µm) [13]	**135**	**116**	**107**	**88**	**79**	**60**
Ideal NMOS device gate capacitance (F/µm) [14]	**5.65E-16**	**4.39E-16**	**4.49E-16**	**3.45E-16**	**3.45E-16**	**2.69E-16**
Parasitic fringe/overlap capacitance (F/µm) [15]	**1.80E-16**	**1.50E-16**	**1.40E-16**	**1.20E-16**	**1.00E-16**	**8.00E-17**
High-performance NMOS intrinsic delay, $\tau = C_{gate} * V_{dd} / I_{d,sat}$ (ps) [16]	**0.39**	**0.30**	**0.26**	**0.18**	**0.15**	**0.11**
Relative NMOS intrinsic switching speed, 1/τ, normalized to 2003 [17]	**3.06**	**4.05**	**4.64**	**6.80**	**8.08**	**10.77**
Nominal logic gate delay (NAND gate) (ps) [18]	**9.88**	**7.47**	**6.52**	**4.45**	**3.74**	**2.81**
NMOSFET power-delay product (J/µm) [19]	**7.45E-16**	**4.77E-16**	**4.77E-16**	**2.98E-16**	**2.85E-16**	**1.71E-16**
NMOSFET static power dissipation due to drain and gate leakage (W/µm) [20]	**1.10E-06**	**9.90E-07**	**2.97E-06**	**2.64E-06**	**4.40E-06**	**3.85E-06**

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



Notes for Tables 47a and 47b:

The linked Microsoft Excel file contains the worksheet in which the details of the model-based scaling are described, including the formulas used in the MOSFET modeling. All the entries in this High-performance Logic Technology Requirements Table are from the worksheet. Please refer to this Excel file for detailed questions about the table.

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 (see Note [16]), while delaying as long as feasible the projected need for major innovations. These include innovations such 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 (see text for more detail on this point).*

[1] This is the final, as-etched length of the bottom of the gate electrode. Values set by ORTC. Gate dimensional control is set by the Lithography and FEP Etch ITWGs, and is assumed to have a three-sigma value of $\pm 10\% \times L_g$. It is expected that meeting this 10% 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. Yellow coloring in 2006 is set by FEP TWG projections of difficulties in achieving adequate thickness control and reliability capability for thin silicon oxy-nitride gate dielectrics. Red coloring for 2007 and beyond is due to projected inability of oxy-nitride gate dielectric to meet the gate leakage current density limits (see Note [5], and see text as well as Figure 25 for detail). Utilization of high-κ gate dielectric is a potential solution. 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 (approximately) for gate electrode depletion and inversion-layer effects, including quantum effects. The portion of the electrical thickness adjustment due to inversion-layer effects is assumed to remain constant at 0.4 nm. For polysilicon gate electrodes, the portion of the electrical thickness adjustment due to gate electrode depletion is dependent on the polysilicon doping. The yellow coloring for 2005 and 2006 reflects the FEP TWG assessment of difficulty in adequately doping polysilicon (particularly for the P^+, boron-doped electrodes) to meet the gate depletion thickness adjustment requirements. The red coloring for 2007 and beyond reflects the projected inability to adequately dope polysilicon gate electrodes to meet the gate depletion thickness adjustment requirements. Introduction of metal-gate electrodes, which reduce the gate depletion effect to zero, is a potential solution.

[4] 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} / (equivalent electrical oxide thickness), 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]) and of the CV^2 dynamic switching energy (see Note [19]). Red/yellow coloring follows that of EOT and Electrical Thickness Adjustment (see Notes [2] and [3] above).

[5] This is the maximum allowed gate leakage at 25°C, and is related to $I_{sd,leak}$, the nominal subthreshold leakage current per micron device width (see Note [8] below). Specifically, gate leakage current density limit = [$I_{sd,leak}$ / (physical gate length)] $\times$ [temp. factor] / [stack and overlap factor]. "Temp factor" = 10, and it accounts for the high operating temperature 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. Stack and overlap factor=3, and accounts for the different effects on $I_{sd,leak}$ and gate leakage current of stacked transistors in logic gates and of transistor gate overlap. The values of both temp factor and stack and overlap factor are rough order of magnitude estimates. The yellow and red coloring follows that of EOT (see Note [2] above).

[6] Nominal power supply voltage 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 (see Note 9), 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] Calculated threshold voltage for minimum nominal gate length transistor with drain bias set equal to V_{dd} (see Note [6]). 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]). Since control of short-channel effects for scaled MOSFETs is a key issue here, the yellow/red coloring follows that of the subthreshold slope adjustment factor (see Note [11] below).

[8] Nominal 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} (see Note [6]) and with the gate, source, and substrate biases set to zero volts. All MOSFET device dimensions are assumed to be at their nominal/target values. Total NMOS off-state current 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 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. Yellow and red coloring follows that of the subthreshold slope adjustment factor, which takes account of the impact of using advanced devices: ultra-thin body, fully depleted MOSFETs and multiple-gate MOSFETs (see Note [11] below). The above subthreshold, gate, and junction leakage current scaling scenario also applies to PMOS devices. Note that subthreshold current values here apply to the fastest MOSFETs only; slower/lower-leakage MOSFETs will also be available, since current and future chips consist/will consist of a mix of both high and lower-leakage devices (see text for further discussion).

[9] Nominal saturation current drive, $I_{d,sat}$, is defined as the NMOSFET drain current per micron device width, at 25°C, with the gate bias and the drain bias set equal to V_{dd} (see Note [6]) and the source and substrate biases set to zero; all MOSFET device dimensions are assumed to be at their nominal/target values. The saturation drive current values have been chosen to continue the historical approximate 17% per year device performance scaling (see Note [17] below). Nominal PMOS saturation current-drive value is assumed to be (40–50)% of the nominal NMOS saturation current-drive value. Yellow/red coloring follows that of three items: the parasitic source/drain series resistance, R_{sd} (see Note [13] below), the equivalent electrical oxide thickness in inversion (see Note [4]), and the required mobility/transconductance improvement factor (see Note [10]). Note that saturation current drive values here apply to the fastest MOSFETs only; lower saturation current drive/lower-leakage MOSFETs will also be available, since current and future chips consist/will consist of a mix of both high and lower-leakage devices (see text for further discussion).

[10] Fundamental device mobility/transconductance improvement (strained Si channel is the current implementation choice) is captured by a factor multiplying the carrier mobility. Such improvement is projected to be needed by 2004 in order to meet the required saturation current drive values (see Note [9]). Yellow coloring in 2004 and beyond reflects projected implementation of strained Si channel devices. Red coloring in 2007 and beyond reflects the difficulty in optimizing this mobility enhancement to a factor of 2.0, and the difficulty of implementing enhanced mobility channels for advanced, ultra-thin body devices.

[11] Subthreshold slope adjustment factor takes account of the impact of using advanced, single-gate, ultra-thin body, fully depleted SOI MOSFETs and eventually, ultra-thin body, multiple-gate MOSFETs, which are needed to control short-channel effects for highly scaled transistors. Specifically, this is a multiplying factor for the subthreshold slope, reducing it towards its minimum ideal value of 60 mV/decade. The factor ranges from 1.0 for classical, planar bulk MOSFETs, to 0.7 to 0.8 for single-gate, ultra-thin body MOSFETs, to 0.6 to 0.5 for ultra-thin body, multiple-gate MOSFETs. These numbers are rough estimates of the impact of these advanced devices. (See Non-Classical CMOS tables in the Emerging Research Devices section for further discussion of the ultra-thin body MOSFETs.) The yellow coloring reflects the projected introduction of the single-gate, ultra-thin body MOSFET in 2008, and the red coloring reflects the projected introduction of the multiple-gate MOSFET in 2010.

[12] This is a multiplying factor for carrier saturation velocity, reflecting quasi-ballistic transport in highly scaled, ultra-thin body MOSFETs, particularly multi-gate MOSFETs. The red coloring in 2010 and beyond reflects the projected need for saturation velocity enhancement in order to meet the required saturation current drive values (see Note [9]).

[13] Rsd 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.

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

[15] This is the parasitic gate overlap/fringing capacitance per micron device width [3× the overlap/fringing capacitance value per side, including the Miller effect]. These values are assumed to be independent of bias conditions.

[16] τ is the intrinsic transistor delay for NMOS devices at 25°C. $\tau = (C_{gate} \times V_{dd}) / I_{d,sat}$, where C_{gate} is the sum of the ideal device gate capacitance per micron device width (Note 14) and the parasitic gate overlap/fringing capacitance per micron device width (see Note [15]). τ for PMOSFETs is assumed to scale similarly, but with PMOS $I_{d,sat}$ ~ (0.4–0.5) × (NMOS $I_{d,sat}$) (see Note [9]). 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 current-drive (see Note [9]) and ideal gate capacitance (see Note [14]).

[17] NMOS performance metric ($1/\tau$, NMOS intrinsic switching speed--see Note [16]), normalized to the year 2003. Maintenance of the historical approximate 17% per year device performance improvement scaling trend is the key scaling goal for high-performance logic. Red/yellow coloring follows that of τ (see Note [16]).

[18] This is the calculated nominal delay for a 2-input, fan-out of 3, NAND gate, which is chosen to represent a typical logic gate (for details of the calculation, see link.). Red/yellow coloring follows that of τ (see Note [16]).

[19] This is the energy dissipated per micron of MOSFET width during a full switching cycle, defined as $C_{gate} \times V_{dd}^2$, where C_{gate} is the sum of the ideal device gate capacitance per micron width (Note [14]) and the parasitic gate overlap/fringing capacitance per micron width (see Note [15]). (This is the same C_{gate} used in calculating τ [see Note [16]). The dynamic power dissipation is directly related to this power-delay product. Red/yellow coloring follows that of $C_{g,ideal}$ (see Note [14]).

[20] This is the static power dissipation per micron of MOSFET width, defined as $V_{dd} \times [\{I_{sd,leak}$ (see Note 8)} + { maximum gate leakage current per micron device width (defined as L_g × gate leakage current density limit, from Note 5)}]. (The junction leakage current is assumed to be much smaller than either the source/drain subthreshold leakage current or the gate leakage current). Yellow/red coloring follows that of nominal gate leakage current density limit (see Note [5]).

