INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

2001 EDITION

PROCESS INTEGRATION, DEVICES, AND STRUCTURES AND EMERGING RESEARCH DEVICES

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PROCESS INTEGRATION, DEVICES, & STRUCTURES

SCOPE

The *Process Integration, Devices, & 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 active device electrical parameters, passive electrical parameters, and reliability criteria. Nominal targets as well as statistical tolerances are discussed. The scope has been broken into four subcategories described below: logic and memory, mixed-signal devices, reliability, and emerging research devices.

One key theme of the *2001 ITRS* is a significant acceleration in the scaling of MOS transistors compared to that in the 1999 edition and the 2000 update of the ITRS. For example, in the year 2005 the physical gate length for transistors in high-performance logic chips is projected to be 32 nm in this edition of the ITRS (see *ORTC chapter, Table 1a and b*), while it was projected to be 65 nm in the 1999 ITRS and 60 nm in the 2000 ITRS. This acceleration reflects the achievements of the industry in implementing very aggressively scaled transistors into production IC processes over the last two years, as well as the achievements of the R&D community as reflected in the literature over the last two years. The acceleration is driven by the need to maintain historical trends of improved transistor performance and power dissipation, and it strongly impacts the overall scaling of the transistor. Another 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, and device performance. This is the fundamental issue associated with the System-on-a-Chip (SoC) application, which is a major trend for future ICs. SoC issues are a larger crosscut problem that will be dealt with as a multidiscipline activity in a separate chapter, but they will also be discussed in this chapter, particularly in the mixed-signal section. 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. In this chapter, the focus will be on those transistors that drive the technology requirements, for example the highest performance transistors in high-performance logic chips.

MEMORY AND LOGIC

Digital circuitry, both memory and logic, forms the major portion of semiconductor device production. High-performance and low-power devices lead the logic category, while memory encompasses DRAM, SRAM, and nonvolatile memory (NVM). In the logic category, transistors for both high-performance and low-power chips are discussed. Process integration optimizes the overall architecture of the full process, including the silicon active devices and the on-chip interconnect hierarchies for power, clock, and signal distribution. The architectural tradeoffs between the devices and interconnect structures are driven by performance, power, density, and reliability requirements. Devices and structures refer to the active transistors, interconnect, and other structures required for logic and memory cell design. The discussion of memory cells, DRAM, SRAM and NVM, is focused on commodity applications.

MIXED-SIGNAL DEVICES

Mixed-signal ICs contain analog and/or radio frequency (RF) circuitry as well as the digital circuitry. The commodity driver applications for mixed-signal ICs are projected to remain in personal computing and communications. Analog refers to "pure" analog circuits such as operational amplifiers, but mixed-signal chips often utilize functions such as analog-to-digital and digital-to-analog converters and digital signal processors. RF refers to circuits operating above 800 MHz, such as for wireless communications and "radio on a chip." Certain analog IC technologies such as high voltage and power ICs are not specifically included, while others, such as microelectromechanical systems (MEMS), are included in the later, long-term years. Such devices and analog and RF transistors must reuse and leverage the mainstream, digital CMOS technology to remain low cost and meet the demands for high-performance and reliability.

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.

EMERGING RESEARCH DEVICES

An *Emerging Research Devices section* is added to the PIDS chapter in the 2000 ITRS Update, and the section is updated and considerably expanded in this chapter. The addition of this section is driven by the increasing difficulty of meeting all the desired requirements for the aggressively scaled technologies projected for later technology nodes. This difficulty only becomes greater for the later nodes, as indicated by the increasing incidence of red cells (no known solutions) in the various technology requirements tables as the years increase. As one approach to dealing with these difficulties, the research and development community has been actively exploring alternative devices to the current classical planar bulk CMOS transistors and current memory devices. (The other approach, to find solutions to the scaling problems with the current device structures, is also being actively pursued.) The Emerging Research Devices section is aimed at giving an overall understanding of the main emerging devices, including their operating principles, their major advantages and disadvantages, and the general time frame in which they might be deployed. The main categories are non-classical MOSFETs, emerging research memory devices, emerging research logic (non-MOSFET) devices, and emerging research novel architectures. Non-classical MOSFETs and some emerging memory devices have the shortest time frame for possible insertion into manufacture, ranging from a few years to the end of this decade and beyond, while non-MOSFET logic devices and emerging research architectures are longer-range, with expected availability ranging from the end of this decade to beyond the end of the Roadmap in 2016 for the more exotic technologies. It is important to understand that the emerging research devices are speculative, with no certainty of being implemented into manufacture.

DIFFICULT CHALLENGES

Table 34a Process Integration Difficult Challenges—Near-term

DIFFICULT CHALLENGES ≥65NM / THROUGH 2007	*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. Implementation of metal gate electrode by about 2007. 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. Very slow scaling of V_{dd} will make overall device scaling difficult.
3. Implementation into manufacturing of non-classical MOSFET devices (for example, double-gate SOI).	It is likely that these transistors will be necessary eventually to control short-channel and other effects in highly scaled devices. See *Emerging Research Devices section, Non-classical CMOS,* for more detail.
4. Ensuring reliability of new materials and structures in a timely manner.	Accelerated reliability ensurance of high κ material for gate stack will be needed for early insertion into manufacturing. Ensuring reliability of new gate electrode materials will be a challenge. Ensuring reliability of new, non-classical CMOS structures will be a challenge. Ensuring reliability of very thin oxy-nitrides with very high leakage current will be critical for high-performance applications. Difficulty of screening with high leakage currents

Table 34a Process Integration Difficult Challenges—Near-term (continued)

DIFFICULT CHALLENGES ≥ 65 nm / THROUGH 2007	SUMMARY OF ISSUES
5. Constructing DRAM, SRAM, and high density nonvolatile memory (NVM) for scaled technologies	DRAM main issues: adequate storage capacitance for devices with reduced feature size; access device design; holding the overall leakage to acceptably low levels; and deploying low sheet resistance materials for bit and word lines to ensure desired speed for scaled DRAMs. Also, the availability of manufacturing worthy 193 nm lithography and integrated DRAM etch capability for 100 nm half pitches in 2003. SRAM: difficult lithography and etch as well as process integration issues. NVM: very difficult scaling issues with tunnel and interpoly dielectrics.
6. High-performance mixed-signal solutions for scaled technologies.	Passive element scaling: embedded inductor densities and Q factor values. Signal isolation. Optimizing RF CMOS devices with scaled technologies: gate leakage is a particularly sensitive issue. Transition to reduced analog supply voltages. Difficulty and cost of integrating analog/RF and high-performance digital functions on a chip.

Table 34b Process Integration Difficult Challenges—Long-term

DIFFICULT CHALLENGES < 65 nm, BEYOND 2007	SUMMARY OF ISSUES
7. Fundamental improvements in MOSFET device effective transconductance needed to maintain device performance scaling trend.	With sharp reductions in V_{dd} and 17% annual increase in intrinsic transistor speed, basic MOSFET device performance will be inadequate to meet circuit speed requirements.
8. Dealing with atomic-level fluctuations and statistical process variations in sub-30 nm MOSFETs.	Fundamental problems of atomic-level statistical fluctuations are not completely understood.
9. New interconnect schemes	Eventually, copper/low κ performances will be inadequate. Solutions (optical, microwave/RF, etc,) are currently unclear.
10 Toward the end of the Roadmap or beyond, implementation of advanced non-CMOS devices and architectures, including memory.	Will drive major changes in process, materials, physics, design, etc. Non-CMOS devices may coexist with CMOS: integration of the two will 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 below 1.0 nm EOT (Equivalent Oxide Thickness). Furthermore, with very rapid transistor scaling, the maximum allowable parasitic series source/drain resistance will decrease, and hence it will become increasingly difficult to meet this requirement. Finally, in the face of increased chip complexity and high leakage with the succeeding years, circuit design and architectural innovation will be needed to design chips with 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 will be required by about 2005 in order to meet stringent leakage specifications. Another major issue is that, since V_{dd} hardly scales, the lateral electric field will become uncomfortably large, resulting in difficulties in controlling short channel effects and possibly reliability problems.

[3] Implementation into manufacturing of non-classical MOSFET devices (for example, double-gate SOI)—With continuing sharp scaling of the transistors, control of short-channel and other effects will become increasingly difficult using classical CMOS transistors. Non-classical CMOS transistors show promise of enabling superior electrical characteristics for very highly scaled transistors, and will likely be implemented into production at some point, perhaps by around 2007. (Refer to the *Emerging Research Devices section, Non-classical CMOS*.)

[4] Ensuring reliability of new materials and structures in a timely manner—Numerous new materials and structures, including high κ gate dielectric, metal gate electrode, and non-classical CMOS transistors, are expected to be

implemented into production in the next six years. For each such new option, reliability issues must be explored, understood, and satisfactorily dealt with before initiation of production.

[5] Constructing high density nonvolatile memory (NVM) , DRAM, and SRAM memories 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. Finally, because of the limited selectivity and etch resistance of 193 nm photoresists, special difficulty is expected in integrating 193 nm lithography and the high aspect ratio, deep etches required for DRAMs. For NVM, the key issue is the difficulty in scaling the tunneling oxide, since this oxide must be thick enough to ensure acceptable retention time, but thin enough to allow ease of program/erase.

[6] High-performance mixed-signal solutions for scaled technologies—Signal isolation, especially between the digital and analog/RF sections of the chip, is a particular issue for scaled technologies. Also, maintaining high Q, well-matched, linear passive devices will be a challenge. Finally, the sheer 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 are expected to increase with scaling.

[7] *Fundamental improvements in MOSFET device effective transconductance needed to maintain device performance scaling trend.* For the long-term years, owing especially to the rapid scaling of V_{dd}, increased device effective mobility will be needed to enable historic 17% per year transistor performance improvement with scaling.

[8] Dealing with atomic-level fluctuations and statistical process variations in sub-30 nm MOSFETs—For very short channel 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. Solutions such as the use of fully depleted devices and non-lithographically defined MOSFETs may help to ameliorate these effects.

[9] New 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 architectural and/or materials solutions.

[10] Toward the end of the Roadmap or beyond, implementation of advanced non-CMOS devices and architectures, including memory—Eventually, toward the end of the Roadmap or beyond, scaling of MOSFETs will become ineffective and/or very costly, and advanced non-CMOS solutions will need to be implemented. (Refer to the *Emerging Research Devices sections*.)

TECHNOLOGY REQUIREMENTS

MEMORY AND LOGIC TECHNOLOGY REQUIREMENTS

LOGIC—HIGH PERFORMANCE AND LOW POWER TECHNOLOGY REQUIREMENTS

The technology requirements tables reflect the MOSFET transistor requirements of both high- performance and low-power digital ICs. High performance refers to chips of high complexity, high performance, and high power dissipation, such as microprocessor unit (MPU) chips for desktop PCs. Low power 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) chips. LOP chips are typically for relatively high-performance mobile applications, such as notebook computers, where the battery is likely to be high capacity, while LSTP chips are typically for lower performance consumer type applications, such as cellular telephones, with lower battery capacity. The transistors for high-performance ICs are highly scaled, and have both the highest performance and the highest leakage current of all. The transistors for LOP chips are less highly scaled and have somewhat lower performance and much lower leakage current, while the transistors for LSTP chips are scaled the same as those for LOP chips, but the performance and the leakage current are lower still compared to transistors for LOP chips.

For high-performance transistors, the physical gate length is much more rapidly scaled in this roadmap compared to the scaling in the *1999 ITRS* (for example, the projected physical gate length was 65 nm in 2005 in the 1999 ITRS, while it is 32 nm in 2005 in the current roadmap). The accelerated scaling in this roadmap reflects the actual progress of the industry over the last two years, both in production chips and in R&D results reported in the literature. The rapid scaling in the industry is a response to the need to maximize the transistor saturation drive current to maximally improve performance. For low-power transistors, the gate length lags behind the high-performance transistor gate length by two years, reflecting historical trends and the need for very low leakage current in mobile applications.

For the logic technology requirements tables, an approach is used in which simplified models embedded in a spreadsheet capture the essentials of the impact of such key input parameters as 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 device delay, $\tau = CV/I$, where C is the total gate capacitance (including parasitic gate overlap and fringing capacitance) per micron transistor width, $V = V_{dd}$, and I is the saturation drive current per micron transistor width. 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 the tables, a target is set for one or more 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 or targets are met, and the final value of the parameters is used for the projected parameter values in the tables. (See *Supplemental files* for the detailed spreadsheets used to generate the Logic technology requirements tables.)

For the high-performance transistor table, the driver is the device 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 table. The NMOSFET saturation drive current, I_{dd}, stays constant at 900 µA/µm until the 65 nm technology node in 2007, and then it increases somewhat in the later years. The subthreshold source/drain leakage current, $I_{sd,leak}$, increases quite rapidly with succeeding years owing to the reduction of the threshold voltage, V_t, with succeeding years. (At > 1µA/µm for the long-term years, the subthreshold leakage current is especially high then.) Since the saturation drive current 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_{dd} up to the specified value. But $(1/I_{sd,leak})$ is exponentially dependent on V_t, and hence $I_{sd,leak}$ increases rapidly as V_t decreases. This increase in $I_{sd,leak}$ causes the static power dissipation per device to increase somewhat with device scaling, despite the reduction in V_{dd} and device dimensions with scaling (see last row of the table), and it has important consequences for the chip power dissipation (to be discussed below). Finally, a reasonable restriction is imposed on the gate leakage current, that it must remain less than $I_{sd,leak}$. Because of the large increases in $I_{sd,leak}$ (and hence in the allowable gate leakage current) with device scaling, *it is projected (see the Front End Processes chapter) that heavily nitrided oxy-nitrides will be sufficient to meet the gate leakage requirements until the end of the roadmap.*

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 presumably poorer device performance (i.e., longer device intrinsic delay, τ) than the high-performance devices. The high-performance device is used just in critical paths or in circuits which 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. Other techniques to curtail static power dissipation include well-biasing, using electrically or dynamically adjustable V_t devices, and using circuit/architectural techniques, including pass gates to cut off access to power/ground rails or other techniques to power down circuit blocks. 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 driver is the leakage current, with a targeted maximum of 100 pA/µm for the LOP transistor and 1 pA/µm for the LSTP transistor. (In both cases, the leakage target increases in the later years.) These targets are set by battery life requirements, and apply to both the gate leakage and the source/drain subthreshold leakage currents. All the other parameter values in the tables are chosen iteratively to meet these targets, as explained above. The resultant average

improvement in the device performance metric, $1/\tau$, is about 14% per year for both LOP and LSTP. One key issue is the relatively slow scaling of V_{dd} for the low-power transistors. This is a result of the slow scaling of V_t required to meet the 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 larger than at least $2 \times V_t$. Since dynamic power dissipation is proportional to $(V_{dd})^2$, the dynamic power dissipation for the low-power devices scales more slowly than for high-performance devices, particularly in the later years. Also, since the lateral electric field ~ (V_{dd}/[gate length]), this field increases sharply with device scaling, and this sharp increase will result in difficulty in controlling short channel effects and possibly in significant reliability problems in the long-term years. A critical issue is the gate leakage current. For the LSTP transistors, given the required EOT in 2005, the leakage current targets cannot be met using oxy-nitride because of direct tunneling *(see the Front End Processes chapter for detailed discussion of this point)*. Hence, high κ gate dielectric will be required in 2005 for LSTP ICs.

The issue of high κ material in place of silicon dioxide or silicon oxy-nitride for the gate dielectric is critically important *(see Front End Processes chapter for detailed discussion).* As mentioned above, *the driver for replacement of silicon oxy-nitride with medium or high κ material for the gate dielectric is LSTP chips in 2005.* The 2005 requirements are:

LSTP: EOT = 1.8 nm, gate leakage current = 1 pA/μm.