Table 48a Low Operating Power (LOP) Logic Technology Requirements—Near-term

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*120*	*107*	*95*	*85*	*76*	*67*	*60*
MPU/ASIC ½ Pitch (nm)	*107*	*90*	*80*	*70*	*65*	*57*	*50*
MPU Printed Gate Length (nm)	*65*	*53*	*45*	*40*	*35*	*32*	*28*
MPU Physical Gate Length (nm)	*45*	*37*	*32*	*28*	*25*	*22*	*20*
Physical gate length low operating power (LOP) (nm) [1]	**65**	**53**	**45**	**37**	**32**	**28**	**25**
EOT: equivalent oxide thickness (physical) for LOP (nm) [2]	**1.6**	**1.5**	**1.4**	**1.3**	**1.2**	**1.1**	**1.0**
Electrical thickness adjustment for gate depletion and inversion layer effects (nm) [3]	**0.8**	**0.8**	**0.7**	**0.7**	**0.7**	**0.4**	**0.4**
Equivalent electrical oxide thickness in inversion (nm) [4]	**2.4**	**2.3**	**2.1**	**2.0**	**1.9**	**1.5**	**1.4**
Nominal gate leakage current density limit (at 25°C) (A/cm^2) [5]	**0.51**	**1.89**	**2.22**	**2.70**	**5.21**	**5.95**	**6.67**
Nominal LOP power supply voltage (V_{dd}) (V) [6]	**1.0**	**0.9**	**0.9**	**0.9**	**0.8**	**0.8**	**0.8**
Saturation threshold voltage (V) [7]	**0.31**	**0.26**	**0.27**	**0.28**	**0.26**	**0.25**	**0.25**
Nominal LOP NMOS sub-threshold leakage current, $I_{sd,leak}$ (at 25°C) ($\mu A/\mu m$) [8]	**1.0E-03**	**3.0E-03**	**3.0E-03**	**3.0E-03**	**5.0E-03**	**5.0E-03**	**5.0E-03**
Nominal LOP NMOS saturation drive current, $I_{d,sat}$ (at V_{dd}, at 25°C) ($\mu A/\mu m$) [9]	**520**	**530**	**580**	**610**	**570**	**730**	**770**
Required "mobility/transconductance improvement" factor [10]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.3**	**1.3**
Sub-threshold slope adjustment factor (full depletion/multiple-gate effects) (0–1) [11]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
Effective saturation carrier velocity enhancement factor (due to quasi-ballistic transport) [12]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
Parasitic source/drain series resistance (R_{sd}) (Ohm-μm) [13]	**180**	**180**	**180**	**180**	**180**	**180**	**180**
Ideal NMOS device gate capacitance ($F/\mu m$) [14]	**9.35E-16**	**7.96E-16**	**7.40E-16**	**6.39E-16**	**5.82E-16**	**6.45E-16**	**6.17E-16**
Parasitic fringe/overlap capacitance ($F/\mu m$) [15]	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.40E-16**
LOP NMOS intrinsic delay, $\tau = C_{gate} * V_{dd} / I_{d,sat}$ (ps) [16]	**2.26**	**1.76**	**1.52**	**1.30**	**1.15**	**0.97**	**0.89**
Relative NMOS intrinsic switching speed, $1/\tau$, normalized to 2003 [17]	**1.00**	**1.29**	**1.49**	**1.74**	**1.96**	**2.33**	**2.54**
Nominal logic gate delay (NAND gate) (ps) [18]	**57.0**	**44.3**	**38.3**	**32.7**	**29.1**	**24.4**	**22.4**
NMOSFET power-delay product ($J/\mu m$) [19]	**1.18E-15**	**8.39E-16**	**7.94E-16**	**7.12E-16**	**5.26E-16**	**5.66E-16**	**5.48E-16**
NMOSFET static power dissipation due to drain and gate leakage ($W/\mu m$) [20]	**2.0E-09**	**5.4E-09**	**5.4E-09**	**5.4E-09**	**8.0E-09**	**8.0E-09**	**8.0E-09**

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

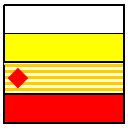


Table 48b Low Operating Power (LOP) Logic Technology Requirements—Long-term

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*54*	*42*	*38*	*30*	*27*	*21*
MPU/ASIC ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU Printed Gate Length (nm)	*25*	*20*	*18*	*14*	*13*	*10*
MPU Physical Gate Length (nm)	*18*	*14*	*13*	*10*	*9*	*7*
Physical gate length low operating power (LOP) (nm) [1]	**22**	**18**	**16**	**13**	**11**	**9**
EOT: equivalent oxide thickness (physical) for LOP (nm) [2]	**0.9**	**0.9**	**0.8**	**0.8**	**0.7**	**0.7**
Electrical thickness adjustment for gate depletion and inversion layer effects (nm) [3]	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**
Equivalent electrical oxide thickness in inversion (nm) [4]	**1.3**	**1.3**	**1.2**	**1.2**	**1.1**	**1.1**
Nominal gate leakage current density limit (at 25°C) (A/cm^2) [5]	**11**	**13**	**21**	**26**	**91**	**111**
Nominal LOP power supply voltage (V_{dd}) (V) [6]	**0.7**	**0.7**	**0.6**	**0.6**	**0.5**	**0.5**
Saturation threshold voltage (V) [7]	**0.22**	**0.23**	**0.21**	**0.19**	**0.16**	**0.17**
Nominal LOP NMOS sub-threshold leakage current, $I_{sd,leak}$ (at 25°C) (μA/μm) [8]	**0.007**	**0.007**	**0.01**	**0.01**	**0.03**	**0.03**
Nominal LOP NMOS saturation drive current, $I_{d,sat}$ (at V_{dd}, at 25°C) (μA/μm) [9]	**770**	**830**	**780**	**900**	**920**	**950**
Required "mobility/transconductance improvement" factor [10]	**1.3**	**2.0**	**2.0**	**2.0**	**2.0**	**2.0**
Sub-threshold slope adjustment factor (full depletion/multiple-gate effects) (0–1) [11]	**0.8**	**0.8**	**0.8**	**0.5**	**0.5**	**0.5**
Effective saturation carrier velocity enhancement factor (due to quasi-ballistic transport) [12]	**1.0**	**1.0**	**1.1**	**1.1**	**1.3**	**1.3**
Parasitic source/drain series resistance (R_{sd}) (Ohm-μm) [13]	**160**	**135**	**126**	**107**	**98**	**80**
Ideal NMOS device gate capacitance (F/μm) [14]	**5.84E-16**	**4.78E-16**	**4.60E-16**	**3.74E-16**	**3.45E-16**	**2.83E-16**
Parasitic fringe/overlap capacitance (F/μm) [15]	**2.20E-16**	**1.80E-16**	**1.60E-16**	**1.40E-16**	**1.30E-16**	**1.10E-16**
LOP NMOS intrinsic delay, $\tau = C_{gate} * V_{dd} / I_{d,sat}$ (ps) [16]	**0.73**	**0.56**	**0.48**	**0.34**	**0.26**	**0.21**
Relative NMOS intrinsic switching speed, 1/τ, normalized to 2003 [17]	**3.1**	**4.1**	**4.7**	**6.6**	**8.7**	**10.9**
Nominal logic gate delay (NAND Gate) (ps) [18]	**18.4**	**14.0**	**12.0**	**8.6**	**6.5**	**5.2**
NMOSFET power-delay product (J/μm) [19]	**3.94E-16**	**3.22E-16**	**2.23E-16**	**1.85E-16**	**1.19E-16**	**9.81E-17**
NMOSFET static power dissipation due to drain and gate leakage (W/μm) [20]	**9.8E-09**	**9.8E-09**	**1.2E-08**	**1.2E-08**	**3.0E-08**	**3.0E-08**

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

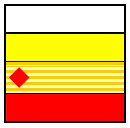


Table 48c Low Standby Power (LSTP) Technology Requirements—Near-term

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*120*	*107*	*95*	*85*	*76*	*67*	*60*
MPU/ASIC ½ Pitch (nm)	*107*	*90*	*80*	*70*	*65*	*57*	*50*
MPU Printed Gate Length (nm)	*65*	*53*	*45*	*40*	*35*	*32*	*28*
MPU Physical Gate Length (nm)	*45*	*37*	*32*	*28*	*25*	*22*	*20*
Physical gate length low standby power (LSTP) (nm) [1]	**75**	**65**	**53**	**45**	**37**	**32**	**28**
EOT: equivalent oxide thickness (physical) for LSTP (nm) [2]	**2.2**	**2.1**	**2.1**	**1.9**	**1.6**	**1.5**	**1.4**
Electrical thickness adjustment for gate depletion and inversion layer effects (nm) [3]	**0.8**	**0.8**	**0.7**	**0.7**	**0.7**	**0.4**	**0.4**
Equivalent electrical oxide thickness in inversion (nm) [4]	**3**	**2.9**	**2.8**	**2.6**	**2.3**	**1.9**	**1.8**
Nominal gate leakage current density limit (at 25°C) (A/cm^2) [5]	**4.4E-03**	**5.1E-03**	**9.4E-03**	**1.5E-02**	**2.3E-02**	**3.1E-02**	**4.8E-02**
Nominal LSTP power supply voltage (V_{dd}) (V) [6]	**1.2**	**1.2**	**1.2**	**1.2**	**1.1**	**1.1**	**1.1**
Saturation threshold voltage (V) [7]	**0.50**	**0.50**	**0.51**	**0.52**	**0.50**	**0.47**	**0.47**
Nominal LSTP NMOS sub-threshold leakage current, $I_{sd,leak}$ (at 25°C) ($\mu A/\mu m$) [8]	**1.0E-05**	**1.0E-05**	**1.5E-05**	**2.0E-05**	**2.5E-05**	**3.0E-05**	**4.0E-05**
Nominal LSTP NMOS saturation drive current, $I_{d,sat}$ (at V_{dd}, at 25°C) ($\mu A/\mu m$) [9]	**410**	**440**	**470**	**510**	**510**	**670**	**700**
Required "mobility/transconductance improvement" factor [10]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.3**	**1.3**
Sub-threshold slope adjustment factor (full depletion/multiple-gate effects) (0–1) [11]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
Effective saturation carrier velocity enhancement factor (due to quasi-ballistic transport) [12]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
Parasitic source/drain series resistance (R_{sd}) (Ohm-μm) [13]	**180**	**180**	**180**	**180**	**180**	**180**	**180**
Ideal NMOS device gate capacitance ($F/\mu m$) [14]	**8.63E-16**	**7.74E-16**	**6.54E-16**	**5.98E-16**	**5.55E-16**	**5.82E-16**	**5.37E-16**
Parasitic fringe/overlap capacitance ($F/\mu m$) [15]	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.40E-16**	**2.40E-16**
LSTP NMOS intrinsic delay, $\tau = C_{gate} * V_{dd} / I_{d,sat}$ (ps) [16]	**3.23**	**2.77**	**2.28**	**1.97**	**1.72**	**1.35**	**1.22**
Relative NMOS intrinsic switching speed, $1/\tau$, normalized to 2003 [17]	**1.00**	**1.17**	**1.42**	**1.64**	**1.88**	**2.39**	**2.64**
Nominal logic gate delay (NAND Gate) (ps) [18]	**81.4**	**69.7**	**57.5**	**49.7**	**43.2**	**34.0**	**30.8**
NMOSFET power-delay product ($J/\mu m$) [19]	**1.6E-15**	**1.5E-15**	**1.3E-15**	**1.2E-15**	**9.6E-16**	**9.9E-16**	**9.4E-16**
NMOSFET static power dissipation due to drain and gate leakage ($W/\mu m$) [20]	**2.4E-11**	**2.4E-11**	**3.6E-11**	**4.8E-11**	**5.5E-11**	**6.6E-11**	**8.8E-11**