In contrast, for LOP chips, it appears that the gate leakage current targets can be met, at least barely, by using heavily nitrided oxy-nitride until the end of the ITRS timeframe (see FEP chapter). Also, as mentioned before, for high-performance chips, the allowable leakage current increases so quickly with device scaling that silicon oxy-nitride appears capable of easily meeting the gate leakage requirements until the end of the 2001 Roadmap timeframe. However, there are serious issues of thickness control and uniformity for very thin oxy-nitride gate dielectrics and there are serious reliability issues, especially with large gate leakage current. Furthermore, it would be desirable from the standpoint of static power dissipation control to reduce the gate leakage below the large values specified in the tables, particularly for the later years in the Roadmap, and numerous companies will probably do so. As a result, a reasonable scenario is that high κ material will be implemented for the gate dielectric in LSTP chips in 2005, with the EOT and leakage requirements above. Several years later, in the 2007 and beyond time frame, after going down the learning curve with the LSTP chips and further development, high κ gate dielectrics will be utilized for LOP and high-performance chips.

Table 35a High-performance Logic Technology Requirements—Near-term

YEAR OF PRODUCTION	2001	2002	2003	2004	2005	2006	2007
DRAM ½ PITCH (nm)	130	115	100	90	80	70	65
MPU / ASIC ½ PITCH (nm)	150	130	107	90	80	70	65
MPU PRINTED GATE LENGTH (nm)	90	75	65	53	45	40	35
MPU PHYSICAL GATE LENGTH (nm)	65	53	45	37	32	28	25
Physical gate length high-performance (HP) (nm) [1]	65	53	45	37	32	28	25
Equivalent physical oxide thickness for high-performance T_{ox} (EOT)(nm) [2]	1.3–1.6	1.2–1.5	1.1–1.6	0.9–1.4	0.8–1.3	0.7–1.2	0.6–1.1
Gate depletion and quantum effects electrical thickness adjustment factor (nm) [3]	0.8	0.8	0.8	0.8	0.8	0.8	0.5
T_{ox} electrical equivalent (nm) [4]	2.3	2.1	2.0	2.0	1.9	1.9	1.4
Nominal power supply voltage (V_{dd}) (V) [5]	1.2	1.1	1.0	1.0	0.9	0.9	0.7
Nominal high-performance NMOS sub-threshold leakage current, $I_{sd,leak}$ (at 25° C) (μA/μm) [6]	0.01	0.03	0.07	0.1	0.3	0.7	1
Nominal high-performance NMOS saturation drive current, I_{dd} (at V_{dd}, at 25° C) (μA/μm) [7]	900	900	900	900	900	900	900
Required percent current-drive "mobility/transconductance improvement" [8]	0%	0%	0%	0%	0%	0%	0%
Parasitic source/drain resistance (Rsd) (ohm-μm) [9]	190	180	180	180	180	170	140
Parasitic source/drain resistance (Rsd) percent of ideal channel resistance (V_{dd}/I_{dd}) [10]	16%	16%	17%	18%	19%	19%	20%
Parasitic capacitance percent of ideal gate capacitance [11]	19%	22%	24%	27%	29%	32%	27%
High-performance NMOS device τ (C_{gate} * V_{dd} / I_{dd}-NMOS)(ps) [12]	1.6	1.3	1.1	0.99	0.83	0.76	0.68
Relative device performance [13]	1.0	1.2	1.5	1.6	2.0	2.1	2.5
Energy per (W/L_{gate}=3) device switching transition (C_{gate}*(3*L_{gate})*V^2) (fJ/Device) [14]	0.347	0.212	0.137	0.099	0.065	0.052	0.032
Static power dissipation per (W/Lgate=3) device (Watts/Device) [15]	5.6E-09	6.7E-09	1.0E-08	1.1E-08	2.6E-08	5.3E-08	5.3E-08

Table 35b High-performance Logic Technology Requirements—Long-term

YEAR OF PRODUCTION	2010	2013	2016
DRAM ½ PITCH (nm)	45	32	22
MPU / ASIC ½ PITCH (nm)	50	35	25
MPU PRINTED GATE LENGTH (nm)	25	18	13
MPU PHYSICAL GATE LENGTH (nm)	18	13	9
Physical gate length high-performance (HP) (nm) [1]	18	13	9
Equivalent physical oxide thickness for high-performance T_{ox} (EOT)(nm) [2]	0.5-0.8	0.4-0.6	0.4-0.5
Gate depletion and quantum effects electrical thickness adjustment factor (nm) [3]	0.5	0.5	0.5
T_{ox} electrical equivalent (nm) [4]	1.2	1.0	0.9
Nominal power supply voltage (V_{dd}) (V) [5]	0.6	0.5	0.4
Nominal high-performance NMOS sub threshold leakage current, $I_{sd,leak}$ (at 25° C) (μA/μm) [6]	3	7	10
Nominal high-performance NMOS saturation drive current , I_{dd} (at V_{dd}, at 25° C) (μA/μm) [7]	1200	1500	1500
Required percent current-drive "mobility/transconductance improvement" [8]	30%	70%	100%
Parasitic source/drain resistance (Rsd) (ohm-μm) [9]	110	90	80
Parasitic source/drain resistance (Rsd) percent of ideal channel resistance (V_{dd}/I_{dd}) [10]	25%	30%	35%
Parasitic capacitance percent of ideal gate capacitance [11]	31%	36%	42%
High-performance NMOS device τ (C_{gate} * V_{dd} / I_{dd}-NMOS)(ps) [12]	0.39	0.22	0.15
Relative device performance [13]	4.3	7.2	10.7
Energy per (W/L_{gate}=3) device switching transition (C_{gate}*(3*L_{gate})*V^2) (fJ/Device) [14]	0.015	0.007	0.002
Static power dissipation per (W/Lgate=3) device (Watts/Device) [15]	9.7E-08	1.4E-07	1.1E-07

White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

Notes for Table 35a and b:

[1] 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 ± 10%. Gate dimension variation is assumed to be the primary factor responsible for driving device parameter variation.

[2] EOT range set by FEP TWG. Yellow/red feasibility coloring set by FEP TWG projections on thickness control and reliability capability. Calculations in rest of PIDS table (and the underlying PIDS workbook) based on approximate midpoint EOT values.

[3] Accounts for gate electrode depletion and inversion-layer quantum effects. Yellow feasibility coloring reflects FEP assessment of polysilicon doping capability; red feasibility coloring reflects the introduction of metal-gate electrodes by 2007 (which reduces the gate depletion value).

[4] Sum of midpoint EOT and Electrical Thickness Adjustment Factor. Used in CV/I performance metric and CV^2 dynamic power metric calculations. Red/yellow feasibility coloring determined by worst-case EOT and Electrical Thickness Adjustment Factor red/yellow feasibility coloring.

[5] Nominal power supply voltage has been set to maintain sufficient voltage over-drive to continue historical approximate 17% per year device performance scaling while still enabling approximate 30% per year switching energy reduction and still maintaining reasonable vertical gate dielectric electric field strengths. Actual power supply voltage values may vary +/- 10%, depending on the particular circuit design application or technology optimization.

[6] Nominal sub-threshold leakage current is defined as the NMOSFET source current at room temperature with the drain bias set equal to the nominal power-supply voltage 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 at room temperature, 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 either the gate or junction leakage current components at either room or high-temperature conditions. The threshold voltage value (and the corresponding sub-threshold current) has been set to maintain sufficient voltage over-drive to continue historical approximate 17% per year device performance scaling. Yellow feasibility coloring by 2007 reflects the potential need for non-classical CMOS or ultra-shallow junction technology in order to control short-channel and/or high-field effects (See Emerging Research Devices section). The above sub-threshold, gate, and junction current scaling scenario also applies to PMOS devices. Note, sub-threshold current value applies to fastest MOS devices only; slower/lower-leakage MOS devices will also be available. Future systems will consist of a mix of both high and lower-leakage devices.

[7] Nominal saturation current drive is defined as the MOSFET drain current at room temperature with the gate bias and the drain bias set equal to the nominal power-supply voltage; 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 feasibility coloring indicates the projected need for fundamental device current-drive (or transconductance/mobility) improvement by 2010 in order to continue historical approximate 17% per year device performance scaling. NMOS/PMOS current-drive targets are approximate with only 1.5 significant digits of accuracy.

[8] Fundamental device mobility/transconductance improvement needed by 2010 in order to continue historical approximate 17% per year device performance scaling. Yellow/red feasibility coloring indicates the difficulty in implementing fundamental device current-drive (transconductance/mobility) improvement required in 2010 and later to continue historical approximate 17% per year device performance scaling.

[9] Total parasitic device source/drain resistance Rsd (sum of the source and drain parasitic resistances). Rsd targets are consistent with FEP TWG projections. Yellow/red feasibility coloring reflects FEP TWG projections on contact resistance, salicide sheet resistance, and drain extension scaling. Similar Rsd values are assumed for the LOP and LSTP devices.

[10] Maximum ratio of the parasitic device source/drain resistance (Rsd) to the ideal channel resistance (Vdd/Idd). Rsd targets are consistent with FEP TWG projections. Yellow/red feasibility coloring reflects FEP TWG projections on contact resistance, salicide sheet resistance, and drain extension scaling.

[11] Maximum ratio of the parasitic gate overlap/fringing capacitance to the ideal gate capacitance. Assume constant C-parasitic value of 2.4E-16F/um. [3× the fringing capacitance value per side, including the Miller effect]; this value is assumed to be independent of bias conditions and/or technology. Parasitic capacitance factor is included in CV/I and CV^2 performance and power metric calculations. Similar parasitic capacitance values are assumed for the LOP and LSTP devices.

[12] τ = CV/I intrinsic delay metric for NMOS device; PMOS CV/I metric assumed to scale proportionally. The CV/I metric provides an indication of the intrinsic switching delay of the device, while $1/\tau$, the reciprocal of CV/I, is a good metric for the intrinsic switching speed of the device. Red/yellow feasibility coloring determined by worst-case saturation current-drive feasibility coloring.

[13] Improvement in $1/\tau$ NMOS performance metric normalized to the year 2001. Maintains historical approximate 17% per year device performance improvement scaling trend. Red/yellow feasibility coloring determined by worst-case, saturation current-drive feasibility coloring.

[14] CV^2 switching energy metric for an NMOS device with dimensions W/Lgate=3. The switching energy metric indicates the amount of dynamic power required to switch the device. Maintains approximate 30% per year device switching energy reduction scaling trend. Red/yellow feasibility coloring determined by worst-case saturation current-drive feasibility and mobility/transconductance improvement feasibility coloring.

[15] Static power dissipation for an NMOS device with dimensions W/Lgate=3. Assume that the device static power is primarily determined by the sub-threshold current (since the other junction and leakage current components are assumed to be lower). Yellow feasibility coloring by 2007 reflects the potential need for non-classical CMOS or ultra-shallow junction technology in order to control short-channel and/or high-field effects (See Emerging Research Devices section, Non-classical CMOS).

Table 36a Low Operating Power (LOP) Logic Technology Requirements—Near-term

YEAR OF PRODUCTION	2001	2002	2003	2004	2005	2006	2007
DRAM ½ PITCH (nm)	130	115	100	90	80	70	65
MPU / ASIC ½ PITCH (nm)	150	130	107	90	80	70	65
MPU PRINTED GATE LENGTH (nm)	90	75	65	53	45	40	35
MPU PHYSICAL GATE LENGTH (nm)	65	53	45	37	32	28	25
Physical gate length low-operating power (LOP) (nm) [1]	90	75	65	53	45	37	32
Equivalent physical oxide thickness for LOP T_{ox} (EOT) (nm) [2]	2.0-2.4	1.8-2.2	1.6-2.0	1.4-1.8	1.2-1.6	1.1-1.5	1.0-1.4
Electrical thickness adjustment factor (gate depletion and quantum effects) (nm) [3]	0.8	0.8	0.8	0.8	0.8	0.8	0.5
T_{ox} electrical equivalent (nm) [4]	3.0	2.8	2.6	2.4	2.2	2.1	1.7
Nominal LOP power supply voltage (V_{dd}) (V) [5]	1.2	1.2	1.1	1.1	1	1	0.9
Nominal LOP NMOS sub-threshold leakage current, $I_{sd,leak}$ (@25C) (pA/μm) [6]	100	100	100	300	300	300	700
Nominal LOP NMOS Saturation drive current, I_{dd} (@V_{dd}, @25C) (μA/μm) [7]	600	600	600	600	600	600	700
Required percent current-drive "mobility/transconductance improvement" [8]	0%	0%	0%	0%	0%	0%	0%
LOP NMOS Device τ (C_{gate} * V_{dd} / Id-NMOS) (ps) [9]	2.55	2.45	2.02	1.84	1.58	1.41	1.14
LOP relative device performance [10]	1.0	1.04	1.3	1.4	1.6	1.8	2.2
Energy per (W/L_{gate}=3) device switching transition (C_{gate}*(3*L_{gate})*V^2) (fJ/Device) [11]	0.496	0.424	0.260	0.193	0.128	0.094	0.069
Static power dissipation per (W/L_{gate}=3) device (Watts/device) [12]	3.2E-11	2.9E-11	2.1E-11	5.2E-11	4.1E-11	3.3E-11	6.0E-11

Table 36b Low Operating Power (LOP) Logic Technology Requirements—Long-term

YEAR OF PRODUCTION	2010	2013	2016
DRAM ½ PITCH (nm)	45	32	22
MPU / ASIC ½ PITCH (nm)	50	35	25
MPU PRINTED GATE LENGTH (nm)	25	18	13
MPU PHYSICAL GATE LENGTH (nm)]	18	13	9
Physical gate length low-operating power (LOP) (nm) [1]	22	16	11
Equivalent physical oxide thickness for LOP T_{ox} (EOT) (nm) [2]	0.8-1.2	0.7-1.1	0.6-1.0
Electrical thickness adjustment factor (gate depletion and quantum effects) (nm) [3]	0.5	0.5	0.5
T_{ox} electrical equivalent (nm) [4]	1.5	1.4	1.3
Nominal LOP power supply voltage (V_{dd}) (V) [5]	0.8	0.7	0.6
Nominal LOP NMOS sub-threshold leakage current, $I_{sd,leak}$ (@25C) (pA/μm) [6]	1000	3000	10000
Nominal LOP NMOS saturation drive current, I_{dd} (@V_{dd}, @25C) (μA/μm) [7]	700	800	900
Required percent current-drive "mobility/transconductance improvement" [8]	10%	30%	70%
LOP NMOS device τ (C_{gate} * V_{dd} / I_{dd}-NMOS) (ps) [9]	0.85	0.56	0.35
LOP relative device Performance [10]	3.0	4.6	7.2
Energy per (W/L_{gate}=3) device switching transition (C_{gate}*(3*L_{gate})*V^2) (fJ/Device) [11]	0.032	0.015	0.006
Static power dissipation per (W/Lgate=3) device (Watts/Device) [12]	5.3E-11	1.0E-10	2.0E-10

White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

Table 36c Low Standby Power (LSTP) Technology Requirements—Near-term

YEAR OF PRODUCTION	2001	2002	2003	2004	2005	2006	2007
DRAM ½ PITCH (nm)	130	115	100	90	80	70	65
MPU / ASIC ½ PITCH (nm)	150	130	107	90	80	70	65
MPU PRINTED GATE LENGTH (nm)	90	75	65	53	45	40	35
MPU PHYSICAL GATE LENGTH (nm)	65	53	45	37	32	28	25
Physical gate length low-standby power (LSTP) (nm) [1]	90	75	65	53	45	37	32
Equivalent physical oxide thickness for LSTP T_{ox} (EOT) (nm) [2]	2.4–2.8	2.2–2.6	2.0–2.4	1.8–2.2	1.6–2.0	1.4–1.8	1.2–1.6
Electrical thickness adjustment factor (gate depletion and quantum effects) (nm) [3]	0.8	0.8	0.8	0.8	0.8	0.8	0.5
T_{ox} electrical equivalent (nm) [4]	3.4	3.2	3.0	2.8	2.6	2.4	1.9
Nominal LSTP power supply voltage (V_{dd}) (V) [5]	1.2	1.2	1.2	1.2	1.2	1.2	1.1
Nominal LSTP NMOS sub-threshold current (at 25 °C) (pA/μm) [6]	1	1	1	1	1	1	1
Nominal LSTP NMOS saturation current drive (I_{dd}) (at V_{dd}, at 25 °C) (mA/μm) [7]	300	300	400	400	400	400	500
Required percent current-drive "mobility/transconductance improvement" [8]	0%	0%	0%	0%	0%	0%	0%
LSTP NMOS device τ (C_{gate} * V_{dd} / Id-NMOS) (ps) [9]	4.61	4.41	2.96	2.68	2.51	2.32	1.81
LSTP relative device performance [10]	1.00	1.05	1.6	1.7	1.8	2.0	2.6
Energy per (W/L_{gate}=3) device switching transition (C_{gate}*(3*L_{gate})*V^2) (fJ/device) [11]	0.448	0.381	0.277	0.204	0.163	0.123	0.095
Static power dissipation per (W/L_{gate}=3) device (Watts/device) [12]	3.2E-13	2.9E-13	2.3E-13	1.9E-13	1.6E-13	1.3E-13	1.1E-13

Table 36d Low Standby Power (LSTP) Technology Requirements—Long-term

YEAR OF PRODUCTION	2010	2013	2016
DRAM ½ PITCH (nm)	45	32	22
MPU / ASIC ½ PITCH (nm)	50	35	25
MPU PRINTED GATE LENGTH (nm)	25	18	13
MPU PHYSICAL GATE LENGTH (nm)	18	13	9
Physical gate length low-standby power (LSTP) (nm) [1]	22	16	11
Equivalent physical oxide thickness for LSTP T_{ox} (EOT) (nm) [2]	0.9-1.3	0.8-1.2	0.7-1.1
Electrical thickness adjustment factor (gate depletion and quantum effects) (nm) [3]	0.5	0.5	0.5
T_{ox} electrical equivalent (nm) [4]	1.6	1.5	1.4
Nominal LSTP power supply voltage (V_{dd}) (V) [5]	1	0.9	0.9
Nominal LSTP NMOS sub-threshold current (at 25 °C) (pA/μm) [6]	3	7	10
Nominal LSTP NMOS saturation current drive (I_{dd}) (at V_{dd}, at 25 °C) (mA/μm) [7]	500	600	700
Required percent current-drive "mobility/transconductance improvement" [8]	10%	30%	50%
LSTP NMOS device τ (C_{gate} * V_{dd} / Id-NMOS) (ps) [9]	1.43	0.91	0.66
LSTP relative device performance [10]	3.2	5.1	7.0
Energy per (W/L_{gate}=3) device switching transition (C_{gate}*(3*L_{gate})*V^2) (fJ/device) [11]	0.047	0.024	0.014
Static power dissipation per (W/L_{gate}=3) device (Watts/device) [12]	2.0E-13	3.0E-13	3.0E-13

White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

Notes for Table 36a through d:

[1] Values set by ORTC.

[1] Values set by ORTC. Assumed to lag high-performance scaling by 2 years. Gate dimensional control is set by the Lithography and FEP (Etch) ITWGs and is assumed to have a three sigma value of ±10%. Gate dimension variation is assumed to be the primary factor responsible for driving device parameter variation. Gate length for the LOP and LSTP devices are assumed to be identical.

[2] EOT value set by FEP ITWG in the Front End Processes chapter, Thermal and Thin Film, Doping, and Etching Technology Requirements tables Yellow/red feasibility coloring set by FEP ITWG projections on gate leakage, thickness control and reliability capability (high κ gate dielectrics will be required around 2005 in order to suppress gate leakage for LSTP). Due to different system applications, the EOT values for the LOP and LSTP devices have been optimized and set independently of each other. Calculations in rest of PIDS table (and the underlying PIDS workbook) based on approximate midpoint EOT values.

[3] Accounts for gate electrode depletion and inversion-layer quantum effects. Yellow feasibility coloring reflects FEP assessment of polysilicon doping capability; red feasibility coloring reflects the introduction of metal-gate electrodes by 2007 (which reduces the gate depletion value).

[4] Sum of midpoint EOT and Electrical Thickness Adjustment Factor. Used in CV/I performance metric and CV^2 dynamic power metric calculations. Red/yellow feasibility coloring determined by worst-case EOT and Electrical Thickness Adjustment Factor red/yellow feasibility coloring.

[5] Nominal power supply voltage which has been set to smallest value to still maintain sufficient voltage over-drive to allow sufficient circuit switching noise margin (approximately 2.3 times the threshold voltage). Actual power supply voltage values may vary ± 10%, depending on the particular circuit design application or technology optimization. Due to different system applications, the power-supply voltages for the LOP and LSTP devices have been optimized and set independently of each other. Note, meeting overall system power dissipation requirements will require the use of circuit/system techniques to "turn-off" or "power-down" various circuit blocks.

[6] Nominal sub-threshold leakage current is defined as the NMOSFET source current at room temperature with the drain bias set equal to the nominal power-supply voltage 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 at room temperature, 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 either the gate or junction current components at either room or high-temperature conditions. The increase in sub-threshold current (and the corresponding threshold-voltage value reduction) has been set at a pace that lags the rate of increase of the high-performance device, but that still increases sufficiently to enable continued device performance scaling; power dissipation due to off-state leakage is assumed to not exceed 10% of the total chip power, which is assumed to be 100mW in 2001 and 200mW in 2010 for LOP. For LOP, the yellow feasibility coloring reflects the difficulty of meeting the gate leakage requirements with thin oxy-nitride films and the potential need for non-classical CMOS or ultra-shallow junction technology by 2007 to control short-channel effects and to limit lateral high-field effects. For LSTP, the yellow feasibility coloring in 2003 and 2004 reflects the difficulty of meeting the gate leakage requirements with thin oxy-nitride gate dielectric films, while the red feasibility color from 2005 on reflects the difficulty of implementing a high κ gate dielectric to meet the gate leakage requirements. The above sub-threshold, gate, and junction current scaling scenario also applies to PMOS devices.

[7] Nominal saturation current drive is defined as the MOSFET drain current at room temperature with the gate bias and the drain bias set equal to the nominal power-supply voltage; 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 feasibility coloring indicates the projected need for fundamental device current-drive (transconductance/mobility) improvement by 2010 in order to continue approximate 14% per year device performance scaling. NMOS/PMOS current-drive targets are approximate with only 1.5 significant digits of accuracy. The sub-threshold slope, parasitic source/drain resistance, and parasitic gate capacitance scaling have been assumed to be similar to that for the high-performance device. (See the enclosed Excel workbook for the detailed calculations).

[8] Fundamental device mobility/transconductance improvement needed by 2010 in order to continue historical approximate 14% per year device performance scaling. Yellow/red feasibility coloring indicates the difficulty of implementing the fundamental device current-drive (transconductance/mobility) improvement required in 2010 and beyond to continue historical approximate 14% per year device performance scaling. The LOP and LSTP required improvement is projected as lagging that required for the high-performance device.

[9] τ = CV/I intrinsic delay metric for NMOS device; PMOS CV/I metric assumed to scale proportionally. The CV/I metric provides an indication of the intrinsic switching delay of the device, while $1/\tau$, the reciprocal of CV/I, is a good metric for the intrinsic switching speed of the device. Red/yellow feasibility coloring determined by saturation current-drive feasibility coloring. The C term includes the effect of parasitic gate capacitance, which has been assumed to be equivalent to that for the high-performance device.

[10] Improvement in $1/\tau$ NMOS performance metric normalized to the year 2001. Maintains approximate 14% per year device performance improvement scaling trend for both LOP and LSTP. Red/yellow feasibility coloring determined by saturation current-drive feasibility coloring.

[11] CV^2 switching energy metric for an NMOS device with dimensions $W/L_{gate}=3$. The switching energy metric indicates the amount of dynamic power required to switch the device. Red/yellow feasibility coloring determined by saturation current-drive feasibility coloring.

[12] Static power dissipation for an NMOS device with dimensions $W/L_{gate}=3$. Assume that the device static power is primarily determined by the sub-threshold current (since the other junction and leakage current components are assumed to be lower). Yellow/red feasibility coloring is determined by the sub-threshold leakage current feasibility coloring.

DRAM TECHNOLOGY REQUIREMENTS

Technical requirements for DRAMs are projected to become increasingly stringent with scaling. Photoresists associated with 193nm exposure wavelength lithography, which is slated for production in 2003 at 100nm DRAM half pitch, have 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. Also, 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. A vertical access device has been proposed,[1] but it has yet to be seen in mainstream production. Other process requirements for DRAMs such as front end isolation, low resistance materials for the word lines and bit 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 Table 37a 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) for the 130nm node and below. 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, on the order of 100 and greater, will be needed at the 65nm node in 2007. All in all, maintaining sufficient storage capacitance will pose an increasingly difficult requirement for continued scaling of DRAM devices.

Table 37a DRAM Technology Requirements—Near-term

YEAR OF PRODUCTION	2001	2002	2003	2004	2005	2006	2007
DRAM 1/2 Pitch (nm)[1]	130	115	100	90	80	70	65
MPU / ASIC ½ PITCH (nm)	150	130	107	90	80	70	65
MPU PRINTED GATE LENGTH (nm)	90	75	65	53	45	40	35
MPU PHYSICAL GATE LENGTH (nm)	65	53	45	37	32	28	25
DRAM cell size (μm^2) [2]	0.135	0.106	0.060	0.049	0.038	0.029	0.025
DRAM storage cell dielectric: equivalent physical oxide thickness, EOT (nm) [3]	2.04	1.80	1.20	1.00	0.45	0.32	0.22
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 37b DRAM Technology Requirements—Long-term

YEAR OF PRODUCTION	2010	2013	2016
DRAM ½ PITCH (nm) [1]	45	32	22
MPU / ASIC ½ PITCH (nm)	50	35	25
MPU PRINTED GATE LENGTH (nm)	25	18	13
MPU PHYSICAL GATE LENGTH (nm)	18	13	9
DRAM cell size (μm^2) [2]	0.0122	0.0041	0.0019
DRAM storage cell dielectric: equivalent physical oxide thickness, EOT (nm) [3]	0.084	0.028	0.010
DRAM retention time (ms) [4]	64	64	64
DRAM soft error rate (fits) [5]	1000	1000	1000

White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

Notes for Table 37a and b:

[1] From ORTC (Overall Roadmap Technology Characteristics) Table 1a and b. These DRAM half pitch numbers are smaller than those in the 2000 ITRS, reflecting a speedup in technology development since 1999.

[1] *C. J. Radens et al., "An Orthogonal 6F2 Trench-Sidewall Vertical Device Cell for 4Gb/16Gb DRAM," IEDM tech. Digest, pp. 349-352, 2000.*

[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 130nm node, 6 for 100nm DRAM half pitch, and 4 for the 32nm node. The "a" value of 8 is probably achievable with current techniques, but the "a" value of 6 will require innovative solutions, as illustrated with yellow zone in this line, while the "a" value of 4 has no known solution for 32nm 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. For the 130nm through 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 κmaterial are needed, so the color is yellow. Finally, for the 65nm 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 85 ºC, 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.

NONVOLATILE MEMORY TECHNOLOGY REQUIREMENTS

Nonvolatile memory (NVM) imposes additional constraints on process integration and structure design. Typically the nonvolatile components are implemented as an add-on to an established CMOS process. As a result, emerging NVM lags behind the current CMOS technology node. To clarify the degree of delay, Table 38a and b identifies both the current CMOS node feature size and the NVM technology feature size. For Flash technologies the delay is about one year, while for FeRAM technologies the delay is more significant.

Flash scaling is complicated because NVM structures that require relatively high voltages must be incorporated into CMOS technology that is low voltage. Nonvolatility is achieved by storing and sensing the charge on a floating gate. The interpoly dielectric 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.

FeRAM scaling is complicated because the ferroelectric materials, buffer materials, and process conditions are still being refined. Nonvolatility is achieved by switching and sensing the polarization state of a ferroelectric capacitor. The ferroelectric material must be physically and chemically isolated from the underlying CMOS. In order to achieve density goals the basic geometry of the cell must be modified while maintaining the desired isolation.

The endurance (erase-write cycle or read-write cycle) ratings and the retention ratings are unique to NVM. These reliability oriented parameters determine whether the product has adequate utility to be of value to an end customer. Stresses imposed during normal operation of the devices may cause degradation of the part, and the endurance and retention ratings identify a “safe” range of use. Understanding subtle failure mechanisms is critical. Testing to confirm endurance and retention takes long times and is a serious impediment to the rapid evolution of the technology.

Table 38a Non-Volatile Memory Technology Requirements—Near-term

YEAR OF PRODUCTION	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU / ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
Flash technology node - F (nm) [1]	150	130	115	100	90	80	70
Flash NOR cell size –area factor a in multiples of F^2 [2]	10–12	10–12	10–12	11–14	11–14	11–14	11–14
Flash NAND cell size —area factor a in multiples of F^2 SLC/MLC [3]	5.5	5.5	4.5	4.5	4.5	4.5/2.3	4.5/2.3
Flash NOR typical cell size (μm^2) [4]	0.248	0.186	0.145	0.125	0.101	0.080	0.061
Flash NOR Lg-stack (physical- μm) [5]	0.29–0.31	0.25–0.27	0.22–0.24	0.21–0.23	0.2–0.22	0.2–0.22	0.19–0.21
Flash NOR highest W/E voltage (V) [6]	8–10	8–10	8–10	8–10	7–9	7–9	7–9
Flash NAND highest W/E voltage (V) [7]	19–21	18–20	18–20	18–20	18–20	17–19	17–19
Flash NOR I_{read} (μA) [8]	36–44	35–43	34–42	33–41	31–39	28–36	29–37
Flash Coupling Ratio [9]	0.65–0.75	0.65–0.75	0.65–0.75	0.65–0.75	0.65–0.75	0.65–0.75	0.6–0.7
Flash NOR tunnel oxide thickness (nm) [10]	9.5–10.5	9.5–10	9–10	9–10	8.5–9.5	8.5–9.5	8.5–9.5
Flash NAND tunnel oxide thickness (nm) [11]	8.5–9.5	8.5–9	8–9	8–9	8–9	7.5–8	7.5–8
Flash NOR interpoly dielectric thickness (nm) [12]	13–15	12–14	11–13	11–13	10–12	9–11	9–11
Flash NAND interpoly dielectric thickness (nm) [13]	14–16	13–15	12–14	12–14	12–14	11–13	10–12
Flash endurance (erase/write cycles) [14]	1E5	1E5	1E5	1E5	1E5	1E5	1E5
Flash nonvolatile data retention (years) [15]	10	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]	500	350	250	220	180	150	130
FeRAM cell size —area factor a in multiples of F^2 [18]	60	40	24	16	10	10	10
FeRAM cell size (μm^2) [19]	15	4.9	1.5	0.518	0.324	0.225	0.169
FeRAM cell structure [20]	2T2C	1T1C	1T1C	1T1C	1T1C	1T1C	1T1C
FeRAM capacitor structure [21]	planar	planar	stack	stack	stack	stack	3D
Ferro capacitor voltage (V) [22]	3.0	3.0	2.5	1.8	1.5	1.3	1.2
FeRAM endurance (read/write cycles) [23]	1E12	1E13	1E14	1E15	>1E16	>1E16	>1E16
FeRAM nonvolatile data retention (years) [24]	10	10	10	10	10	10	10