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

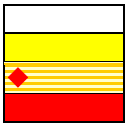
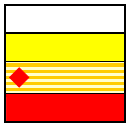


Table 48d Low Standby Power (LSTP) Technology Requirements—Long-term

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*54*	*42*	*38*	*30*	*27*	*21*
MPU/ASIC ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU Printed Gate Length (nm)	*25*	*20*	*18*	*14*	*13*	*10*
MPU Physical Gate Length (nm)	*18*	*14*	*13*	*10*	*9*	*7*
Physical gate length low standby power (LSTP) (nm) [1]	**25**	**20**	**18**	**14**	**13**	**10**
EOT: equivalent oxide thickness (physical) for LSTP (nm) [2]	**1.3**	**1.2**	**1.1**	**1.1**	**1.0**	**0.9**
Electrical thickness adjustment for gate depletion and inversion layer effects (nm) [3]	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**
Equivalent electrical oxide thickness in inversion (nm) [4]	**1.7**	**1.6**	**1.5**	**1.5**	**1.4**	**1.3**
Nominal gate leakage current density limit (at 25°C) (A/cm^2) [5]	**8.00E-02**	**1.00E-01**	**1.48E-01**	**1.90E-01**	**2.56E-01**	**3.33E-01**
Nominal LSTP power supply voltage (V_{dd}) (V) [6]	**1.0**	**1.0**	**0.9**	**0.9**	**0.8**	**0.8**
Saturation threshold voltage (V) [7]	**0.39**	**0.43**	**0.34**	**0.38**	**0.36**	**0.40**
Nominal LSTP NMOS sub-threshold leakage current, $I_{sd,leak}$ (at 25°C) ($\mu A/\mu m$) [8]	**6.0E-05**	**6.0E-05**	**8.0E-05**	**8.0E-05**	**1.0E-04**	**1.0E-04**
Nominal LSTP NMOS saturation drive current, $I_{d,sat}$ (at V_{dd}, at 25°C) ($\mu A/\mu m$) [9]	**760**	**790**	**880**	**870**	**860**	**990**
Required "mobility/transconductance improvement" factor [10]	**1.3**	**1.3**	**1.3**	**1.3**	**2.0**	**2.0**
Sub-threshold slope adjustment factor (full depletion/multiple-gate effects) (0–1) [11]	**0.8**	**0.8**	**0.5**	**0.5**	**0.5**	**0.5**
Effective saturation carrier velocity enhancement factor (due to quasi-ballistic transport) [12]	**1.0**	**1.0**	**1.0**	**1.0**	**1.1**	**1.3**
Parasitic source/drain series resistance (R_{sd}) (Ohm-μm) [13]	**180**	**144**	**135**	**116**	**107**	**88**
Ideal NMOS device gate capacitance ($F/\mu m$) [14]	**5.08E-16**	**4.32E-16**	**4.14E-16**	**3.22E-16**	**3.21E-16**	**2.66E-16**
Parasitic fringe/overlap capacitance ($F/\mu m$) [15]	**2.40E-16**	**1.90E-16**	**1.70E-16**	**1.50E-16**	**1.40E-16**	**1.20E-16**
LSTP NMOS intrinsic delay, $\tau = C_{gate} * V_{dd} / I_{d,sat}$ (ps) [16]	**0.98**	**0.79**	**0.60**	**0.49**	**0.43**	**0.31**
Relative NMOS intrinsic switching speed, $1/\tau$, normalized to 2003 [17]	**3.28**	**4.10**	**5.40**	**6.61**	**7.54**	**10.36**
Nominal logic gate delay (NAND gate) (ps) [18]	**24.80**	**19.83**	**15.06**	**12.31**	**10.80**	**7.85**
NMOSFET power-delay product ($J/\mu m$) [19]	**7.48E-16**	**6.22E-16**	**4.73E-16**	**3.83E-16**	**2.95E-16**	**2.47E-16**
NMOSFET static power dissipation due to drain and gate leakage ($W/\mu m$) [20]	**1.20E-10**	**1.20E-10**	**1.44E-10**	**1.44E-10**	**1.60E-10**	**1.60E-10**

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



Notes for Tables 48a through 48d:

Microsoft Excel file links contain the worksheets in which the details of the model-based scaling are described, including the formulas used in the MOSFET modeling. All the entries in these LOP and LSTP Logic Technology Requirements Tables are from the worksheets. Please refer to these Excel files for detailed questions about the tables.

The scaling of the numbers in each of the tables reflects a particular scaling scenario in which we have attempted to optimally scale to meet the key goal for low-power logic while delaying as long as feasible the projected need for such major innovations as metal gate electrodes, high-κ gate dielectric, and novel doping and annealing techniques to reduce the value of the parasitic series source/drain resistance. The key goal consists of meeting the targets for sub-threshold leakage current (see Note [8]) in the tables. However, there are numerous parameters (such as EOT, V_{dd}, $I_{sd,leak}$, Rsd, etc.) which can be varied and traded off, and different scaling scenarios are possible by making different choices on the scaling of these parameters (see text for more detail on this point).

[1] This is 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 three years (LSTP)(See text for further discussion.). Gate dimensional control is set by the Lithography and FEP Etch ITWGs, and is assumed to have a three-sigma value of $\pm 10\% \times L_g$. It is expected that meeting this 10% 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. For both LOP and LSTP, red coloring for 2006 and beyond is due to projected inability of oxy-nitride gate dielectric to meet the gate leakage current density limits (see Note [5], and see text as well as Figures 23 and 24 for detail). Utilization of high-κ gate dielectric is a potential solution.

[3] Accounts (approximately) for gate electrode depletion and inversion-layer effects, including quantum effects. The portion of the electrical thickness adjustment due to inversion-layer effects is assumed to remain constant at 0.4 nm. For polysilicon gate electrodes, the portion of the electrical thickness adjustment due to gate electrode depletion is dependent on the polysilicon doping. The red coloring for 2008 and beyond reflects the projected inability to adequately dope polysilicon gate electrodes to meet the gate depletion thickness adjustment requirements. Introduction of metal-gate electrodes, which reduce the gate depletion effect to zero, is a potential solution.

[4] 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} / (equivalent electrical oxide thickness), 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]) and of the CV^2 dynamic switching energy (see Note [19]). Red/yellow coloring follows that of EOT and Electrical Thickness Adjustment (see Notes [2] and [3] above).

[5] This is the maximum allowed gate leakage at 25°C, and is related to $I_{sd,leak}$, the nominal sub-threshold leakage current per micron device width, (see Note [8] below). Specifically, gate leakage current density limit = [$I_{sd,leak}$ / (physical gate length)] $\times$ [temp. factor] / [stack and overlap factor]. "Temp factor" = 1, because LOP and LSTP logic are expected to operate near room temperature, and temp. factor adjusts for high-temperature operation. "Stack & overlap factor" = 3, and accounts for the different effects on $I_{sd,leak}$ and gate leakage current of stacked transistors in logic gates and of transistor gate overlap. The values of both temp factor and stack and overlap factor are rough order of magnitude estimates. The yellow and red coloring follows that of EOT (see Note [2] above).

[6] Nominal power supply voltage has been set to the smallest value sufficient to maintain acceptable circuit switching noise margin (approximately 2.3 times the threshold voltage). Target power supply voltage values for actual ICs may vary $\pm 10\%$ (or more) from the values in this table, depending on the particular circuit design application or technology optimization. Due to different system applications, the power-supply voltage values for LOP and LSTP logic have been optimized and set independently of each other. In particular, the LOP values for V_{dd} are minimized as much as possible to enable the dynamic power dissipation, which is proportional to V_{dd}^2, to be minimized. Note that meeting overall system power dissipation requirements will require the use of circuit/system techniques to "turn-off" or "power-down" various circuit blocks (see text for further discussion of this point). Another issue is maintaining sufficient voltage over-drive [V_{dd} – saturation threshold voltage (see Note [7])] in order to meet the saturation current drive values (see Note [9]).

[7] Calculated threshold voltage for minimum nominal gate length transistor with drain bias set equal to V_{dd} (see Note [6]). The threshold voltage values have been chosen to meet the sub-threshold leakage current requirements (see Note [8]). Since control of short-channel effects for scaled MOSFETs is a key issue here, the yellow/red coloring follows that of the Sub-threshold Slope Adjustment Factor (see Note [11] below).

[8] Nominal sub-threshold 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} (see Note [6]) and with the gate, source, and substrate biases set to zero volts. All MOSFET device dimensions are assumed to be at their nominal/target values. Total NMOS off-state leakage current is the NMOSFET drain current per micron of device width at 25°C, and is the sum of the NMOS sub-threshold, gate, and junction leakage current components. The sub-threshold 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. For LOP and especially LSTP logic, the sub-threshold leakage current is considerably less than for high-performance logic; attaining the required low values of sub-threshold leakage current is the main goal of scaling for LOP and LSTP logic. Yellow and red coloring follows that of the Subthreshold Slope Adjustment Factor, which takes account of the impact of using advanced devices: ultra-thin body, fully depleted MOSFETs and multiple-gate MOSFETs (see Note [11] below). The above subthreshold, gate, and junction leakage current scaling scenario also applies to PMOS devices. Note that subthreshold current values here apply to the fastest MOSFETs only; slower/lower-leakage MOSFETs will also be available, since current and future chips consist/will consist of a mix of both high and lower-leakage devices (see text for further discussion).

[9] Nominal saturation current drive, $I_{d,sat}$, is defined as the NMOSFET drain current per micron device width, at 25°C, with the gate bias and the drain bias set equal to V_{dd} (see Note [6]) and the source and substrate biases set to zero; all MOSFET device dimensions are assumed to be at their nominal/target values. Nominal PMOS saturation current-drive value is assumed to be (40–50)% of the nominal NMOS saturation current-drive value. Yellow/red coloring follows that of parasitic source/drain series resistance, R_{sd} (see Note [13] below) and the equivalent electrical oxide thickness in inversion (Note [4]. Note that saturation current drive values here apply to fastest MOSFETs only; lower saturation current drive/lower leakage MOSFETs will also be available, since current and future chips consist/will consist of a mix of both high and lower-leakage devices (see text for further discussion).