White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

Table 38b Non-Volatile Memory Technology Requirements—Long-term

YEAR OF PRODUCTION	2010	2013	2016
DRAM ½ PITCH (nm)	45	32	22
MPU / ASIC ½ PITCH (nm)	50	35	25
MPU PRINTED GATE LENGTH (nm)	25	18	13
MPU PHYSICAL GATE LENGTH (nm)	18	13	9
Flash technology node - F (nm) [1]	50	35	25
Flash NOR cell size –-area factor a in multiples of F^2 [2]	12–15	13–16	14–17
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
Flash NOR typical cell size (μm^2) [4]	0.034	0.018	0.010
Flash NOR Lg-stack (physical- μm) [5]	0.17–0.19	0.14–0.16	0.12–0.14
Flash NOR highest W/E voltage (V) [6]	7–9	7–9	7–9
Flash NAND highest W/E voltage (V) [7]	17–19	16–18	16–18
Flash NOR I_read (μA) [8]	27–33	25–31	22–28
Flash Coupling Ratio [9]	0.6–0.7	0.6–0.7	0.6–0.7
Flash NOR tunnel oxide thickness (nm) [10]	8–9	8	8
Flash NAND tunnel oxide thickness (nm) [11]	6–7	6–7	6–7
Flash NOR interpoly dielectric thickness (nm) [12]	8–10	6–8	4–6
Flash NAND interpoly dielectric thickness (nm) [13]	10–12	9–11	9–11
Flash endurance (erase/write cycles) [14]	1E6	1E6	1E7
Flash nonvolatile data retention (years) [15]	10–20	20	20
Flash maximum number of bits per cell (MLC) [16]	8	8	8
FeRAM technology node - F (nm) [17]	100	70	50
FeRAM cell size –-area factor a in multiples of F^2 [18]	8	8	8
FeRAM cell size (μm^2) [19]	0.080	0.039	0.020
FeRAM cell structure [20]	1T1C	1T1C	1T1C
FeRAM capacitor structure [21]	3D	3D	3D
Ferro capacitor voltage (V) [22]	1.0	0.7	0.7
FeRAM endurance (read/write cycles) [23]	>1E16	>1E16	>1E16
FeRAM nonvolatile data retention (years) [24]	10	10	10

White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

Notes for Table 38a and b:

[1] Flash devices tend to lag the current CMOS technology node by one year. This entry provides the F value for designs in the indicated time period.

[2] Flash NOR cell size is presented in terms of F^2 multiples of the Flash implementation technology node, a=cell area/F^2. Note the lack of long term scaling.

[3] a=cell area/F^2. Flash NAND enjoys a smaller cell size because much of the cell structure is shared among a group of cells. (SLC single level cell, MLC multilevel cell)

[4] The expected midrange "typical" Flash NOR cell size is presented in terms of micrometers squared.

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

[9] Ratio: (control gate capacitane to floating gate capacitance)/(total floating gate capacitance 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 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 with scaling down is the major issue

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

[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] FeRAM devices tend to significantly lag the CMOS current technology node. This entry provides the F value for designs in the indicated time period.

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

[19] FeRAM cell size is presented in terms of micrometers squared.

[20, 21] Typical cell designs are moving to one transistor and one capacitor. Ferroelectric material selection, capacitor geometry and positioning relative to the transistor are intertwined design decisions that enable achievement of the cell size objective.

[22] Low voltage operation is a difficult key design issue.

[23] For FeRAM to compete with DRAM and SRAM the cycle endurance should be about 1E15. Test time is a serious issue. Note that 100MHz x 10 years = 1E16 cycles.

[24] Unpowered data retention is usually specified at 85°C.

MIXED-SIGNAL DEVICE TECHNOLOGY REQUIREMENTS

The previously identified trends leading to higher integration levels of logic and RF with mixed-signal circuitry have continued and materialized into new application areas. The steadily increasing digital processing capabilities enable more and more signal treatment to be done in the digital domain. In addition, the use of multiple logic gate oxides, to accommodate higher voltages continues to support interfacing to the outside world and the signal-to-noise requirements for mixed-signal, although at the cost of, e.g., matching and 1/f noise performance. Moreover, the acceleration of the CMOS roadmap in the past years also has accelerated the integration possibilities of RF in logic processes. As before, continuous focus on 1/f noise, passive component density and device matching is imperative to satisfy the increasing demands on power and area efficiency. Emerging issues from this increased integration level are with RF device modeling and protection against electrostatic discharge.

Performance and cost considerations will continue to drive modularity of process features in order to adapt the technology to specific SoC architectures. However, the ever more stringent mixed-signal and/or RF transistor requirements may force the addition of process complexity to achieve integration goals. CMOS technology is expected to gain importance in the field of mixed-signal at the cost of bipolar and Si or SiGe based BiCMOS processes, which will however continue to be strong in the high-performance application areas that require high-linearity, high-speed and/or low noise, especially at low power. This strength in high performance comes about because the bipolar RF devices are carefully optimized and hence have intrinsic, but non-scaling, advantages in gain, noise and matching. In contrast, the CMOS RF devices come as they are from the baseline logic process, with extremely good frequency behavior combined with lesser, but improving, performance on the other parameters. This continued parallelism of technologies has been expressed in the mixed-signal table by having separate CMOS and bipolar device parameter requirements sections, making performance comparison possible. Refer to Table 39a and b.

Table 39a Mixed-signal Device Technology Requirements—Near-term

YEAR OF PRODUCTION		2001	2002	2003	2004	2005	2006	2007	OWNER
DRAM ½ PITCH (nm)		130	115	100	90	80	70	65	ORTC
MPU / ASIC ½ PITCH (nm)		150	130	107	90	80	70	65	ORTC
MPU PRINTED GATE LENGTH (nm)		90	75	65	53	45	40	35	ORTC
MPU PHYSICAL GATE LENGTH (nm)		65	53	45	37	32	28	25	ORTC
ASIC/Low Power Physical Gate Length (nm) [1]		90	75	65	53	45	37	32	ORTC
Minimum Supply Voltage	Digital Design (V)[2]	1.2	1.1	1.0	1.0	0.9	0.9	0.7	PIDS
	Analog Design (V) [3]	3.3–1.8		2.5–1.8					DESIGN
Frequency Range	RF (GHz) [4]	0.5–10			0.5–20			0.5–30	PIDS
	Analog (GHz)[5]	0.1–2			0.1–4			0.1–6	PIDS
Bipolar RF Device	Current (μA) [6]	100	100	75	75	75	50	50	PIDS
	fmax (GHz) [7]	90	100	110	120	130	140	160	PIDS
	ft (GHz) * [8]	45	50	55	60	65	70	80	PIDS
	Gm / Gce @We-min ** [9]	1250	1250	1200	1200	1200	1150	1150	PIDS
	1/f Noise ($\mu V^2 \cdot \mu m^2$ / Hz) [10]	10	10	5	5	5	2.5	2.5	PIDS
Bipolar Analog Device	Current (μA) [11]	65	60	55	50	45	40	35	PIDS
	1/f Noise ($\mu V^2 \cdot \mu m^2$ / Hz)[12]	10	10	5	5	5	3	3	PIDS
	3σ current matching (%) [13]	1	1	1	1	1	1	1	PIDS
NMOS RF Device	T_{ox} (nm) [14]	1.3–1.6	1.2–1.5	1.1–1.6	0.9–1.4	0.8–1.3	0.7–1.2	0.6–1.1	PIDS
	fmax (GHz) [15]	160	165	170	175	180	185	190	DESIGN
	ft (GHz) *** [16]	132	149	183	225	264	322	372	PIDS
	Gm / Gds @Lmin-digital [17]	20	20	20	20	20	20	20	DESIGN
	@10·Lmin-digital [18]	100	100	100	100	100	100	100	DESIGN
	1/f Noise ($\mu V^2 \cdot \mu m^2$ / Hz) [19]	500	500	300	300	300	200	200	DESIGN
	3σ V_t matching (mV·μm) [20]	5	5	5	5	4	4	3	DESIGN
NMOS Analog Device	T_{ox} (nm) [21]	7–2.5	7–2.5	5–2.5	5–2.5	5–2.5	5–2.5	5–2.5	PIDS
	Analog V_t (V) [22]	0.5–0.3	0.5–0.2	0.5–0.2	0.5–0.2	0.4–0.2	0.4–0.2	0.4–0.2	DESIGN
	Gm / Gds @10·Lmin-digital [23]	200	200	200	200	200	200	200	DESIGN
	1/f Noise ($\mu V^2 \cdot \mu m^2$ / Hz) [24]	1000	500	500	500	300	300	300	DESIGN
	3σ V_t matching (mV·μm)[25]	21	21	15	15	15	15	15	DESIGN
Analog Capacitor	Density (fF/μm^2) [26]	2	3	3	3	4	4	4	DESIGN
	Q (1 / $k\Omega^2 \cdot \mu m^2 \cdot$GHz) [27]	200	300	300	300	450	450	450	DESIGN
	Voltage linearity (ppm / V^2) [28]	100	100	100	100	100	100	100	DESIGN
	Leakage (fA / [pF·V])[29]	7	7	7	7	7	7	7	DESIGN
	3σ Matching (%·μm^2)[30]	4.5	3	3	3	2.5	2.5	2.5	DESIGN
RF Bypass Capacitor	Density (fF/μm^2)[31]	7	7.5	8	9	10	11	12	PIDS
	Q (1 / $k\Omega^2 \cdot \mu m^2 \cdot$GHz) [32]	22	25	27	29	30	30	30	PIDS
	Voltage linearity (ppm / V) [33]	1000	1000	1000	1000	1000	1000	1000	PIDS
Resistor	Resistance (Ω/sq.)[34]	100	100	100	100	100	100	100	DESIGN
	Q ($k\Omega^2 \cdot \mu m^2 \cdot$GHz)[35]	1000	1500	1500	1500	2000	2000	2000	DESIGN
	Temp. linearity (ppm / °C)[36]	60	60	50	50	50	40	40	DESIGN
	3σ Matching (%·μm) [37]	9	8	8	8	7	7	7	DESIGN
	1/f current noise per current^2 (1 / [μm^2·Hz]) [38]	10^{-18}	10^{-18}	10^{-18}	10^{-18}	10^{-18}	10^{-18}	10^{-18}	DESIGN
Inductor	Density (nH/μm^2)[39]	0.03	0.03	0.03	0.03	0.03	0.03	0.03	DESIGN
	Q_{3dB} [40]	12	15	17	18	19	20	20	DESIGN
Signal Isolation ****	Transmission coefficient S21 (dB) [41]	-100	-100	-100	-100	-120	-120	-120	PIDS

* *Value optimized for frequency range of application [4], not the maximum possible as for RF-CMOS [15,16].*
** *Unusual measure for the voltage gain of bipolar devices, see notes. Added for 1:1 comparison with Gm/Gds of CMOS transistors [17].*
*** *Value completely determined by ASIC/LP physical gate length [1], this in contrast to bipolar devices.*
**** *For reference purposes only. Signal isolation is a complex of noise avoidance and noise suppression both in design and by technology.*

White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

Table 39b Mixed-signal Device Technology Requirements—Long-term

YEAR OF PRODUCTION		2010	2013	2016	OWNER
DRAM ½ PITCH (nm)		45	32	22	ORTC
MPU / ASIC ½ PITCH (nm)		50	35	25	ORTC
MPU PRINTED GATE LENGTH (nm)		25	18	13	ORTC
MPU PHYSICAL GATE LENGTH (nm)		18	13	9	ORTC
ASIC/Low Power Physical Gate Length (nm) [1]		22	16	11	ORTC
Minimum Supply Voltage	Digital Design (V) [2]	0.6	0.5	0.4	PIDS
	Analog Design (V) [3]	1.8–1.0		1.5–1.0	DESIGN
Frequency Range	RF (GHz)[4]	0.5–50	0.5–75	0.5–100	PIDS
	Analog (GHz)[5]	0.1–10	0.1–15	0.1–20	PIDS
Bipolar RF Device	Current (μA) [6]	50	50	50	PIDS
	fmax (GHz) [7]	190	220	250	PIDS
	ft (GHz) * [8]	95	110	125	PIDS
	Gm / Gce @We-min ** [9]	1100	1050	1000	PIDS
	1/f Noise ($\mu V^2 \cdot \mu m^2$ / Hz) [10]	1.3	0.7	0.4	PIDS
Bipolar Analog Device	Current (μA) [11]	30	20	15	PIDS
	1/f Noise ($\mu V^2 \cdot \mu m^2$ / Hz) [12]	1.3	0.7	0.4	PIDS
	3σ current matching (%) [13]	1	1	1	PIDS
NMOS RF Device	T_{ox} (nm) [14]	0.5–0.8	0.4–0.6	0.4–0.5	PIDS
	fmax (GHz) [15]	200–230	230–260	260–290	DESIGN
	ft (GHz) *** [16]	541	744	1082	PIDS
	Gm / Gds @Lmin-digital [17]	20	20	20	DESIGN
	@10·Lmin-digital [18]	100	100	100	DESIGN
	1/f Noise ($\mu V^2 \cdot \mu m^2$ / Hz) [19]	150	100	75	DESIGN
	3σ V_t matching (mV·μm) [20]	3	2.5	2.5	DESIGN
NMOS Analog Device	T_{ox} (nm) [21]	3–1.3	3–1.3	2.5–1.3	PIDS
	Analog V_{th} (V) [22]	0.3–0.1	0.3–0.1	0.2–0.1	DESIGN
	Gm / Gds @10·Lmin-digital [23]	200	200	200	DESIGN
	1/f Noise ($\mu V^2 \cdot \mu m^2$ / Hz) [24]	150	100	100	DESIGN
	3σ V_{th} matching (mV·μm) [25]	9	9	7.5	DESIGN
Analog Capacitor	Density (fF/μm^2) [26]	7	10	15	DESIGN
	Q (1 / $k\Omega^2 \cdot \mu m^2 \cdot$GHz) [27]	700	1000	1500	DESIGN
	Voltage linearity (ppm / V^2) [28]	100	100	100	DESIGN
	Leakage (fA / [pF·V]) [29]	7	7	7	DESIGN
	3σ Matching (%·μm^2) [30]	2	1.5	1	DESIGN
RF Bypass Capacitor	Density (fF/μm^2) [31]	17	20	23	PIDS
	Q (1 / $k\Omega^2 \cdot \mu m^2 \cdot$GHz) [32]	35	40	40	PIDS
	Voltage linearity (ppm / V) [33]	1000	1000	1000	PIDS
Resistor	Resistance (Ω /) [34]	100	100	100	DESIGN
	Q ($k\Omega^2 \cdot \mu m^2 \cdot$GHz) [35]	3000	4500	6000	DESIGN
	Temp. linearity (ppm / °C) [36]	30	30	30	DESIGN
	3σ Matching (%·μm) [37]	6	6	6	DESIGN
	1/f current noise per current^2 (1 / [μm^2·Hz]) [38]	$4x10^{-19}$	$3x10^{-19}$	$2x10^{-19}$	DESIGN
Inductor	Density (nH/μm^2) [39]	0.025	0.02	0.01	DESIGN
	Q_{3dB} [40]	30	40	50	DESIGN
Signal Isolation	Transmission coefficient S21 (dB) [41]	-120	-120	-120	PIDS

* *Value optimized for frequency range of application [4], not the maximum possible as for RF-CMOS [15,16].*

** *Unusual measure for the gain of bipolar devices added for 1:1 comparison with Gm/Gds for CMOS transistors [17].*

*** *Value completely determined by ASIC/LP physical gate length [1], this in contrast to bipolar devices.*

**** *For reference purposes only. Signal isolation is a complex of noise avoidance and noise suppression*

White–Manufacturable Solutions Exist, and Are Being Optimized
Yellow--Manufacturable Solutions are Known
Red–Manufacturable Solutions are NOT Known