[10] Fundamental device mobility/transconductance improvement (strained Si channel is the current implementation choice) is captured by a factor multiplying the carrier mobility. Such improvement is projected to be needed by 2008 in order to meet the required saturation current drive values (see Note [9]). Yellow coloring in 2008 and beyond reflects difficulty in projected implementation of strained Si channel devices. Red coloring in 2012 (LOP) and in 2016 (LSTP) and beyond reflects the difficulty in optimizing this mobility enhancement to a factor of 2.0, and the difficulty in implementing enhanced mobility channels for advanced ultra-thin body devices.

[11] Sub-threshold Slope Adjustment Factor takes account of the impact of using advanced, single-gate, ultra-thin body, fully depleted SOI MOSFETs and eventually, ultra-thin body, multiple-gate MOSFETs, which are needed to control short-channel effects for highly scaled transistors. Specifically, this is a multiplying factor for the sub-threshold slope, reducing it towards its minimum ideal value of 60 mV/decade. The factor ranges from 1.0 for classical, planar bulk MOSFETs, to 0.7 to 0.8 for single-gate, ultra-thin body MOSFETs, to 0.6 to 0.5 for ultra-thin body, multiple-gate MOSFETs. These numbers are rough estimates of the impact of these advanced devices. (See the Non-Classical CMOS tables in the Emerging Research Devices section for further discussion of the ultra-thin body MOSFETs.) The yellow coloring reflects the projected introduction of the single-gate, ultra-thin body MOSFET in 2010, and the red coloring reflects the projected introduction of the multiple-gate MOSFET in 2013 (LSTP) and in 2015 (LOP).

[12] This is a multiplying factor for carrier saturation velocity, reflecting quasi-ballistic transport in highly scaled, ultra-thin body MOSFETs, particularly multi-gate MOSFETs. The red coloring in 2013 and beyond (LOP) and in 2016 and beyond (LSTP) reflects the projected need for saturation velocity enhancement in order to meet the required saturation current drive values (see Note [9]).

[13] R_{sd} is the maximum allowable parasitic series source plus drain resistance per micron device width. In particular, the R_{sd} 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.

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

[15] This is the parasitic gate overlap/fringing capacitance per micron device width [3× the overlap/fringing capacitance value per side, including the Miller effect]. These values are assumed to be independent of bias conditions.

[16] τ is the intrinsic transistor delay for NMOS devices at 25°C. $\tau = (C_{gate} \times V_{dd}) / I_{d,sat}$, where C_{gate} is the sum of the ideal device gate capacitance per micron device width (Note [14]) and the parasitic gate overlap/fringing capacitance per micron device width (see Note [15]). τ for PMOSFETs is assumed to scale similarly, but with PMOS $I_{d,sat}$ ~ (0.4–0.5) × (NMOS $I_{d,sat}$) (see Note [9]). 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 saturation current drive (see Note [9]) and ideal gate capacitance (see Note [14]).

[17] NMOS performance metric ($1/\tau$, NMOS intrinsic switching speed—see Note [16]), normalized to the year 2003. The historical approximate 17% per year device performance improvement scaling trend is maintained, even while meeting the key scaling goal for LSTP and LOP logic of low source/drain sub-threshold leakage current (see Note [8]). Red/yellow coloring follows that of τ (see Note [16]).

[18] This is the calculated nominal delay for a 2-input, fan-out of 3, NAND gate, which is chosen to represent a typical logic gate (for details of the calculation refer to the supplemental link). Red/yellow coloring follows that of τ (see Note [16]). See the LOP and LSTP spreadsheets.

[19] This is the energy dissipated per micron of MOSFET width during a full switching cycle, defined as $C_{gate} \times V_{dd}^2$, where C_{gate} is the sum of the ideal device gate capacitance per micron width (Note [14]) and the parasitic gate overlap/fringing capacitance per micron width (see Note [15]). (This is the *same C_{gate} used in calculating τ [see Note [16]). The dynamic power dissipation is directly related to this power-delay product. Red/yellow coloring follows that of $C_{g,ideal}$ (see Note [14]).*

[20] This is the static power dissipation per micron of MOSFET width, defined as $V_{dd} \times [\{I_{sd,leak}$ (see Note [8])} + {maximum gate leakage current per micron device width (defined as $L_g \times$ gate leakage current density limit, from Note [5])}]. (The junction leakage current is assumed to be much smaller than either the source/drain subthreshold leakage current or the gate leakage current). Yellow/red coloring follows that of nominal gate leakage current density limit (see Note [5]).

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) or in some cases, yellow coloring (manufacturable solutions are 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 the potential solutions listed in the figure involve major, significant changes, and to successfully accomplish these changes, including understanding any new and different reliability, yield, and process integration issues, the qualification/pre-production interval has been set to one and a half years.

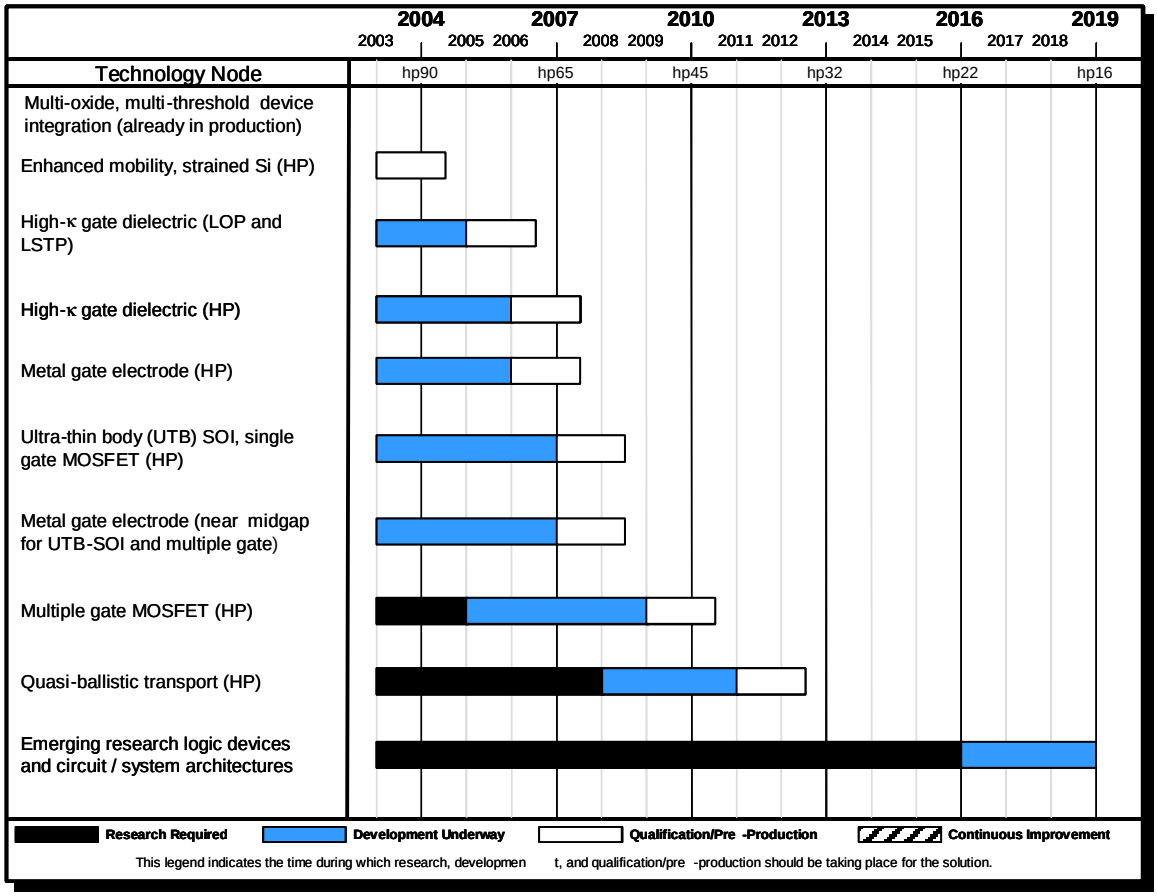
The first potential solution, integrating multiple transistor types with several threshold voltages and several equivalent oxide thicknesses, is currently in production for high-performance logic. As discussed in the text on logic technology requirements,

this is important to controlling the chip static power dissipation, and has the added benefit of improving the flexibility for system-on-a-chip (SoC) applications. The next potential solution, enhanced mobility using strained silicon channels, is needed to enhance the saturation current drive, and is projected to be implemented by 2004 for high-performance logic. It is projected to be implemented considerably later for low-power logic, perhaps in about 2008, because of the slower scaling and lower device performance compared to high-performance logic.

In order to scale the basic MOSFET structure significantly below the 90 nm node in 2004 (physical gate length of 37 nm), 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 continues to be reduced, a corresponding reduction in the gate dielectric equivalent oxide thickness (EOT) is necessary in order to control short-channel effects, including drain-induced barrier-lowering-induced off-state current and degraded subthreshold slope. However, continued thinning of the conventional gate dielectric, oxy-nitride, results in a significant increase in gate leakage current due to sharply increasing 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. See the section on logic technology requirements, including Tables 47a and b and 48a through 48d, and Figures 23–25, for more detail on these points. For LOP and LSTP logic, high-κ gate dielectric material is a potential solution projected to solve the problem of high gate leakage current in 2006 and beyond. The EOT is relatively large and the gate leakage limit is relatively low, particularly for LSTP logic (EOT = 1.9 nm, nominal gate leakage current density limit = 0.015 A/cm^2 in 2006). For high-performance logic, high-κ gate dielectric is projected for 2007, when the EOT = 0.9 nm and oxy-nitride could not meet the nominal gate leakage current density limit of 930 A/cm^2. Metal gate electrodes are also projected for 2007, in order to effectively prevent gate electrode depletion and hence allow acceptable scaling of the equivalent electrical oxide thickness. Because the gate electrode needs a work function near the silicon valence band for PMOSFETs and near the silicon conduction band for NMOSFETs, different metals will probably be needed for the PMOSFET and NMOSFET. Alternatively, a single metal type that can be tuned to the appropriate work functions for both PMOSFETs and NMOSFETS, e.g., by doping or other methods, is being researched and may be utilized. Metal gate electrodes are likely to be implemented somewhat later for LOP and LSTP logic, because the EOT scaling lags behind that of high-performance logic.