Notes for Table 39a and b

[1] *Year of first digital product for a given technology generation as given in ORTC Table 1a and b. Lithographic drivers for key technologies at each node are indicated in the ORTC tables. Year of first mixed-signal product at the same technology may lag by as much as one generation.*

[2] *Nominal supply voltage, V_{dd}, from high-performance (HP) roadmap. Repeated here for reference with e.g., RF CMOS data.*

[3] *Analog supply voltage is expected to lag digital by at least two or more generations. Additional voltage headroom is needed to avoid excessive power dissipation under reduced signal swing conditions. Analog CMOS designs may use thick gate oxide and low V_t techniques. The analog power supply reduction trend may lag digital backward compatibility trend for I/O such that a common thick gate oxide solution is not feasible.*

[4] *Higher frequency bands will drive performance requirements, but applications at the lower frequencies will continue throughout the duration of the roadmap (lower frequency bands are not "obsoleted" and will continue to be used).*

[5] *General purpose analog frequency (such as DSP, audio/video, ADC) trend is in parallel with RF. Very clean, jitter-free, clock frequency generation is required.*

[6] *Absolute collector current, Ic, at which the key bipolar RF transistor parameters of the section are required. Its scaling, reflecting the trend towards lower power, is primarily due to area reduction at constant current density. Determines the color coding in this bipolar RF device section, reflecting the expected difficulty in obtaining the key parameters at these low currents.*

[7] *Assumes a "rule of thumb" of approximately 5–10× the transmit/receive frequency. This is an extrinsic device parameter that depends on parasitics and lay-out.*

[8] *Assumes that ft generally follows the progression of fmax. This is an intrinsic device parameter, together with all other key parameters optimized to fit the application.*

[9] *Measure for the voltage gain of bipolar RF transistors with minimum emitter area: $Gm \sim q\, I_c / kT$ and $1/R_{on} = Gce \sim I_c / V_{Early}$, so to first order, Gm/Gce is equal to the ratio of $V_{Early}/(kT/q)$, and not on the collector current specified in item [6]. Can be traded off against ft/fmax. SiGe technology assumed for all nodes. To be compared with Gm/Gds at minimum gate length for CMOS, items [17] and [18], which are significantly less.*

[10] *1/f noise spectral density, at a frequency of 1 Hz, normalized to an active emitter area of 1 μm^2. The 1/f noise is less prominent in vertical (bipolar) or sub-surface devices (JFETs) than in baseline CMOS which has carrier recombination in traps at the insulator-semiconductor interface.*

[11] *See [6]. Determines the color coding in this bipolar analog device section, reflecting the expected difficulty in obtaining the key parameters at these low currents.*

[12] *See [10]*

[13] *Current matching specified for an active emitter area of 20μm^2. Assumes "near neighbor" devices at minimum practical separation. Careful layout and photolithographic uniformity, e.g. by using dummy structures, is imperative. In contrast to CMOS it does not scale with technology.*

[14] *SiO_2 equivalent physical CMOS gate dielectric thickness associated with the digital high-performance (HP) roadmap. Determines the color coding in this NMOS RF device section, since its realization is crucial for obtaining the key parameters.*

[15] *Maximum oscillation frequency of digital NMOS transistor. Related to its transit frequency [16]. Determines maximum speed for signal processing.*

[16] *Highest transit frequency of digital NMOS transistor. Determined by the transit time of carriers between source and drain and therefore by the physical gate length.*

[17] *Measure for the voltage amplification of a minimum length digital CMOS transistor, important in most mixed-signal designs. Operation point taken at 100mV above the threshold voltage V_t. Can be compared to the bipolar equivalent in item [9].*

[18] *Measure for the amplification of a 10× minimum length digital CMOS transistor. Using different lengths is an extra degree of freedom in mixed signal designs. Long devices have better Gds amplification (at low frequencies). Operation point taken at 100mV above the threshold voltage V_t. Can be compared to the bipolar equivalent in item [9].*

[19] *1/f spectral density of digital NMOS device at a frequency of 1Hz, critical for the minimum input signal in mixed-signal circuits (lower boundary for dynamic range) in low frequency circuits, mixers and VCOs. The 1/f noise is more prominent in MOS devices than in vertical (bipolar) or sub-surface devices (JFETs) due to carrier recombination in traps at the insulator-semiconductor interface (surface effect). Operation point at 100mV above the threshold voltage V_t.*

[20] *Matching specification for the NMOS transistor's threshold voltage, assuming "near neighbor" devices at minimum practical separation. Careful layout and photolithographic uniformity, e.g., by using dummy structures, are required. Optimum situation obtained when scaling as gate oxide thickness: $\Delta V_t = T_{ox}$ [nm] mV·µm, i.e., 5mV·µm for T_{ox} = 5nm. Statistical dopant fluctuations start limiting further improvement, and SiO_2 matching behavior of new high κ gate dielectrics very insecure. Lower boundary for size of transistor in a mixed-signal circuit for a given accuracy and therefore often also for DC power consumption and speed. Also reliability related.*

[21] *SiO_2 equivalent physical CMOS gate dielectric thickness associated with the analog supply voltage roadmap in item (3). Lags two generations or more behind digital, see item [14].*

[22] *Together with the analog supply voltage, item [3], the threshold voltage V_t determines the maximum analog signal (upper dynamic range) the analog circuit is able to handle.*

[23] *Measure for the amplification of an analog NMOS transistor with 10× minimum physical gate length of the technology node (minimum physical gate length not practical for analog applications). Using different lengths is an extra degree of freedom in mixed signal designs. Long devices have better Gds amplification (at low frequencies). Determines color coding of this analog NMOS device section, reflecting the expected difficulty in obtaining the key parameters with devices optimized for lower supply voltages.*

[24] *See [19]*

[25] *See [20]*

[26] *Capacitors are mostly needed for weighting and comparing different analog signals. As digital content increases, chip size decreases, and capacitors occupy a larger percentage of the chip. Choice of implementation is driven by complexity/chip size tradeoff (cost). MicroElectro-Mechanical Systems (MEMS) implementation for filter applications may be favorable at densities ≥7 fF/μm^2. Emerging approach is integration on separate substrate in low-cost technology. The coloring of this and all the analog capacitor rows (through Row 30) is from the Design TWG.*

[27] *Q is a measure for parasitic effects, the distributed resistance and capacitance to the substrate. Q is important for this capacitor to provide bypass response at high frequency. Large parallel, parasitic implementations are expected to dominate for the next years.*

[28] *The capacitors' second order voltage linearity is critical for the dynamic range of analog circuits. The first order component can be cancelled out by differential techniques.*

[29] *Leakage is driven by feedback capacitor applications, where a long time constant is required, and low frequency switched capacitor applications. Requirement is relaxed with increasing analog clock frequency. Highest quality dielectric is suggested.*

[30] *Lower boundary for size of capacitor for a given accuracy. Also reliability related.*

[31] *As digital filter solutions are expected to dominate beyond 2003, bypass capacitor applications will drive density. High κ dielectrics may be cost-effective as chip size decreases in the 5–15 year timeframe. Density is driven by low frequency requirements and chip size considerations. The coloring of this and all the RF bypass capacitor rows (through Row 33) is from the Design TWG.*

[32] *Q is important for this capacitor to provide bypass response at high frequency. Large parallel, parasitic implementations are expected to dominate for the next five years.*

[33] *First order voltage linearity is driven by this type of capacitor.*

[34] *Square resistance at which the following properties, [35-38] are defined. Resistors are needed, e.g., for stable voltage references. The square resistance defines the area, and properties such as voltage linearity, for a given absolute resistance value. A 10–100-1000 Ω/ range is desirable. The coloring of this and all the Resistor rows (through Row 38) is from the Design TWG.*

[35] *Q is a measure for the parasitic effects in an analog resistor.*

[36] *Temperature coefficient is important for the analog resistor, e.g., in references. Low TC films or TC canceling techniques may be utilized*

[37] *Matching is important for the analog resistor e.g., in references. Careful layout and photolithographic uniformity, e.g. using dummy structures, are required. Minimum dimensions assumed to be larger than minimum technology dimensions.*

[38] *1/f noise is critical for the minimum input signal in mixed-signal designs (lower boundary of dynamic range), such as low frequency circuits, mixers and VCOs. It is assumed that low 1/f solutions other than polysilicon coincide with high-Q RF-resistor solutions.*

[39] *Inductors are needed in RF-circuits to realize impedance-matching of the circuit and the transmission line. Since they can generally not be placed above active circuitry, their area is relatively expensive and density in nH/um^2 very important. Emerging approach is integration on separate substrate in low-cost technology. The coloring of this and all the Inductor rows (through Row 40) is from the Design TWG.*

[40] *Q_{3dB}, the quality factor at half bandwidth, is a measure for parasitical effects, mainly affected by distributed resistance and capacitance to substrate. Typical on-chip inductances range between 1–10 nH. Low series R implementations, such as with thick copper, assumed. The increased substrate doping trend in CMOS conflicts with the integration of inductors. Color coding indicative of difficulty in obtaining values >20 on non-isolating substrates.*

[41] *Signal isolation is added here for reference purposes only. It is defined as the transmission efficiency (S21 in dB) between a noise source and a noise sensor in a given frequency range. Its target value depends on the type of circuitry, increasing from plain CMOS, via analog to RF. Values quoted here are for the intermediate, i.e. analog, case. Only with the right combination of technology solutions such as substrate resistivity (high Ohmic, SOI) and metal interconnect scheme (low-κ), design choices like system partitioning, metal spacings and decoupling capacitances, and optimized package design, can the specified values be obtained.*

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 15 years in spite of the reliability risk inherent in massive technology changes. 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.

In addition, as System-on-a-Chip evolves to integrating more and more new technologies (such as MEMS, optoelectronics) on a single chip, there is a need to manage not only the reliability of these new technologies but also any reliability interactions between the various technologies. Defect screening may be required to meet the reliability requirements, especially during the first year of production when yields are ramping up. Finally, failure analysis cycle time improvements are required to support rapid yield learning and design debugging.

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, building-in-reliability, reliability qualification and defect screening. There are some significant gaps in these capabilities today. Furthermore, these gaps will become even larger with the introduction of new materials and new device structures. Inadequate reliability tools lead to unnecessary performance penalties and/or unnecessary risks.

Table 40a Reliability Technology Requirements—Near-term

YEAR OF PRODUCTION	2001	2002	2003	2004	2005	2006	2007	DRIVER
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65	
MPU / ASIC ½ Pitch (nm)	150	130	107	90	80	70	65	
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35	
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25	
Customer reliability expectations*								
Early Failures (ppm) (First 4000 operating hours)** [1], [3]	50–2000	50–2000	50–2000	50–2000	50–2000	50–2000	50–2000	CUSTOMER NEEDS; NEW MATERIALS
Long term reliability (FITS = failures in 1E9 hours) [2, [3]]	10–100	10–100	10–100	10–100	10–100	10–100	10–100	CUSTOMER NEEDS; NEW MATERIALS
Soft Error Rate (FITs) [4]	1000	1000	1000	1000	1000	1000	1000	SCALING
Relative Failure Rate per Transistor (normalized to 130nm) [3],[5]	1	0.7	0.5	0.35	0.25	0.18	0.13	NUMBER OF TRANSISTORS
Relative Failure Rate per m of interconnect (normalized to 130nm node)[6]	1	0.82	0.74	0.66	0.55	0.5	0.45	CUSTOMER NEEDS; J11LENGTH OF INTERCONNECT
ESD protection voltage (V/μm) [7]	10.5	10.5	12	12	13	13.5	14	SCALING
ESD protection circuit effectiveness (V/μm^2) [8]	3	3	3.5–4.0	3.5–4.0	4.0–4.5	4.0–4.5	4.5–5.0	SCALING

White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

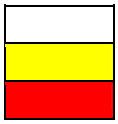
Table 40b Reliability Technology Requirements—Long-term

YEAR OF PRODUCTION	2010	2013	2016	DRIVER
DRAM ½ PITCH (nm)	45	32	22	
MPU / ASIC ½ PITCH (nm)	50	35	25	
MPU PRINTED GATE LENGTH (nm)	25	18	13	
MPU PHYSICAL GATE LENGTH (nm)	18	13	9	
Early Failures (ppm) (First 4000 operating hours)** [1], [3]	50–2000	50–2000	50–2000	CUSTOMER NEEDS; NEW MATERIALS
Long term reliability (FITS = failures in 1E9 hours) [2], [3]	10–100	10–100	10–100	CUSTOMER NEEDS; NEW MATERIALS
Soft Error Rate (FITs) [4]	1000	1000	1000	SCALING
Relative Failure Rate per Transistor (normalized to 130nm) [3], [5]	0.04	0.02	0.006	NUMBER OF TRANSISTORS
Relative Failure Rate per m of interconnect (normalized to 130nm) [6]	0.39	0.25	0.18	CUSTOMER NEEDS; J11LENGTH OF INTERCONNECT
ESD protection voltage (V/μm) [7]	15	17.5	20	SCALING
ESD protection circuit effectiveness (V/μm^2) [8]	5.5–6.0	7.5–10	9.5–14	SCALING

White–Manufacturable Solutions Exist, and Are Being Optimized

Yellow–Manufacturable Solutions are Known

Red–Manufacturable Solutions are NOT Known



Notes for Tables 40a and b

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

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

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

[3] Driven red by need for low-power logic to have an alternate gate dielectric in 2005.

[4] Soft error rate refers to data upset caused by high energy radiation from cosmic rays, radioactive decay, etc

[5] While the overall IC failure rate does not change with time, as the number of transistors increase, the relative failure rate per transistor must decrease

[6] As the length of interconnect increases, the failure rate per m of interconnect must decrease. Even more important from a reliability viewpoint is to reduce the failure rate per via/contact

[7] ESD V/μm in terms of NMOS capability per micron width. Particularly critical in NMOS ESD circuits and self-protected circuits.

[8] ESD V/μm^2 in terms of protection circuit effectiveness per unit area of the protection circuit including guard ring.

POTENTIAL SOLUTIONS

MEMORY AND LOGIC POTENTIAL SOLUTIONS

In order to scale the basic MOSFET structure significantly below the 90 nm node, 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 and to limit drain-induced barrier-lowering-induced off-state current. However, continued thinning of conventional gate oxides results in a significant increase in gate leakage current due to the onset of tunneling current. In addition, the effectiveness of continued EOT reduction begins to become limited due to the non-scalability of gate electrode depletion and quantum-mechanical effects in the MOSFET inversion layer.

To address these scaling problems, existing silicon oxide and nitride based gate dielectrics need to be eventually replaced by alternate high κ materials, which should exhibit significantly lower gate leakage for the same EOT value. In addition, the existing silicided polysilicon gate electrode needs to be replaced by complementary metal-gate electrodes, which should exhibit significantly reduced gate depletion effects as well as lower gate electrode resistance. Refer to Figure 21.

Based on the device scaling scenarios in the technology requirement tables, low-power devices (specifically, the LSTP) will be the driving force behind the need for high- κ gate dielectrics (sometime around the 2005 timeframe). Although high-performance device scaling might be possible without utilizing these materials, high κ gate dielectrics will likely be used for high-performance and LOP devices sometime after they become available for LSTP devices. For both high-performance and low-power devices, complementary metal-gate electrodes are projected to be utilized around the 2007 time-frame.

Continued MOSFET scaling below the 90 nm node will also require technical innovations and advancement involving the device source/drain junctions (including the source/drain halo, extensions, deep junctions, and silicides). In order to control short-channel effects and sub-threshold off-state current, the junction depth needs to be continuously reduced. The parasitic resistance of the source/drain junctions also needs to be controlled and minimized. To address this scaling problem, technological advances in ultra-shallow and ultra-low resistance junction technology are required; alternatively, non-classical CMOS device structures (such as vertically-configured devices) may be required (For more detail, *see the Emerging Research Devices section, Non-classical CMOS*).