As scaling proceeds to the 65 nm node in 2007 (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 junction leakage current. 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-mid-gap work functions (slightly above the silicon mid-gap for NMOSFETs and slightly below the mid-gap for PMOSFETs) will likely be used, and because of the different work functions in this case, these will presumably be different than those utilized for planar bulk MOSFETs. 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 2010. 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, which will result in enhanced saturation velocity and hence enhanced saturation current drive.

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*). It is expected that such solutions will be integrated 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.



HP—high-performance

Figure 26 Logic Potential Solutions

MEMORY TECHNOLOGY REQUIREMENTS AND POTENTIAL SOLUTIONS

DRAM TECHNOLOGY REQUIREMENTS AND POTENTIAL SOLUTIONS

Technical requirements for DRAMs are projected to become increasingly stringent with scaling. Photoresists associated with 193 nm exposure wavelength lithography, which is slated for production in 2003 at 100 nm DRAM half pitch, had serious limitations on their etch selectivity and resistance. These limitations result in significant process flow issues for both trench or stack capacitor structures, since process steps such as capacitor formation or high aspect ratio contact etches require a resist that can stand up for a prolonged etch time. As a result of effort during 2002 and at the early part of 2003, the technology level related photoresists has been improved, resulting in early production of 100 nm DRAM but continuous improvement is still necessary. 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 (pass gate) device for the 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. 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 sharply to maintain adequate storage capacitance with scaling. To scale EOT as projected in Tables 49a and b, dielectric materials having high relative dielectric constant (κ) will be needed. Several manufactures are pursuing MIS (Metal Insulator Semiconductor) capacitors using Ta_2O_5 and Al_2O_3 (κ~ 10–25) through the 90 nm node. Eventually, beyond the 90 nm node in 2004, MIM (Metal Insulator Metal) structures and dielectric materials with even higher κ 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 at the 70 nm node in 2007. And also the

physical thickness of the high-κ insulator should be scaled down to fit minimum feature size. All in all, maintaining sufficient storage capacitance will pose an increasingly difficult requirement for continued scaling of DRAM devices.

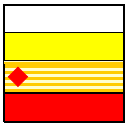
Table 49a DRAM Technology Requirements—Near-term

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm) [1]	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm)	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
DRAM cell size (μm^2) [2]	**0.082**	**0.065**	**0.048**	**0.036**	**0.028**	**0.019**	**0.015**
DRAM storage cell dielectric: equivalent physical oxide thickness, EOT (nm) [3]	**3.5**	**2.3**	**1.8**	**1.3**	**0.8**	**0.8**	**0.8**
Minimum DRAM retention time (ms) [4]	**64**	**64**	**64**	**64**	**64**	**64**	**64**
DRAM soft error rate (FITs) [5]	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**

Table 49b DRAM Technology Requirements—Long-term

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm) [1]	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm)	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
DRAM cell size (μm^2) [2]	**0.0122**	**0.0077**	**0.0061**	**0.0038**	**0.0025**	**0.0016**
DRAM storage cell dielectric: equivalent physical oxide thickness, EOT (nm) [3]	**0.70**	**0.58**	**0.53**	**0.42**	**0.37**	**0.25**
Minimum DRAM retention time (ms) [4]	**64**	**64**	**64**	**64**	**64**	**64**
DRAM soft error rate (FITs) [5]	**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 49a and 49b:

[1] From ORTC (Overall Roadmap Technology Characteristics) Table 1a and b. These DRAM half pitch numbers are the same as those in the 2002 ITRS due to no further speed up in the pace of DRAM half pitch scaling during 2002 and the early part of 2003.

[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 used by FEP are based on the ORTC Tables 1a and 1b. Since the FEP 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 the 90 nm node, real 6 for the 55 nm DRAM half pitch in 2008, and 5 (need $4F^2$ layout) for the 32 nm node. The reason for the introduction delay of "a" value of 6 is due to no clear solutions, as illustrated with yellow zone in this line. The "a" value of 5 has no known solution for 32 nm node and beyond, as illustrated with red zone.

[3] The 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 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 90 nm nodes, the dielectric material is based on Al_2O_3 or Ta_2O_5 with MIS structure, and hence the color is white. Beyond the 90 nm node, breakthroughs such as MIM structure and higher κ material are needed, so the color is yellow. Finally, for the 65 nm node and beyond, there are no known solutions with demonstrated credibility, and hence the color is red. The actual EOT required for each node also depends on the 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] 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.

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

DRAM POTENTIAL SOLUTIONS

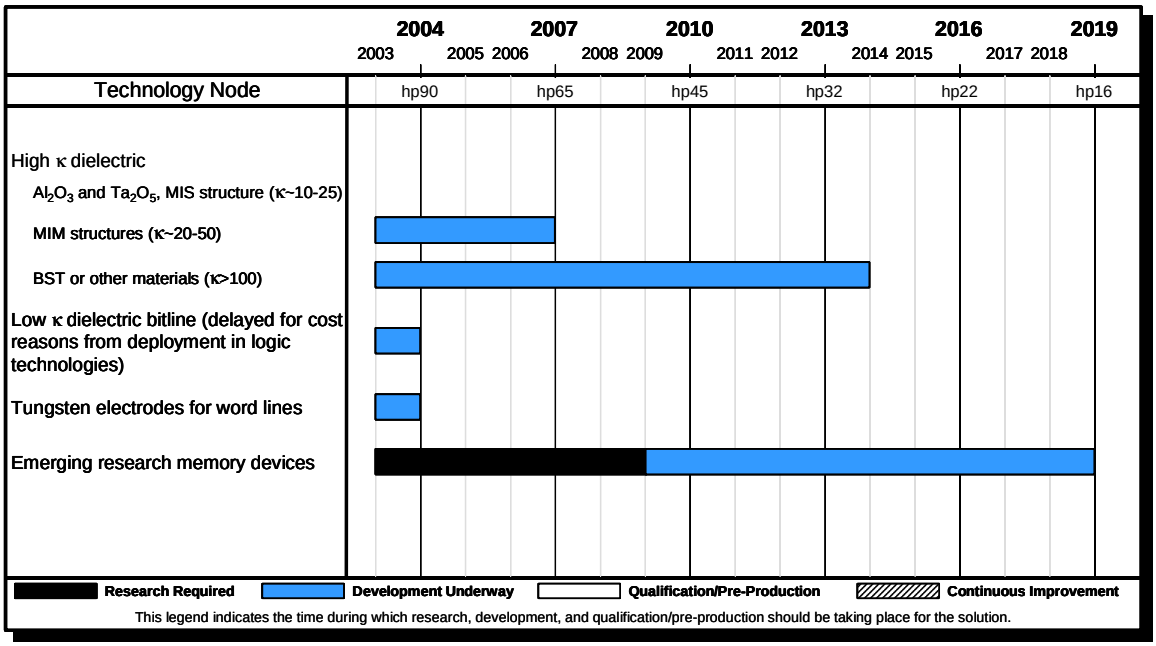


Figure 27 DRAM Potential Solutions

NON-VOLATILE MEMORY TECHNOLOGY REQUIREMENTS

Nonvolatile memory (NVM) technologies are in a general sense a combination of a CMOS structure and a memory element structure. The progression through sequential technology nodes 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 50a and 50b for Flash (NOR and NAND), FeRAM, SONOS, and MRAM technologies.

Historically the NVM devices in a given time period have not been referenced to the then-current CMOS node, but have lagged behind by one or more years. The tables identify both the current CMOS node feature size and the feature size actually used to form the NVM cells (i.e. the NVM technology node "F" in nanometers). Depending on the particular NVM technology, the time lag from the CMOS node to the NVM node is expected to reduce and eventually vanish.

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 (FeRAM) operates by switching and sensing the polarization state of a ferroelectric capacitor. Silicon-oxide-nitride-oxide-silicon (SONOS) memory transfers charge to and from traps in a nitride layer. Magnetic RAM (MRAM) 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.

Table 50a Non-Volatile Memory Technology Requirements—Near-term

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*120*	*107*	*95*	*85*	*76*	*67*	*60*
MPU/ASIC ½ Pitch (nm)	*107*	*90*	*80*	*70*	*65*	*57*	*50*
MPU Printed Gate Length (nm)	*65*	*53*	*45*	*40*	*35*	*32*	*28*
MPU Physical Gate Length (nm)	*45*	*37*	*32*	*28*	*25*	*22*	*20*
Flash technology node – F (nm) [1]	**107**	**90**	**80**	**70**	**65**	**55**	**50**
Flash NOR cell size – area factor a in multiples of F^2 [2]	**10–12**	**11–14**	**11–14**	**11–14**	**11–14**	**12–14**	**12–15**
Flash NAND cell size – area factor a in multiples of F^2 SLC/MLC [3]	**5.5**	**5.5**	**5.5**	**5.5**	**4.5**	**4.5**	**4.5**
Flash NOR typical cell size (μm^2) [4]	**0.135**	**0.101**	**0.08**	**0.061**	**0.053**	**0.039**	**0.034**
Flash NOR L_g-stack (physical – μm) [5]	**0.22–0.24**	**0.2–0.22**	**0.2–0.22**	**0.19–0.21**	**0.19–0.21**	**0.18–0.20**	**0.18–0.20**
Flash NOR highest W/E voltage (V) [6]	**8–10**	**7–9**	**7–9**	**7–9**	**7–9**	**7–9**	**7–9**
Flash NAND highest W/E voltage (V) [7]	**18–20**	**17–19**	**17–19**	**17–19**	**15–17**	**15–17**	**15–17**
Flash NOR I_{read} (μA) [8]	**34–42**	**31–39**	**29–37**	**28–36**	**27–35**	**26–34**	**25–33**
Flash coupling ratio [9]	**0.65–0.75**	**0.65–0.75**	**0.65–0.75**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**	**0.6–0.7**
Flash NOR tunnel oxide thickness (nm) [10]	**9–10**	**8.5–9.5**	**8.5–9.5**	**8.5–9.5**	**8–9**	**8–9**	**8–9**
Flash NAND tunnel oxide thickness (nm) [11]	**7–8**	**7–8**	**7–8**	**7–8**	**6–7**	**6–7**	**6–7**
Flash NOR interpoly dielectric thickness (nm) [12]	**11–13**	**10–12**	**9–11**	**9–11**	**8.5–10.5**	**8.5–10.5**	**8.5–10.5**
Flash NAND interpoly dielectric thickness (nm) [13]	**13–15**	**13–15**	**13–15**	**13–15**	**10–13**	**10–13**	**10–13**
Flash endurance (erase/write cycles) [14]	**1.00E+05**	**1.00E+05**	**1.00E+05**	**1.00E+05**	**1.00E+05**	**1.00E+05**	**1.00E+05**
Flash nonvolatile data retention (years) [15]	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**
Flash maximum number of bits per cell (MLC) [16]	**2**	**2**	**4**	**4**	**4**	**4**	**4**
FeRAM technology node – F (nm) [17]	**250**	**180**	**180**	**150**	**130**	**120**	**110**
FeRAM cell size – area factor a in multiples of F^2 [18]	**24**	**16**	**12**	**12**	**12**	**10**	**10**
FeRAM cell size (μm^2) [19]	**1.500**	**0.518**	**0.389**	**0.270**	**0.203**	**0.144**	**0.121**
FeRAM cell structure [20]	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**
FeRAM capacitor structure [21]	**stack**	**stack**	**stack**	**stack**	**3D**	**3D**	**3D**
FeRAM capacitor footprint (μm^2) [22]	**0.50**	**0.26**	**0.13**	**0.09**	**0.07**	**0.06**	**0.05**
FeRAM capacitor active area (μm^2) [23]	**0.50**	**0.26**	**0.13**	**0.09**	**0.09**	**0.09**	**0.08**
FeRAM cap active area/footprint ratio [24]	**1.00**	**1.00**	**1.00**	**1.00**	**1.29**	**1.50**	**1.60**
Ferro capacitor voltage (V) [25]	**2.5**	**1.8**	**1.5**	**1.5**	**1.2**	**1.2**	**1.2**
FeRAM minimum switching charge density ($\mu C/cm^2$) [26]	**11.2**	**17.2**	**34.5**	**34.5**	**40**	**40**	**40**
FeRAM endurance (read/write cycles) [27]	**1.00E+13**	**1.00E+14**	**1.00E+15**	**>1E16**	**>1E16**	**>1E16**	**>1E16**
FeRAM nonvolatile data retention (years) [28]	**10**	**10**	**10**	**10**	**10**	**10**	**10**

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

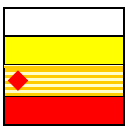


Table 50a Non-Volatile Memory Technology Requirements—Near-term (continued)

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*120*	*107*	*95*	*85*	*76*	*67*	*60*
MPU/ASIC ½ Pitch (nm)	*107*	*90*	*80*	*70*	*65*	*57*	*50*
MPU Printed Gate Length (nm)	*65*	*53*	*45*	*40*	*35*	*32*	*28*
MPU Physical Gate Length (nm)	*45*	*37*	*32*	*28*	*25*	*22*	*20*
SONOS/NROM technology node – F (nm) [29]	**130**	**115**	**100**	**90**	**70**	**65**	**55**
SONOS/NROM cell size – area factor a in multiples of F^2 [30]	**5**	**5**	**5.5**	**5.5**	**6**	**6**	**6**
SONOS/NROM typical cell size (mm^2) [31]	**0.085**	**0.066**	**0.055**	**0.045**	**0.029**	**0.025**	**0.018**
SONOS/NROM maximum number of bits per cell (MLC) [32]	**2**	**2**	**2**	**2**	**2**	**2**	**2**
SONOS/NROM area per bit (mm^2) [33]	**0.042**	**0.033**	**0.028**	**0.022**	**0.015**	**0.013**	**0.009**
SONOS L_g-stack (physical – µm) [34]	**0.18**	**0.18**	**0.17**	**0.17**	**0.16**	**0.16**	**0.16**
SONOS highest W/E voltage (V) [35]	**6.0–7.0**	**6.0–7.0**	**5.0–6.0**	**5.0–6.0**	**5.0–5.5**	**5.0–5.5**	**5.0–5.5**
SONOS/NROM I_{read} (µA) [36]	**35–45**	**33–43**	**31–41**	**29–39**	**27–37**	**25–35**	**25–35**
SONOS/NROM tunnel oxide thickness (nm) [37]	**5**	**5**	**4.5**	**4**	**3.5**	**3.5**	**3.5**
SONOS/NROM nitride dielectric thickness (nm) [38]	**6**	**5**	**5**	**4.5**	**4**	**4**	**4**
SONOS/NROM blocking (top) oxide thickness (nm) [39]	**5**	**5**	**4.5**	**4.5**	**4**	**4**	**4**
SONOS/NROM endurance (erase/write cycles) [40]	**1.00E+06**	**1.00E+06**	**1.00E+07**	**1.00E+07**	**1.00E+07**	**1.00E+07**	**1.00E+07**
SONOS/NROM nonvolatile data retention (years) [41]	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**
MRAM technology node F (nm) [42]	**180**	**130**	**90**	**90**	**65**	**55**	**50**
MRAM cell size area factor a in multiples of F^2 [43]	**24.7**	**23.7**	**24.7**	**22.2**	**22.2**	**22.2**	**22.2**
MRAM typical cell size (μm^2) [44]	**0.80**	**0.40**	**0.20**	**0.18**	**0.09**	**0.07**	**0.06**
MRAM switching field (Oe) [45]	**50**	**60**	**80**	**70**	**90**	**100**	**110**
MRAM write energy (pJ) [46]	**150**	**100**	**70**	**70**	**50**	**45**	**40**
MRAM active area per cell (μm^2) [47]	**0.28**	**0.20**	**0.11**	**0.10**	**0.05**	**0.04**	**0.03**
MRAM resistance-area product (Kohm-μm^2) [48]	**3.0**	**2.5**	**2.0**	**1.7**	**1.5**	**1.3**	**1.2**
MRAM magnetoresistance ratio(%) [49]	**45**	**45**	**50**	**50**	**60**	**60**	**60**
MRAM nonvolatile data retention (years) [50]	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**
MRAM write endurance (read/write cycles) [51]	**>1e15**	**>1e15**	**>1e15**	**>1e15**	**>1e15**	**>1e15**	**>1e15**
MRAM endurance – tunnel junction reliability (years at bias) [52]	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**

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

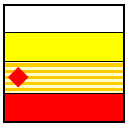


Table 50b Non-Volatile Memory Technology Requirements—Long-term

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*54*	*42*	*38*	*30*	*27*	*21*
MPU/ASIC ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU Printed Gate Length (nm)	*25*	*20*	*18*	*14*	*13*	*10*
MPU Physical Gate Length (nm)	*18*	*14*	*13*	*10*	*9*	*7*
Flash technology node – F (nm) [1]	**50**	**39**	**35**	**28**	**25**	**20**
Flash NOR cell size – area factor a in multiples of F^2 [2]	**12–15**	**12–15**	**13–16**	**14–17**	**14–17**	**15–18**
Flash NAND cell size – area factor a in multiples of F^2 SLC/MLC [3]	**4.5/2.3**	**4.5/2.3**	**4.5/2.3**	**4.5/2.3**	**4.5/2.3**	**4.5/2.3**
Flash NOR typical cell size (μm^2) [4]	**0.034**	**0.021**	**0.018**	**0.012**	**0.010**	**0.007**
Flash NOR L_g-stack (physical – μm) [5]	**0.17–0.19**	**0.15–0.17**	**0.14–0.16**	**0.13–0.15**	**0.12–0.14**	**0.11–0.13**
Flash NOR highest W/E voltage (V) [6]	**7–9**	**7–9**	**7–9**	**7–9**	**7–9**	**7–9**
Flash NAND highest W/E voltage (V) [7]	**15–17**	**15–17**	**15–17**	**15–17**	**15–17**	**15–17**
Flash NOR I_{read} (μA) [8]	**27–33**	**26–32**	**25–31**	**23–29**	**22–28**	**20–26**
Flash Coupling Ratio [9]	**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 (nm) [10]	**8–9**	**8–9**	**8**	**8**	**8**	**8**
Flash NAND tunnel oxide thickness (nm) [11]	**6–7**	**6–7**	**6–7**	**6–7**	**6–7**	**6–7**
Flash NOR interpoly dielectric thickness (nm) [12]	**8–10**	**7–9**	**6–8**	**5–7**	**4–6**	**3–5**
Flash NAND interpoly dielectric thickness (nm) [13]	**10–13**	**10–13**	**9–10**	**9–10**	**9–10**	**9–10**
Flash endurance (erase/write cycles) [14]	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+06**	**1.00E+07**	**1.00E+07**
Flash nonvolatile data retention (years) [15]	**10–20**	**10–20**	**20**	**20**	**20**	**20**
Flash maximum number of bits per cell (MLC) [16]	**8**	**8**	**8**	**8**	**8**	**8**
FeRAM technology node – F (nm) [17]	**100**	**80**	**70**	**57**	**50**	**45**
FeRAM cell size – area factor a in multiples of F^2 [18]	**8**	**8**	**8**	**8**	**8**	**8**
FeRAM cell size (μm^2) [19]	**0.080**	**0.051**	**0.039**	**0.026**	**0.020**	**0.016**
FeRAM cell structure [20]	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**	**1T1C**
FeRAM capacitor structure [21]	**3D**	**3D**	**3D**	**3D**	**3D**	**3D**
FeRAM capacitor footprint (μm^2) [22]	**0.030**	**0.019**	**0.015**	**0.010**	**0.0075**	**0.0061**
FeRAM capacitor active area (μm^2) [23]	**0.076**	**0.065**	**0.06**	**0.052**	**0.048**	**0.045**
FeRAM cap active area/footprint ratio [24]	**2.53**	**3.42**	**4.00**	**5.20**	**6.40**	**7.38**
Ferro capacitor voltage (V) [25]	**1**	**1**	**0.7**	**0.7**	**0.7**	**0.7**
FeRAM minimum switching charge density ($\mu C/cm^2$) [26]	**40**	**40**	**40**	**40**	**40**	**40**
FeRAM endurance (read/write cycles) [27]	**>1E16**	**>1E16**	**>1E16**	**>1E16**	**>1E16**	**>1E16**
FeRAM nonvolatile data retention (years) [28]	**10**	**10**	**10**	**10**	**10**	**10**

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

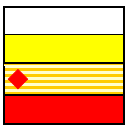
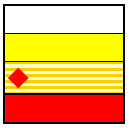


Table 50b Non-Volatile Memory Technology Requirements—Long-term (continued)