At each technology node, the device channel doping, source/drain junction halo and extension doping profiles need to be optimized to allow maximum saturation current-drive within a specified off-state current limit. However, for continued MOSFET scaling to continue significantly below the 90 nm node, multiple threshold-voltage and multiple-oxide-thickness devices also need to be available in order to enable more comprehensive circuit/system-level power/performance optimization. Dynamic (electrically adjustable) threshold-voltage devices will likely be required to enable an even greater level of power/performance optimization. No single device type (either high-performance or low-power) will be able to meet simultaneously all of the differing performance and power-dissipation requirements; the availability of a variety of different device types will allow maximum flexibility to meet future circuit/system needs.

For the long-term technology nodes (2010 and beyond), in order to maintain the historical rate of device performance improvement (approximately 17% per year), fundamental technological improvement in the MOSFET mobility/transconductance is required. The performance improvement through gate length, gate dielectric, and power supply scaling will no longer be sufficient. New high-mobility materials (such as strained SiGe layers) will be needed. These materials will be required first by the high-performance device, then eventually be the low-power devices.

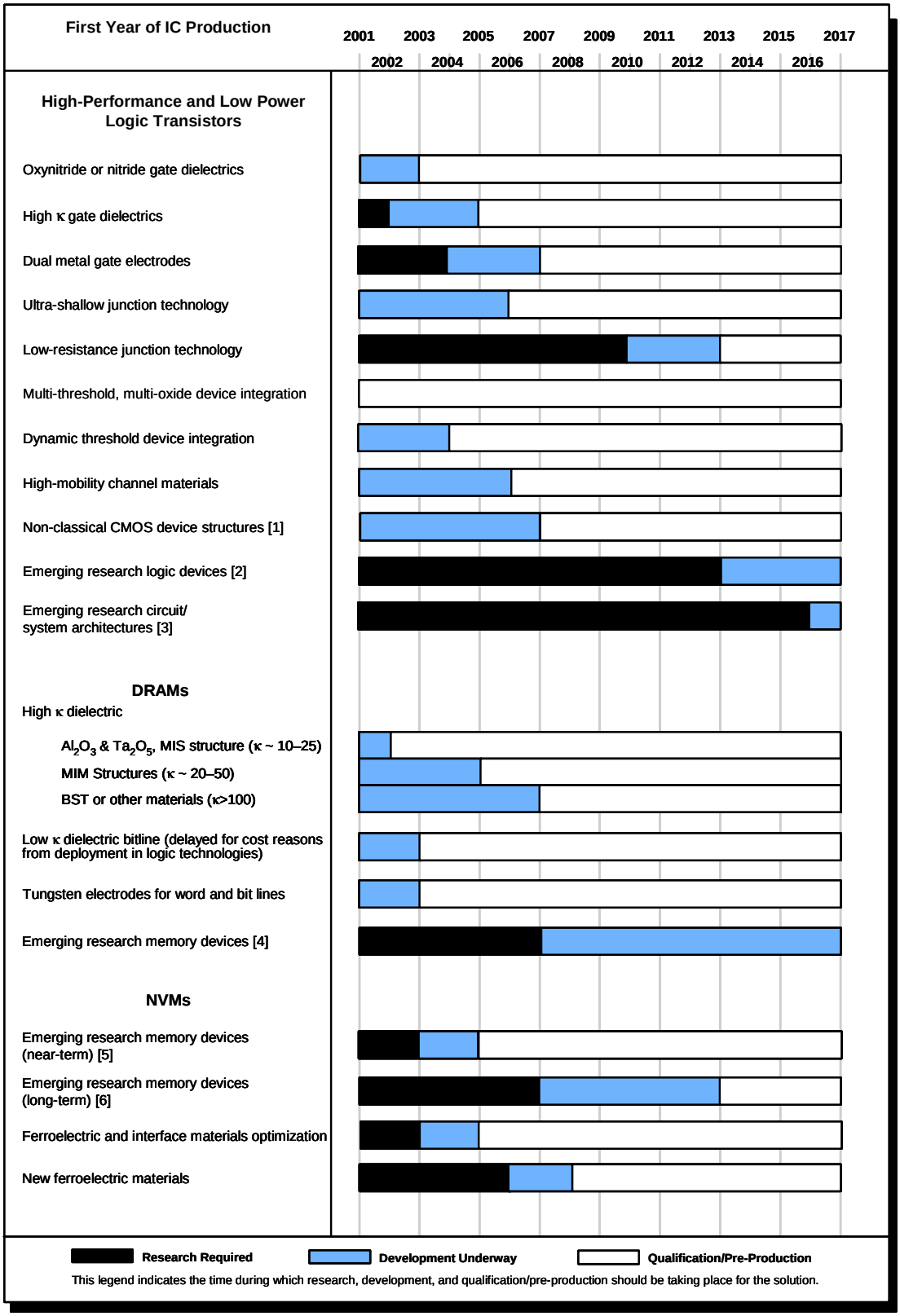
Ultimately, for the very long-term technology nodes, more radical alternatives to MOSFET devices need to be considered to carry out the basic device logic and memory functions. In addition, fundamental changes may be required in overall circuit design and architecture. The *Emerging Research Devices sections* describes these options in greater detail.

Scaling of flash technology options involves the optimization of a tunnel dielectric, floating gate, interpoly dielectric stack to support charge transfer through the dielectrics in erase and write modes while suppressing charge leakage in other modes. For each technology node the problem of integration of the necessarily high voltage erase and write modes with a much lower voltage underlying CMOS technology becomes more serious. The intractability of scaling has led to more active research into alternative forms of nonvolatile memory. Some of these approaches are identified in the *Emerging Research Devices Memory table*.

Scaling of FeRAM depends on the geometric arrangement of the cell, and more critically on the characteristics of the ferroelectric and interface materials. Cell geometry has been visualized as rapidly moving from planar structures to more efficient 3 dimensional arrangements. To facilitate these changes it is necessary to modify and optimize the PZT or SBT materials and related interface materials in such a way as to maintain process compatibility and operation at lower voltages. It is also recognized that in the longer term new ferroelectric materials will be required for lower voltage operation and for compatibility with processing 3 dimensional structures.

For DRAMs, improvements in density, cost, speed, and noise immunity with scaling are critical. High κ dielectrics are expected to be employed to increase the capacitance per unit area of the storage cell. Al_2O_3 or Ta_2O_5 is expected to be the first such material deployed into production DRAMs, using metal-insulator-semiconductor (MIS) structures. Later, both metal-insulator-metal structures and other, higher κ materials may be deployed, perhaps together, to increase the capacitance per unit area. Finally, materials with $\kappa > 100$ will need to be deployed. Barium strontium titanate (BST) is such a material, but it has serious problems that make it doubtful whether BST can be effectively utilized as the storage capacitor dielectric. The most serious problem is that it is very difficult to scale BSTs thickness into the desired range without incurring high leakage.

Another potential solution for DRAMs is the use of low κ dielectrics similar to those in logic IC interconnects to lower the bitline capacitance and hence to improve the sensing noise margin. However, due to cost considerations, the deployment of low κ dielectrics for DRAM chips will be delayed by several years from their deployment for logic chips. Eventually, toward the end of the roadmap and beyond, more exotic memory devices may be utilized (see *Emerging Research Devices Memory table* for more detail).



Emerging Research Devices sections: *[1] Non-classical CMOS* *[2] Logic Devices* *[3] Architectures* *[4], [5], [6] Memory Devices*

Figure 21 Memory and Logic Potential Solutions

MIXED-SIGNAL DEVICE POTENTIAL SOLUTIONS

In this section, the potential solutions to the challenges in mixed-signal are discussed (Refer to Figure 22). The solutions here are different from, or in addition to, those already reviewed for memory and logic. Successful mixed-signal technologies will leverage the baseline digital platform while integrating value-added features and functions. Key ingredients to successful mixed-signal integration are the addition of precision high-Q passive elements, adequate signal isolation, and compatible active devices.

With the steady improvement in high-frequency performance and decrease in speed-power product of CMOS, this technology will gradually gain on traditional BiCMOS and bipolar implementations. The onset of this trend was visible at the 180nm node, with the integration of baseband and IF functions. Integration of RF and IF in CMOS will follow in wireless personal network ICs. Portable wireless products will continue to drive low operating and stand-by power technologies and design architectures (for long battery lifetime). SOI based processes will be one key solution in obtaining high signal isolations and low parasitic capacitance values.

The mixed-signal supply voltage continues to lag that of high-performance digital by two or more generations. A combination of multiple gate oxide thickness, multiple thresholds, and DC-DC conversion is needed to support the ever increased mixed-signal requirements. Solutions in active threshold regulation, substrate biasing, and novel design architecture will be required to extend the trend for lower supply voltages for mixed-signal applications. An alternative to full integration is the use of Multi-Chip-Modules (MCM) which combines circuits on different technologies and is optimized for the desired functions. Ultimately, full-digital implementations in CMOS will replace most analog designs.

The 1/f noise is less prominent in vertical (bipolar) or sub-surface devices (JFETs) than in baseline CMOS which has carrier traps at the insulator-semiconductor interface. In CMOS, however, 1/f noise decreases with each new technology node, although the future introduction of high κ gate dielectrics represents an uncertainty here.

Matching requirements have become as stringent for mixed-signal as for some digital designs (SRAM, clock trees). Precise active and passive device matching will be achieved through careful layout and low thermal budget processes. Device matching is critical for differential circuit design and active circuit compensation.

The trend of moving discrete passive elements from board-level to chip-level will continue. Solutions for achieving discrete-equivalent precision on-chip passive components are expected. Very low resistance metal interconnects (thick) are required for high-Q integrated passives. New high κ dielectrics are needed to reduce integrated capacitor area. Alternatively, some passives may be integrated into the printed board or package as a method of cost reduction/simplification. The need for low-loss, high-Q inductors may justify use of 3D or micromachined structures.

As the integration density increases and the operation frequency rises, protection of noise sensitive analog circuits from "noisy" digital circuits will become increasingly difficult. Signal isolation is managed through a combination of substrate (e.g., SOI), interconnect, and package solutions. Today, circuit blocks are protected by oxide isolation, guard rings and buried wells (triple well). Integrated shielding structures are required for protection of circuits and interconnect. Innovative optical and micromachining techniques may be employed as solutions in the future. Novel device structures and design architectures may be employed to enhance circuit signal/noise performance. It is required that any cost-effective solution addressing these problems and challenges must be compatible with the mainstream CMOS technology of the time.

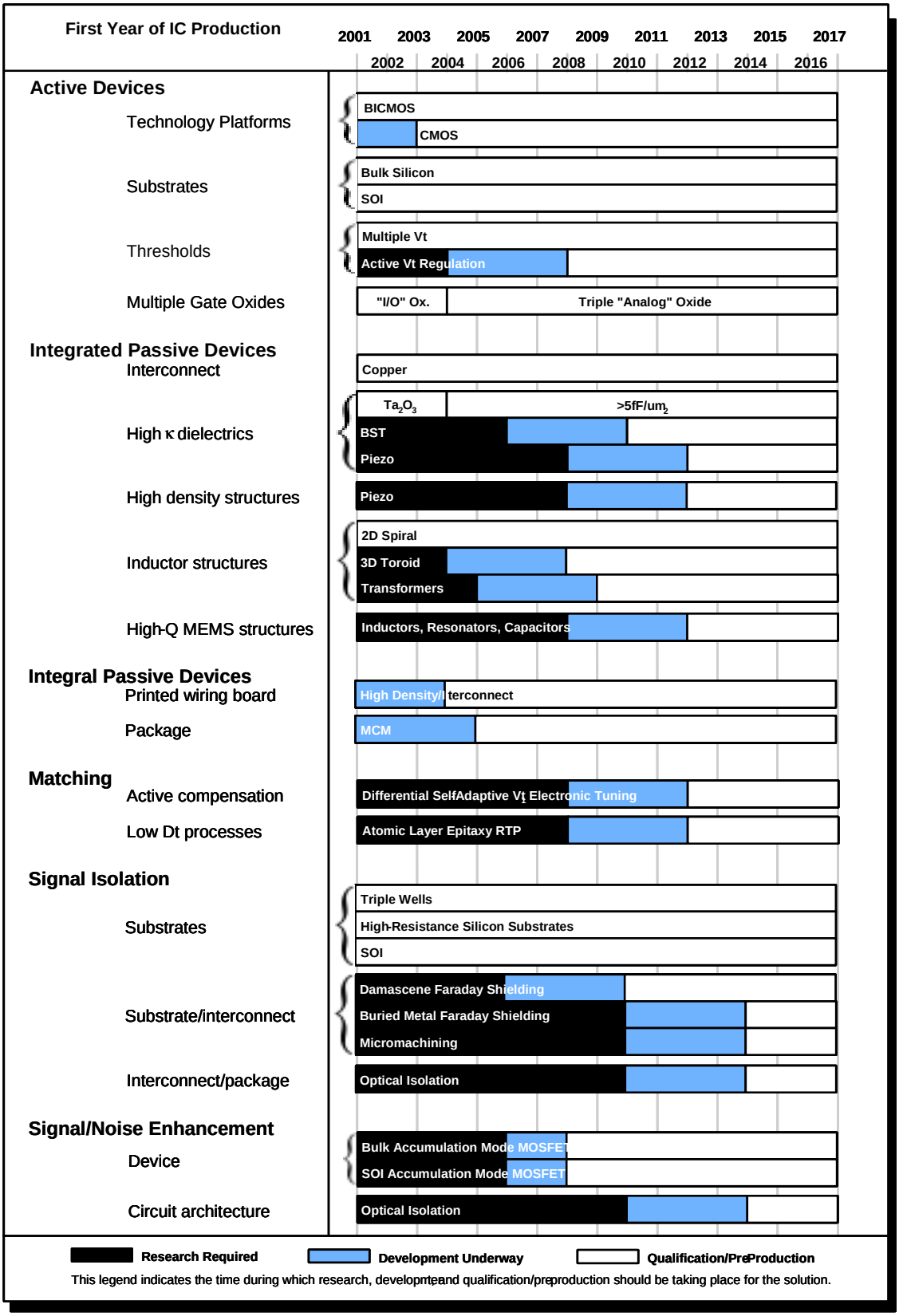


Figure 22 Mixed-signal Device Potential Solutions

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 will 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 23 cover the major technical discontinuities over the lifetime of the Roadmap. Because these are major discontinuities with serious reliability issues it takes several years to conduct the R&D to identify and model the failure modes (black bars), turn these results into practical reliability engineering capabilities (blue bars), and, finally to perform the qualification of a new technology node (white bars). 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 23. 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.