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	*54*	*42*	*38*	*30*	*27*	*21*
MPU/ASIC ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU Printed Gate Length (nm)	*25*	*20*	*18*	*14*	*13*	*10*
MPU Physical Gate Length (nm)	*18*	*14*	*13*	*10*	*9*	*7*
SONOS/NROM technology node – F (nm) [29]	**50**	**40**	**35**	**28**	**25**	**20**
SONOS/NROM cell size – area factor a in multiples of F^2 [30]	**6**	**6**	**6.5**	**6.5**	**7**	**7**
SONOS/NROM typical cell size (mm^2) [31]	**0.015**	**0.010**	**0.008**	**0.005**	**0.004**	**0.003**
SONOS/NROM maximum number of bits per cell (MLC) [32]	**4**	**4**	**4**	**4**	**4**	**4**
SONOS/NROM area per bit (mm^2) [33]	**0.0038**	**0.0024**	**0.0020**	**0.0013**	**0.0011**	**0.0007**
SONOS L_g-stack (physical – µm) [34]	**0.16**	**0.16**	**0.15**	**0.15**	**0.14**	**0.14**
SONOS highest W/E voltage (V) [35]	**5.0–5.5**	**5.0–5.5**	**5.0–5.5**	**5.0–5.5**	**4.5.–5.0**	**4.0–4.5**
SONOS/NROM I_{read} (µA) [36]	**25–35**	**24–34**	**23–33**	**22–32**	**21–31**	**20–30**
SONOS/NROM tunnel oxide thickness (nm) [37]	**3**	**3**	**2.5**	**2.5**	**2**	**2**
SONOS/NROM nitride dielectric thickness (nm) [38]	**4**	**4**	**4**	**4**	**3.5**	**3.5**
SONOS/NROM blocking (top) oxide thickness (nm) [39]	**4**	**4**	**4**	**4**	**4**	**4**
SONOS/NROM endurance (erase/write cycles) [40]	**1.00E+08**	**1.00E+08**	**1.00E+08**	**1.00E+08**	**1.00E+09**	**1.00E+09**
SONOS/NROM nonvolatile data retention (years) [41]	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**	**10–20**
MRAM technology node – F (nm) [42]	**45**	**35**	**32**	**25**	**22**	**18**
MRAM cell size – area factor a in multiples of F^2 [43]	**22.2**	**22.2**	**22.2**	**22.2**	**22.2**	**22.2**
MRAM typical cell size ($µm^2$) [44]	**0.04**	**0.03**	**0.02**	**0.01**	**0.01**	**0.01**
MRAM switching field (Oe) [45]	**120**	**120**	**120**	**120**	**120**	**120**
MRAM write energy (pJ) [46]	**35**	**30**	**25**	**23**	**20**	**18**
MRAM active area per cell ($µm^2$) [47]	**0.03**	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**
MRAM resistance-area product (Kohm-$µm^2$) [48]	**1.1**	**0.95**	**0.8**	**0.7**	**0.6**	**0.55**
MRAM magnetoresistance ratio(%) [49]	**60**	**65**	**70**	**70**	**70**	**70**
MRAM nonvolatile data retention (years) [50]	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**
MRAM write endurance (read/write cycles) [51]	**>1e15**	**>1e15**	**>1e15**	**>1e15**	**>1e15**	**>1e15**
MRAM endurance – tunnel junction reliability (years at bias) [52]	**>10**	**>10**	**>10**	**>10**	**>10**	**>10**

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



Notes for Tables 50a and 50b:

[1] In the past Flash devices tended to lag behind the current CMOS technology node, 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 node F^2. Note the lack of long term scaling.

[3] The area factor "a" = cell area/F^2, so this entry presents the Flash NAND cell area in multiples of F^2 the implementation technology node. 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.)

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

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

[6, 7] This is the highest voltage relative to ground seen in the cell array. It is not usually an external supply.

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

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

[10, 11] 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.

[12, 13] 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.

[14] 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.

[15] 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.

[16] Cell read out distinguishes between four levels of charge storage to provide two storage bits. Progression to 16 and 256 levels is anticipated. (MLC multilevel cell).

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

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

[19] FeRAM cell size is presented in terms of micrometers squared. It is the product $a \times F^2$.

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

[21] 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.

[22] 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 2003–2004 $8F^2$, for 2005–2009 $4F^2$, and for 2010–2018 F^2 are assumed.

[23] This is the actual effective area of the capacitor. It can be larger than the footprint because of the utilization of area in the third dimension.

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

[25] This is the operating voltage (V_{op}) applied to the capacitor. Low voltage operation is a difficult key design issue.

[26] The minimum switching charge density in $\mu C/cm^2$ 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}.

[27] 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.

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

[29] SONOS/NROM devices have recently been introduced into the commercial market and will tend to lag the current CMOS technology node by one year; however, similar to Flash floating-gate technology, this gap will close rapidly due to low programming voltages combined with CMOS compatibility (i.e. single-level polysilicon). This entry provides the F value for designs in the indicated time period.

[30] The area factor "a" = cell area/F^2. This entry depicts the expected SONOS/NROM NOR cell area in multiples of the implementation technology node F^2. The reduced ONO gate stack thickness and low programming voltages should permit long range scaling of this technology.

[31] The expected "typical" SONOS/NROM NOR cell size is presented in terms of micrometers squared.

[32] 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, thereby, increasing the bit storage per cell.

[33] The expected "typical" SONOS/NROM NOR area per bit is presented in terms of micrometers squared.

[34] 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.

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

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

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

[38] 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.

[39] 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.

[40] E/W endurance (erase/write cycles) requirements vary with the specifics of an application, but 1E5cycles 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.

[41] 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.

[42] MRAM devices are expected to lag the CMOS current technology node up until the 65 nm node in 2007. This entry provides the F value for designs in the indicated time period.

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

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

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

*[46] 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).*

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

*[48] 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.*

[49] 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.

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

[51] 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.

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

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 requires 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 figure 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 nodes.

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. As shown in the potential solutions figure, material selection and development must proceed rapidly if the needs of the 65 nm node are to be met.

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 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 control of the spread of charge in the nitride will become an issue. The potential solutions table points out that the stack refinement effort now underway must serve to meet the needs of the 90 nm node.

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 and development 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 current density electromigration limits are being approached. The potential solutions table provides a snapshot of the current materials research activity that must converge to a manufacturable solution for the 65 nm node.

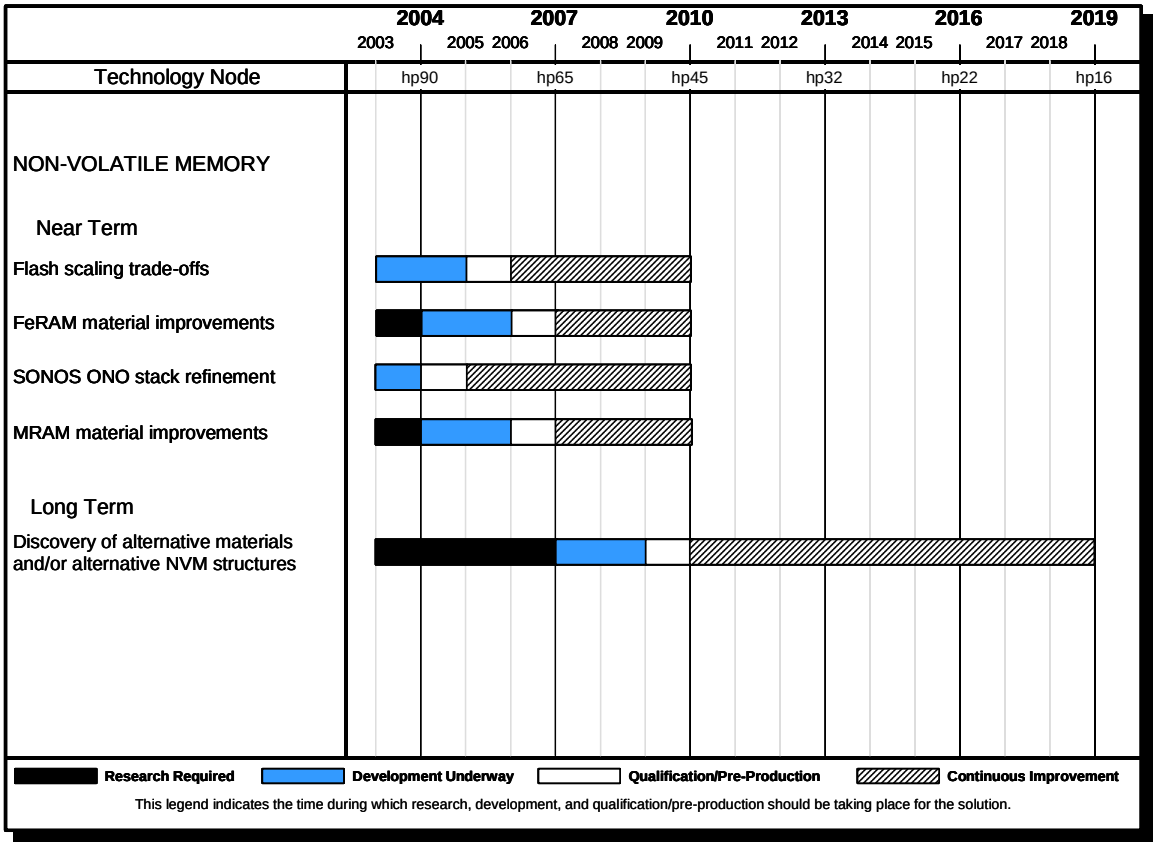


Figure 28 Non-volatile Memory Potential Solutions

RELIABILITY

RELIABILITY DIFFICULT CHALLENGES

There are widespread reliability challenges associated with evolutionary changes in scaling current materials and devices. There are even more profound reliability challenges associated with revolutionary changes associated with new materials and new devices. The semiconductor industry's experience with implementing of Cu metallization and low-κ interlevel dielectrics demonstrates the significant reliability challenges that can be encountered with technology changes. Refer to Table 51.

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 International Sematech. The report identifies reliability concerns and research needs in thirteen 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.

Table 51 indicates the RTAB consensus on the most critical five near-term reliability challenges. It expands on the PIDS overall Difficult Challenges, "*Assuring the reliability and implementing into manufacturing of multiple material, process, and structural changes in a relatively short period of time,*" described at the beginning of this chapter. The replacement of silicon dioxide with a higher dielectric constant (κ) insulator 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 (NBTI). To put this change into perspective, one needs 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.

As mentioned above, the move to Cu/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 sensitive to interface issues.

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

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.

Table 51 Reliability Difficult Challenges

Difficult Challenges ≥45 nm/ Through 2010	*Summary of Issues*
High-κ Gate Dielectrics	Dielectric breakdown characteristics (hard and soft breakdown) Influence of charge trapping and NBTI on threshold voltage stability Stability and number of fixed charges
Metal Gate	Impact of metal-ion drift and/or diffusion on gate dielectric reliability Work function control and stability Metal susceptibility to oxidation Thermo-mechanical issues due to large thermal expansion mismatch Impact of implantation
Copper/Low-κ Interconnects	Stress migration of Cu vias and lines Cu via and line electromigration performance Thermal-mechanical stability of the interfaces between metals, barriers and interlevel dielectrics and resulting line-to-line leakage Time Dependent Dielectric Breakdown (TDDB) of the Cu/low-κ system Reliability impact of lower thermal conductivity of low-κ dielectric Reliability issues due to the porous nature of the low-κ dielectrics and moisture Reliability impact of the lower mechanical strength in the Cu/low-κ system, including the impact of packaging
Packaging	Ability of bumps to withstand thermal and mechanical stresses while providing sufficient current carrying capability Solder joints fracture at 1st and 2nd level interconnects Electromigration in package traces, vias and bumps Impact of increasing Coefficient of Thermal Expansion (CTE) mismatch between low-κ, silicon and organic packages
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

RELIABILITY TECHNOLOGY 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 specification 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 physics of all relevant failure modes and the existence of powerful reliability engineering capabilities for design-for reliability, and defect screening. There are some significant gaps in these capabilities today. Furthermore, these gaps will become even larger with the introduction of new materials and new device structures. Inadequate reliability tools lead to unnecessary performance penalties and/or unnecessary risks. Finally, as trade-offs between reliability and performance become more difficult, excess reliability margins need to be eliminated.

Reliability qualification always involves some risk. There is a risk of qualifying a technology that does not, in fact, meet reliability requirements or 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 represents the reliability risk associated with incomplete knowledge and tools for new materials and devices. The progression from yellow to striped to red indicates a growing reliability risk. The requirements first turn to yellow, "Manufacturing Solutions are Known," in 2004 indicating a relative smaller risk associated with the introduction into manufacturing of strained silicon substrates.

The requirements first turn to striped, "Interim Solutions are Known," in 2007 due to the increased reliability risk resulting from introduction of high-κ gate insulators. The reliability knowledge base and tools will not be as well developed as for oxides. As an interim solution, the techniques that are currently used to qualify oxides can be used with high-κ insulators. 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 is not considered a showstopper (i.e., a red box). The introduction of high κ will be followed by other changes (such as metal gate, fully depleted SOI) that will also be in the striped category.

Red is now reserved for the reliability requirements nearer to the end of the Roadmap. Red is used where a reliability solution is not known because the technology choices have not been specified (for example, "what specifically will be used for post Cu interconnect?"). When the details of these technology changes are understood, a better assessment of the reliability risk will be possible.

Table 52a Reliability Technology Requirements—Near-term

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*	
Technology Node		*hp90*			*hp65*			
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*	
*Early failures (ppm) (First 4000 operating hours)** [1]*	**50–2000**	**50–2000**	**50–2000**	**50–2000**	**◆50–2000**	**◆50–2000**	**◆50–2000**	*Customer needs; new materials (High κ in 2004)*
Long term reliability (FITS = failures in 1E9 hours) [2]	**10–100**	**10–100**	**10–100**	**10–100**	**◆10–100**	**◆10–100**	**◆10–100**	*Customer needs; new materials (High κ in 2004)*
Soft error rate (FITs)	**1000**	**1000**	**1000**	**1000**	**◆1000**	**◆1000**	**◆1000**	*Scaling*
Relative failure rate per transistor (normalized to 130 nm) [3]	**1**	**0.8**	**0.63**	**0.5**	**◆ 0.4**	**TBD**	**TBD**	*Number of transistors*
Relative failure rate per m of interconnect (normalized to 130 nm node) [4]	**1**	**0.84**	**0.64**	**0.58**	**◆ 0.52**	**◆ 0.41**	**◆ 0.37**	*Customer needs; J11 length of interconnect*

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

Table 52b Reliability Technology Requirements—Long-term

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
*Early failures (ppm) (First 4000 operating hours)**[1]*	◆ 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
Soft error rate (FITs)	◆ 1000	1000	1000	1000	1000	1000
Relative failure rate per transistor (normalized to 130 nm) [3]	◆ 0.2		0.1		0.04	
Relative failure rate per m of interconnect (normalized to 130 nm) [4]	◆ 0.36	0.26	0.25	0.16	0.17	0.11

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

Notes for Tables 52a and 52b:

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 mode 0. 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 increase [from ORTC], the relative failure rate per transistor must decrease.

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

RELIABILITY POTENTIAL SOLUTIONS

The most effective way to meet requirements is to have complete built-in-reliability and design-for-reliability solutions available at the start of the development of each new technology node. 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 mode 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 node 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 29 include the major significant technical changes through the lifetime of the Roadmap. (There are a wide variety of changes not listed in the figure that also could impact reliability.) Because these serious reliability issues are involved, 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 node (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 29.

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 need 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.

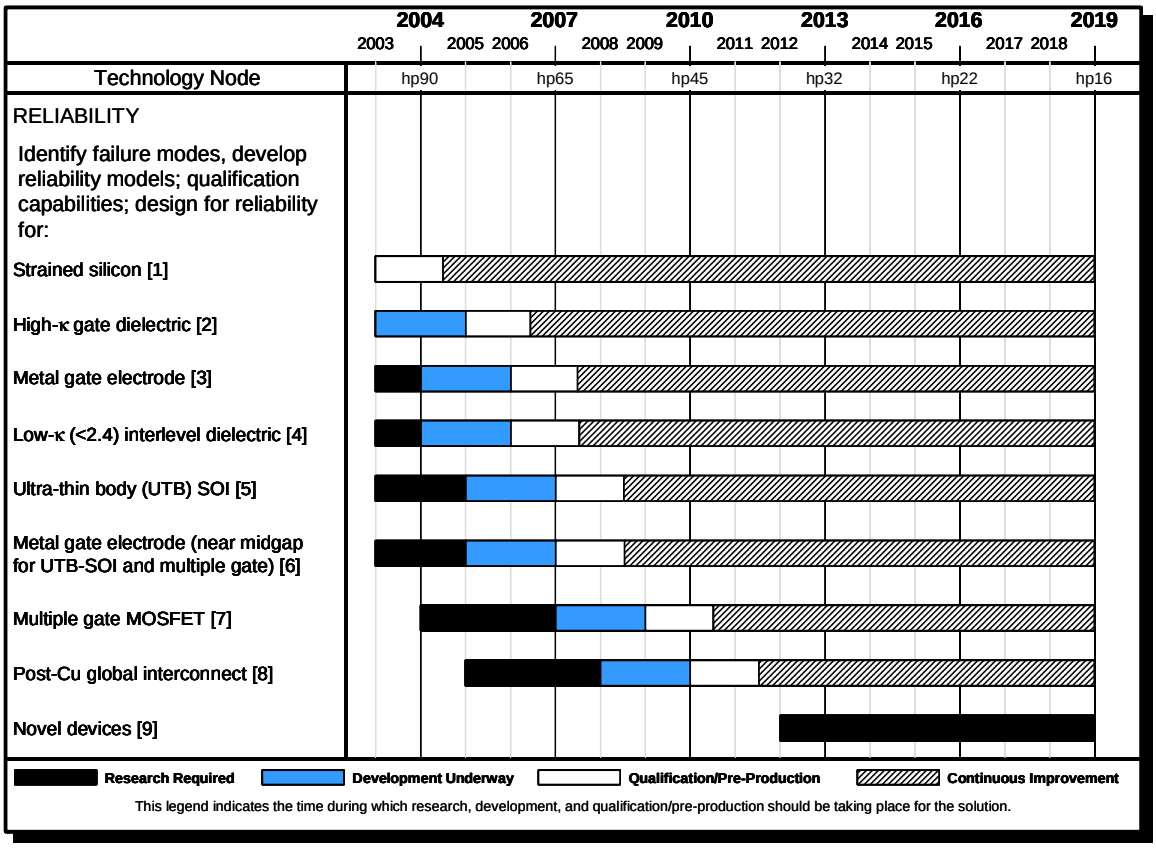


Figure 29 Reliability Potential Solutions

Notes for Figure 29

[1] Driven by requirement for strained silicon in 2004 for high-performance (HP) logic (see PIDS Logic Potential Solutions Figure).

[2] Driven by requirement for high-κ gate insulator in 2006 for Low Operating Power (LOP) and Low Standby Power (LSTP) logic (see PIDS Logic Potential Solutions Figure).

[3] Driven by requirement for metal gate electrodes in 2007 for HP logic (see PIDS Logic Potential Solutions Figure).

[4] Driven by requirement for interlevel dielectric constant, κ, < 2.4 in 2007 (see MPU Interconnect Technology Requirements Table).

[5] Driven by requirement for Ultra Thin Body SOI in 2008 for HP (see PIDS Logic Potential Solutions Figure).

[6] Driven by requirement for near mid-gap work function metal gate for Ultra Thin Body SOI in 2008 for HP logic (see PIDS Logic Potential Solutions Figure).

[7] Driven by requirement for multiple-gate MOSFET in 2010 for high-performance logic (see PIDS Logic Potential Solutions Figure).

[8] Driven by requirement to implement optical interconnect in 2009–2011 (see Global Interconnect Potential Solutions Figure).

[9] Timing needs to be refined in future roadmaps. Put here to indicate that research on reliability issues of novel devices will be needed in the future. (See Emerging Research Devices Section.)

CROSS-CUT ITWG ISSUES

MODELING AND SIMULATION

The key innovations requested by the *PIDS* section include enhanced mobility (leading to strained Si), high-κ dielectrics, metal gate, advanced MOSFETs (such as fully depleted SOI), and enhanced saturation current which requests ballistic transport. Other more long-term issues include atomic-level fluctuations, statistical process variations, new interconnect schemes, mixed-signal device technology, and generally advanced non-CMOS devices and architectures which will drive major changes in process, materials, physics, design, etc. 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."

Concerning *Emerging Research Devices*, near-term requirements on Modeling and Simulation result from non-classical CMOS, which have in this issue of the ITRS also been discussed in the Logic Technology Requirements and Logic Potential Solutions sections of the PIDS chapter. These requirements have been integrated into the Modeling and Simulation chapter via its requirements tables and its subchapter on Device Modeling.

The requirements on simulation from the AMS and RF technologies for Wireless Communications Working Group within PIDS mainly consist of the extension to include not only silicon-based substrates but also III-V compounds, certain device architectures beyond MOSFET, and the capability to simulate frequencies up to 100 GHz. It increases the need for coupled device/circuit/system simulation of system-on-chip (SoC), accurate, fast and predictive analog/RF compact models, and computationally efficient physical models for carrier transport. These aspects are to some extent already addressed in the "Device Modeling", the "Interconnect and Integrated Passives Modeling" and the "Circuit Element Modeling" of the *Modeling and Simulation* chapter of this ITRS, and will in future be further worked out and implemented in that chapter.