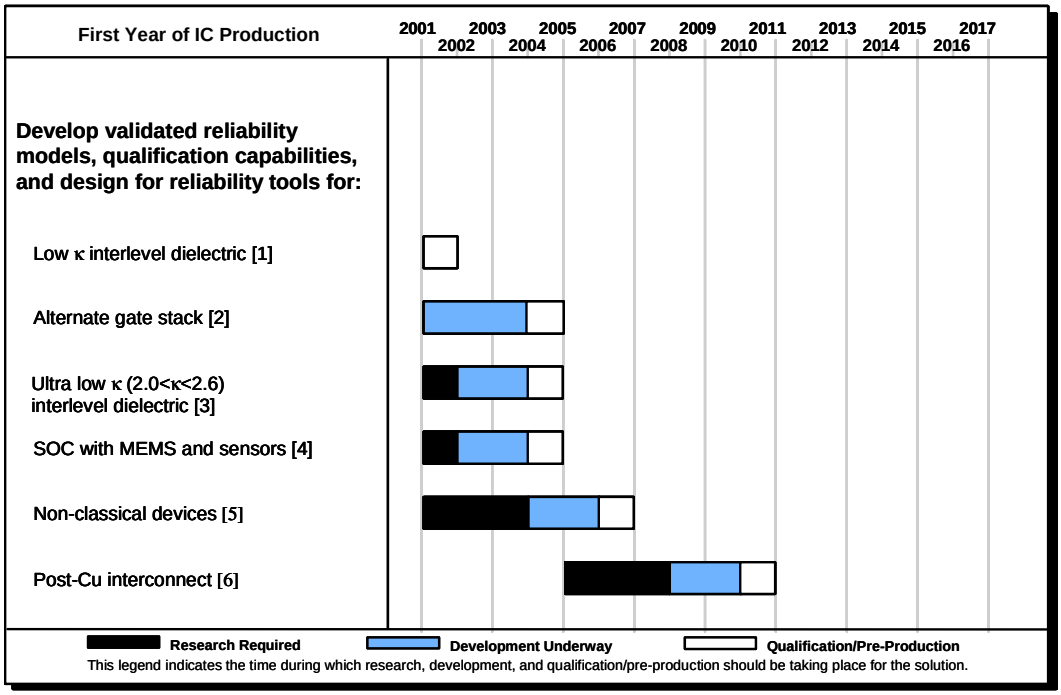


Figure 23 Reliability Potential Solutions

Notes for Figure 23:

Model need to be able to predict failure rates for each significant failure mode as a function of operating conditions

Dates are driven by when new materials/devices will be introduced. Sufficient resources to meet these dates may not be available.

There are several possible variants in each of these categories. Models are needed for each variant going into production.

Less disruptive changes can be qualified in less time

[1] Assumes Low κ enters volume production in 2002 (from Interconnect chapter, Dielectric Potential Solutions)

[2] Assumes alternate gate enters volume production in 2005 (see LSTP Technology Requirements Table)

[3] Assumes ultra low κ enters volume production in 2005 (driven by Interconnect chapter, Dielectric Potential Solutions)

[4] Assumes enters volume production in 2005

[5] Assumes enters volume production in 2007 (driven by high power logic, refer to the Memory and Logic Potential Solutions).

[6] Assumes enters volume production in 2012 (post Cu interconnect used only for global Interconnect, refer to the Interconnect chapter, Conductor Potential Solutions)

EMERGING RESEARCH DEVICES

SCOPE

The primary goal of this new section on Emerging Research Devices is to stimulate invention and research leading to feasibility demonstration for one or more Roadmap-extending concepts. In addressing this goal, this section serves two purposes. First is to introduce advanced non-classical CMOS structures and memory technologies aimed at extending microelectronic technologies to the end of this Roadmap timeframe. The second is to introduce (without endorsement) completely new technological and architectural concepts for information and signal processing beyond the end of the Roadmap timeframe.

The quickening pace of MOSFET scaling is accelerating introduction of new technologies to extend bulk CMOS beyond the 65 nm node. These technologies include both new materials and advanced MOSFET structures. The *Front End Processes chapter* discusses new materials required, for example, for the gate stack and for source/drain contacts. The *Process Integration, Devices and Structures section identifies technology requirements* for CMOS structures. This new Emerging Research Devices section serves as the bridge between bulk CMOS and the realm of microelectronics beyond the end of the Roadmap.

This discussion is divided into four categories: Non-classical CMOS, Memory Devices, Logic Devices and information processing Architectures. The discussion of Non-classical CMOS and Memory Devices, given their shorter time frame, provide some detail regarding their operation principles, advantages, disadvantages, and maturity. The subsection on Memory Devices also provides some estimates of performance upon their introduction to manufacturing and a few speculative estimates for ultimate performance. The discussions of Logic Devices and Architectures, given their "beyond-the-Roadmap" time frame, focus on principles of operation, major advantages and challenges, and their maturity or state of exploration and development. This section on Emerging Research Devices ends with a preliminary but interesting comparison of the performance projections and cost attributes for several speculative new approaches to information and signal processing. An interesting observation of this comparison is that the emerging devices, technologies and architectures, given their successful development, would extend applications of microelectronics to domains not accessible to CMOS, rather than competing directly with CMOS in the same domain.

Accepting the risk of including ideas that eventually will be shown as non-functional or impractical, the intent of this Section is to "cast a broad net" to gather in one place substantive, alternative concepts for memory, logic and information processing architectures that would, if successful, substantially extend the Roadmap beyond CMOS. As such, this section will provide a window into candidate approaches. Regarding inclusion of concepts in this section, the idea is to identify, but not endorse; to include but not to validate. Concepts included in this section are illustrative, but perhaps not comprehensive. In summary, including a particular concept in this section does not in any way constitute advocacy or endorsement. Rather, inclusion does point out that existing research efforts are exploring a variety of basic technology and architectural concepts for information and signal processing.

DIFFICULT CHALLENGES

Table 41 Emerging Research Devices Difficult Challenges

DIFFICULT CHALLENGES ≥ 65 nm, THROUGH 2007	*SUMMARY OF ISSUES*
Implementation into manufacturing of non-classical (non-bulk) MOSFET devices (for example, dual-gate SOI).	Select most promising choice of device structure Timely development, process integration, and qualification.
DIFFICULT CHALLENGES < 65 nm, BEYOND 2007	
Toward the end of the Roadmap or beyond, implementation of novel, non-CMOS devices and architectures, including interconnect and memory.	Will drive major changes in process, materials, physics, design, etc. Novel devices may coexist with CMOS: integration of the two.

DIFFICULT CHALLENGES

The microelectronics community is facing two major difficult challenges related to scaling microelectronics to and beyond the end of the *2001 ITRS* timeframe. The near-term challenge is to develop a manufacturable non-classical MOSFET structure. The longer-term challenge is invention of a feasible new information and signal processing technology addressing "beyond CMOS" applications. Solution to the first may be critically important to extension of CMOS to and beyond the 65 nm node, and solution to the latter could provide many new opportunities for microelectronics.

EMERGING TECHNOLOGY SEQUENCE

Figure 24 shows an overview of the organization of the Emerging Research Devices section and illustrates the relationship of particular new concepts to the four functional categories that they each address—Non-classical CMOS, Memory, Logic, and Architectures. A category for Architectures is included to emphasize the point that because new systems architectures and new device technologies each will drive development of the other, synergistic/collaborative development of the two together can be very rewarding. This figure illustrates one simplified example of a richly diverse set of emerging application specific concepts and technologies addressing different functions, and it includes several highly speculative approaches. Many of these concepts likely will not mature to manufacturing or application. The important message here is that the emergence of many new ideas and technologies, several of which are suitable for only certain function(s) and do not have broad application, may be signaling a coming dispersion of microelectronics technologies to address an increasingly diverse set of market-driven applications. Integration of Systems on Chip (SoC) and in a package at low cost and within a prescribed form factor will undoubtedly continue use of CMOS as the functional integration platform. This further illustrates the need to integrate dissimilar technologies and functions in a high-performance, low-cost fashion with CMOS platforms.

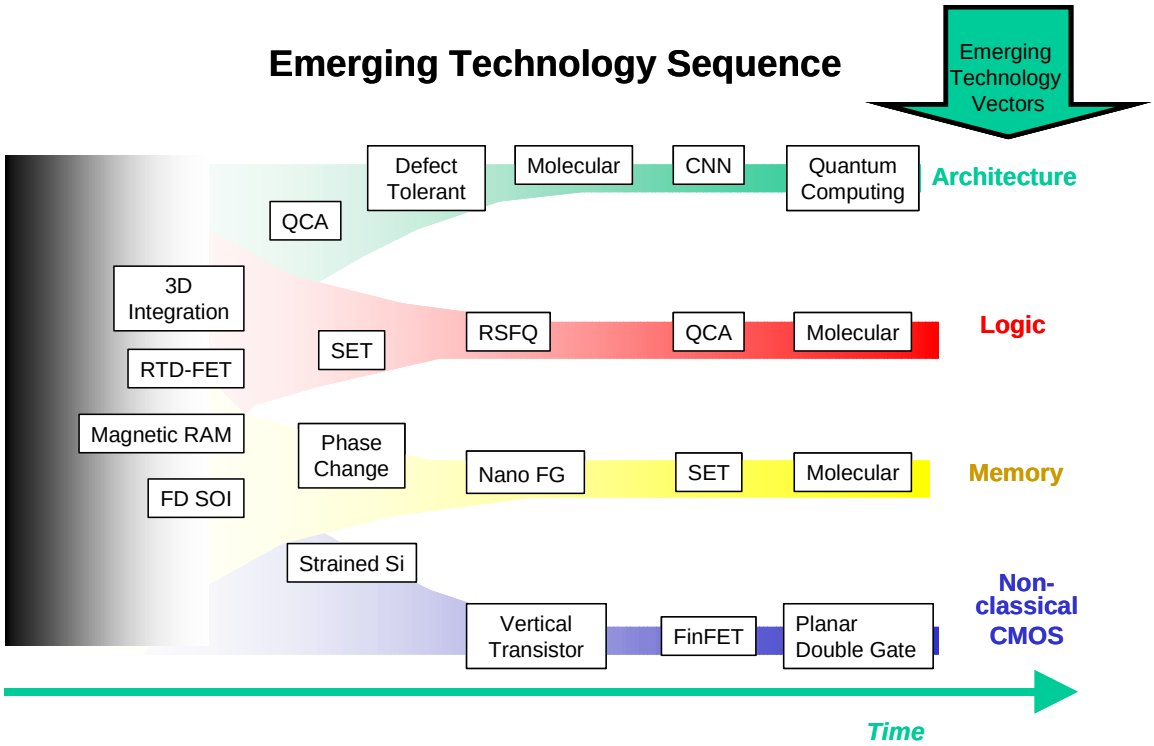


Figure 24 Emerging Technology Sequence

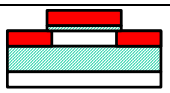
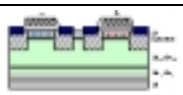
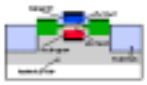
EMERGING RESEARCH TECHNOLOGIES

NON-CLASSICAL CMOS

INTRODUCTION

Non-classical CMOS includes those advanced MOSFETs that provide a path to scaling CMOS to the end of the Roadmap using new transistor structural designs. Scaling MOSFETs to and below the 65 nm node exacerbates several well-known, emerging challenges. For digital applications, these challenges include exponentially increasing leakage currents (gate, channel, source/drain junctions), short channel effects controlling V_t, continuing to increase I_{on}, and control of V_t over the die. For analog/RF applications, the challenges additionally include sustaining linearity, low noise figure, power-added-efficiency, and transistor matching. The industry is pursuing two fundamentally distinct approaches to managing these scaling challenges—bulk transistors enhanced using new materials for the gate stack, etc., and new transistor structures. The first approach is discussed in the *Front End Processes* chapter and the latter approach is discussed in this part of the *Process Integration* chapter. This subsection's entries include ultra-thin body SOI, band-engineered transistor, and three entries for double gate structures (vertical transistor, FinFET or delta transistor and the planar double-gate transistor). Refer to Table 42.

Table 42 Non-Classical CMOS

					
DEVICE	*ULTRA-THIN BODY SOI*	*BAND-ENGINEERED TRANSISTOR*	*VERTICAL TRANSISTOR*	*FINFET*	*DOUBLE-GATE TRANSISTOR*
CONCEPT	**Fully depleted SOI**	**SiGe or Strained Si channel; bulk Si or SOI**	**Double-gate or surround-gate structure (No specific temporal sequence for these three structures is intended)**		
APPLICATION/DRIVER	**Higher performance, Higher transistor density, Lower power dissipation**				
ADVANTAGES	**-Improved subthreshold slope -V_t controllability**	**-Higher drive current -Compatible with bulk and SOI CMOS**	**-Higher drive current -Lithography independent L_g**	**-Higher drive current -Improved subthreshold slope -Improved short channel effect -Stacked NAND**	**-Higher drive current -Improved subthreshold slope -Improved short channel effect -Stacked NAND**
SCALING ISSUES	**-Si film thickness -Gate stack -Worse short channel effect than bulk CMOS**	**-High mobility film thickness, in case of SOI -Gate stack -Integration**	**-Si film thickness -Gate stack -Integrability -Process complexity -Accurate TCAD including QM effect**	**-Si film thickness -Gate stack -Process complexity -Accurate TCAD including QM effect**	**-Gate alignment -Si film thickness -Gate stack -Integrability -Process complexity -Accurate TCAD including QM effect**
DESIGN CHALLENGES	**-Device characterization -Compact model and parameter extraction**	**-Device characterization**	**-Device characterization -PD versus FD -Compact model and parameter extraction -Applicability to mixed signal applications**		
MATURITY	**Development**				
TIMING	**Near Future** ⟶				

NON-CLASSICAL CMOS – DEFINITION AND DISCUSSION OF TABLE ENTRIES

Ultra-thin body SOI [2,3]—Currently in production, an SOI CMOS transistor is partially depleted when there is a quasi-neutral body or substrate region under all operating bias conditions. This quasi-neutral body is usually left floating with no external electrical connection. Alternatively, the quasi-neutral body is electrically connected to the source or to an externally accessible body contact. In contrast, a fully depleted SOI CMOS transistor has no quasi-neutral body region because the entire body or substrate region is depleted of mobile carriers under all operating bias conditions. Electrical connection to the body is not possible. Ultra-thin body scaling provides the extremely thin channel dimensions (<5 nm) required to scale CMOS to the 22 nm node. Recently a new structure has been reported [3] utilizing a thin Si channel [5–20 nm] isolated from the substrate by a thin localized buried dielectric layer [10–30 nm]. This structure combines the best features of bulk CMOS (e.g., deep source/drain regions) with the best features of SOI (e.g., ultra-thin channel and dielectric insolating layer).

Band-engineered transistors [4, 5, 6]—The concept of a band-engineered transistor is to enhance the mobility of electrons and/or holes in the channel by modifying the band structure of silicon in a way such that the physical structure of the transistor remains substantially unchanged. This enhanced mobility increases the transistor g_m and $I_{on.}$ A Si-Ge layer or a strained-Si on relaxed Si-Ge layer is used as the enhanced-mobility channel layer. The device structure can be a bulk transistor or an SOI transistor.

Vertical transistor [7]—A vertical transistor is one having surface conduction channels on two or more vertical surfaces and having current flow in the vertical direction. The channel length is given by the vertical separation between source and drain, which is usually determined by the thickness of an epitaxial layer and not by a lithographic step.

FinFET [8]—A FinFET is another form of a double gate transistor having surface conduction channels on two opposite vertical surfaces and having current flow in the horizontal direction. The channel length is given by the horizontal separation between source and drain and is usually determined by a lithographic step combined with a side-wall spacer etch process.

Double-gate transistor [9]—A double-gate transistor is one having surface conduction channels on two opposite horizontal surfaces and having current flow in the horizontal direction. The channel length is given by the horizontal separation between source and drain and is defined by a lithographic step combined with an etch process.

---

[2] *S. Cristoloveanu, "SOI Technology: Renaissance or Science Fiction?", in: S. Luryi, J. Xu, and A. Zaslavsky (Eds.), Future Trends in Microelectronics (1999 John Wiley & Sons, Inc), pp. 105-114*

[3] *M. Jurczak, T. Skotnicki, M. Paoli, B. Tormen, J. Martins, J. Regolini, D. Dutartre, P. Ribot, D. Lenoble, R. Pantel and S. Monfray, "Silicon-on-Nothing (SON) – An Innovative Process for Advanced CMOS, IEEE Trans. Elect. Dev. 47, 2179 (2000).*

[4] *R. Hartmann, U. Gennser, H. Sigg, D. Grützmacher, and G. Dehlinger, "Si/SiGeC Heterostructures: A path towards high mobility channels", in: S. Luryi, J. Xu, and A. Zaslavsky (Eds.), Future Trends in Microelectronics (1999 John Wiley & Sons, Inc), pp. 133-141*

[5] *J. Alieu, T. Skotnicki, P. Bouillon, J. L. Regollini, A. Souifi, G. Guillot, and G. Bremond, "Potential of SiGe-Channel MOSFETs for a submicron CMOS technology", in: S. Luryi, J. Xu, and A. Zaslavsky (Eds.), Future Trends in Microelectronics (1999 John Wiley & Sons, Inc), pp. 143-153*

[6] *S. Takagi, T. Mizuno, N. Sugiyama, T. Tezuka, A. Kurobe, "Strained-Si-on-insulator (strained-SOI) MOSFETs – Concept, structures and device characteristics", IEICE Trans. Electronics E84C (2001) 1043-1050*

[7] *S-H. Oh, J. M. Hergenrother, T. Nigam, D. Monroe, F. P. Klemens, A. Kornblit, W. M. Mansfield, M. R. Baker, D. L. Barr, F. H. Baumann, K. J. Bolan, T. Boone, N. A. Ciampa, R. A. Cirelli, D. J. Eaglesham, E. J. Ferry, A. T. Fiory, J. Frackoviak, J. P. Garno, H. J. Gossmann, J. L. Grazul, M. L. Green, S. J. Hillenius, R. W. Johnson, R. C. Keller, C. A. King, R. N. Kleiman, J. T.-C. Lee, J. F. Miner, M. D. Morris, C. S. Rafferty, J. M. Rosamilia, K. Short, T. W. Sorsch, A. G. Timko, G. R. Weber, G. D. Wilk, and J. D. Plummer, "50-nm Vertical Replacement-Gate (VRG) pMOSFETs," IEDM Tech. Dig., p. 65, 2000*

[8] *D. Hisamoto, W. C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T. J. King, J. Bokor, C. M. Hu, "FinFET – A self-aligned double-gate MOSFET scalable to 20 nm," IEEE Trans. Electron. Dev. 47 (2000) 2320-2325*

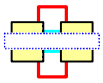
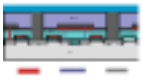
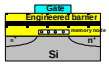
[9] *F. Allibert, T. Ernst, J. Pretet, N. Hefyene, C. Perret, A. Zaslavsky, S. Cristoloveanu, "From SOI materials to innovative devices", Solid State Electron. 45 (2001) 559-566*

MEMORY DEVICES

INTRODUCTION

The memory technologies tabulated below in Table 43 are an incomplete sample of published 2001 research efforts selected to describe some attractive alternative approaches. Historically, very few memory research options yield practical memory devices, and including a particular approach here does not in any way constitute advocacy or endorsement. Conversely, not including a particular concept in this subsection does not in any way constitute rejection of that approach. This listing points out that existing research efforts are exploring a variety of basic memory mechanisms. These mechanisms include charge isolated by surrounding dielectrics, charge held in place by Coulomb blockade potentials, chemical phenomena, magnetic phenomena, and material phase changes. A strong theme is to merge each of these memory options into a CMOS technology platform in a seamless manner. Fabrication is viewed as some type of modification or addition to a CMOS platform technology. A goal is to present the end user with a device that looks like a familiar silicon memory chip. Because all of these approaches attempt to mimic and improve on the capabilities of present day memory technologies, parameters are provided for DRAM and Flash NOR technologies as benchmarks. DRAM and Flash NOR are the current dominant volume produced memories.

Table 43 Emerging Research Memory Devices

STORAGE MECHANISM	*BASELINE 2002 TECHNOLOGIES*		*MAGNETIC RAM*		*PHASE CHANGE MEMORY*	*NANO FLOATING GATE MEMORY*	*SINGLE/FEW ELECTRON MEMORIES*	*MOLECULAR MEMORIES*
								
DEVICE TYPES	*DRAM*	*NOR FLASH*	*PSEUDO-SPIN- VALVE*	*MAGNETIC TUNNEL JUNCTION*	*OUM*	*-ENGINEERED TUNNEL BARRIER -NANOCRYSTAL*	*SET*	*-BISTABLE SWITCH -MOLECULAR NEMS -SPIN BASED MOLECULAR DEVICES*
AVAILABILITY	**2002**		**~2004**	**~2004**	**~2004**	**>2005**	**>2007**	**>2010**
INITIAL F VALUE	**130 nm**	**150 nm**	**350 nm**	**130 nm**	**100 nm**	**80 nm**	**65 nm**	**45 nm**
CELL SIZE	**$8F^2$ 0.14 µm^2 1T**	**$10F^2$ 0.19 µm^2 1T**	**~$40F^2$ 4.9 µm^2 2T**	**20–$40F^2$ 0.68 µm^2 2T**	**$6F^2$ 0.06 µm^2 1T**	**4–$10F^2$ 0.04 µm^2**	**4–$9F^2$ ~0.04 µm^2**	**~$2F^2$ 0.004 µm^2**
ACCESS TIME	**<20 ns**	**~80 ns**	**<25 ns**	**<10 ns**	**<100 ns**	**<10 ns**	**<10 ns**	**~10 ns**
STORE TIME	**<20 ns**	**~1 ms**	**<25 ns**	**<10 ns**	**<100 ns**	**<10 ns**	**<100 ns**	**~10 ns**
RETENTION	**64 ms**	**>10 yrs**	**>10 yrs**	**>10 yrs**	**>10 yrs**	**>10 yrs**	**Seconds to minutes[1]**	**Days**
E/W CYCLES	**Infinite**	**>1E5**	**>1E15**	**>1E13**	**>1E13**	**>1E6**	**>1E9**	**>1E15**
GENERAL ADVANTAGES	**Density Economy**	**Non-volatile**	**Non-volatile, High endurance, Fast read and write, Radiation hard, NDRO**		**Non-volatile, Low power, NDRO, Radiation hard**	**Non-volatile, Fast read and write**	**Density Power**	**Density, Power, Identical Switches, Larger I/O difference, Opportunities for 3D, Easier to interconnect, Defect tolerant circuitry**
CHALLENGES	**Scaling**	**Scaling**	**Integration issues, Material quality, Control magnetic properties for write operations**		**New materials and integration**	**Material Quality**	**Dimensional Control (Room temperature operation), Background Charge**	**Volatile Thermal Stability**
MATURITY	**Production**		**Development**		**Development**	**Demonstrated**	**Demonstrated**	**Demonstrated**

1. Optimistic projection for retention time for Single/Few Electron Memory is > 10 years.

MEMORY DEVICES – DEFINITION AND DISCUSSION OF TABLE ENTRIES

Magnetic RAM (MRAM)[10]—MRAM structures are based on the magnetoresistive effects in magnetic materials and structures that exhibit a resistance change when an external magnetic field is applied. In the MRAM, data are stored by applying magnetic fields that cause magnetic materials to be magnetized into one of two possible magnetic states. Reading data is performed by measuring resistance changes in the cell compared to a reference. Passing currents nearby or through the magnetic structure creates the magnetic fields applied to each cell. Two magnetoresistive effects are used in MRAM—giant magnetoresistance (GMR) and tunneling magnetoresistance (TMR). Correspondingly, two types of MRAM are explored: GMR (or its modified version Pseudo Spin Valve) MRAM and magnetic tunnel junction (MTJ) MRAM. The PSV cell providing GMR is made of two magnetic layers (e.g., Ni, Co, Fe) separated by a thin conductive nonmagnetic layer (e.g., Cu). In PSV MRAM devices the GMR is sensed with the current flowing along the center conducting layer which is coupled to both of the magnetic cladding layers. The resistance to current flow in the thin nonmagnetic conducting layer depends upon the alignment of magnetic spin polarization of each of the two cladding magnetic metal layers with each other and with the electrons in the center conducting layer. If the magnetic spins of the two magnetic cladding layers are aligned with each other, those electrons flowing in the center conducting layer which possess magnetic spins aligned to these layers will experience little large angle scattering and will, therefore, flow with little resistance. If the magnetic spins of the two cladding magnetic layers are not aligned with each other, then the magnetic spins of all electrons in the center nonmagnetic conducting layer will be misaligned with one or both of the cladding layers. Consequently, for this case a fractionally larger resistance will impede current flow in the center nonmagnetic conducting layer. The difference between 0 and 1 levels correspond to about 12% of the cell resistance. Disadvantages of PSV cells are low impedance and small signal voltage difference of ±3 mV during a read operation resulting in larger read time.

The MTJ cell is made of two ferromagnetic layers separated by a thin insulating layer that acts as a tunnel barrier. In contrast to giant magnetoresistive (GMR) structures in which the sense current usually flows parallel to the layers of the structure, the current is passed perpendicular to the layers of the MTJ sandwich. Similar to GMR, the resistance of the MTJ sandwich depends on the magnetic arrangement of the magnetic moments of the two ferromagnetic layers. Typically, the resistance of the MTJ is lowest when these moments are aligned parallel to one another, and is highest when antiparallel, thereby giving rise to magnetoresistance (MR). The read operation is performed by measuring spin dependent tunneling current between the magnetic layers, thus the impedance is high and the difference between 0 and 1 level can achieve 50% of cell resistance. Scalability of both PSV and MTJ memory cells is difficult because of increased write current density for smaller cell size. Numerical data in the Memory Devices, Table 43, correspond to MRAM parameters at the year of introduction.

Phase change memory [11,12]—Phase Change Memory also called Ovonic Unified Memory (OUM) is based on rapid reversible phase change effect in some materials under influence of electric current pulses. The OUM uses the reversible structural phase-change in thin-film material (e.g., chalcogenides) as the data storage mechanism. The small volume of active media acts as a programmable resistor between a high and low resistance with > 40× dynamic range. 1 and 0 are represented by crystalline versus amorphous phase states of active material. Phase states are programmed by the application of a current pulse through a MOSFET that drives the memory cell into a high or low resistance state, depending on current magnitude. Reading data is performed by measuring resistance changes in the cell. OUM cells can be programmed to intermediate resistance values, e.g., for multi-state data storage.

Potential advantage of OUM is a relatively simple system based on "smart" material properties rather than on an elaborate multi-material layered structure. Also, since the energy required for phase transformation decreases with cell size, the write current scales with cell size, thus facilitating memory scaling. Numerical data in the Memory Technologies correspond to OUM parameters at the year of introduction.

Nanofloating gate memory (NFGM) [13, 14, 15, 16, 17, 18, 19]—NFGM includes several possible evolutions of conventional floating gate memory. There are two major approaches proposed to improve performance of floating gate memory cells—

[10] *K. Inomata, "Present and future of magnetic RAM technology", IEICE Trans. Electron. v. E84-C, No6 (2001) 740-746*
[11] *Tyler Lowrey, "Ovonic Unified Memory", http://www.ovonic.com*
[12] *Will Wade and David Lammers, "Intel grooms pair of substitutes for flash", EE Times, July 16, 2001*
[13] *K. K. Likharev 1998, "Layered tunnel barriers for nonvolatile memory devices", Appl. Phys. Lett. 73 (1998) 2137-2139*
[14] *K. Nakazato, K. Itoh, H. Mizuta, and H. Ahmed, "Silicon stacked tunnel transistor for high-speed and high-density random access memory gain cells", Eelctronics Lett. 35 (1999) 848-850*
[15] *S. Tiwari, J. A. Wahl, H. Silva, F. Rana, and J. J. Welser, "A silicon nanocrystals based memory", Appl. Phys. Lett. 68 (1996) 1377-1379*

engineered tunnel barrier[13, 14] and nano-sized memory node.[15–19] Engineered tunnel barrier includes crested barrier floating gate memory[13] and phase-state low-electron-number drive memory (PLED). [14] The crested barrier concept uses a stack of insulating materials to create a special barrier shape enabling effective Fowler-Nordheim tunneling into/out of the storage node. In the PLED memory electrons are injected into the memory node through stacked multiple tunnel junctions with double gate. Engineered tunnel barriers serve to increase the read/write performance of memory cells while keeping long retention time typical for floating gate memories. The approach of NFGM with engineered tunnel barriers is currently in *concept* stage, since no memory operation has been experimentally demonstrated. In the NFGM with nano-sized memory node, the floating gate consists of multiple[15] or single[16] silicon nanocrystal dots. The multiple floating dots are separated and independent, and electrons are injected to the dots via different paths. The endurance problem can be much improved in multidot (nanocrystal) memory.[15] In both multidot and single dot memories the very small size of dots creates an additional geometrical factor enhancing tunneling. NFGM with nano-sized memory node is sometimes referred as single electron memory.[17] While most NFGM use standard MOSFETs for reading, one concept of NFGM uses silicon nanowires in which conductivity is modulated with small charges trapped in localized areas naturally formed within the channel. While this approach suffers lack of reproducibility, the elimination of the bulk MOSFET allows a significant reduction of the cell size. A 128 Mbit memory based on silicon nanowires was demonstrated.[19] All concepts of NFGM with nano-sized memory node, i.e., multidot, single dot, and nanowire memories have been experimentally demonstrated.

Single/few electron memory [20, 21]—In single electron devices electron movement (e.g., the addition or subtraction of an electron to a small 3-dimensional "island", or quantum dot) is controlled with integer electron precision. Injection of each electron onto the quantum dot occurs through a tunneling barrier and is controlled by a separate gate electrode via the "Coulomb blockade" effect. In such quantum dots, electrons are confined electrostatically in all three dimensions, forming a small island of electrons, bounded on all sides by potential walls. This electron island can accommodate only an integer number of electrons, and these electrons can occupy only certain discrete energy states. Connected through tunneling barriers, the conductance of the dot exhibits strong oscillations as the voltage of a gate electrode is varied. Each successive conductance maximum corresponds to the discrete addition of a single electron to the dot. A basic component of single electron memory is the single electron transistor (SET). The SET is composed of a quantum dot connected to an electron source and to a separate electron sink through tunnel junctions with electron injection controlled by a gate electrode. Several concepts of single electron memory have been experimentally demonstrated,[20] including an SET/FET hybrid.[21] Two major disadvantages of all single electron memories reported so far is very low operating temperature of 4.2–20K and background charges.

Molecular memory [22, 23]—Molecular Memory is a broad term combining different proposals for using individual molecules as building blocks of memory cells in which one bit of information can be stored in the space of an atom, molecule or cell. One experimentally demonstrated approach is based on rapid reversible change of effective conductance of a molecule attached between two electrodes controlled by an applied voltage.[22] In this molecular memory data are stored by applying external voltage that cause the transition of the molecule into one of two possible conduction states. Data is read by measuring resistance changes in the molecular cell. There are also concepts for combining molecular components with current memory technology e.g., DRAM and floating gate memory. In this case the molecular element acts as a nano-sized resonant tunnel diode or ultimately small memory node.[23] A major drawback of experimentally demonstrated approaches in molecular memory is their inherent two-terminal nature. A number of advanced molecular components have been proposed, such as 3-terminal molecular devices (molecular transistor),

---

[16] *X. Tang, X. Baie, J. P. Colinge, A. Crahay, B. Katschmarsyj, V. Scheuren, D. Spote, N. Reckinger, F. Van de Wiele, and V. Bayot, "Self-aligned silicon-on-insulator nano flash memory device", Solid-State Electronics 44 (2000) 2259-2264*

[17] *S. M. Sze, "Evolution of nonvolatile semiconductor memory: from floating-gate concept to single-electron memory cell", in:* S. Luryi, J. Xu, and A. Zaslavsky (Eds.), Future Trends in Microelectronics (1999 John Wiley & Sons, Inc), pp. 291-303

[18] *K. Yano, T. Ishii, T. Hashimoto, F. Murai, and K. Seki, "Room-temperature single-electron memory", IEEE Trans. Electron. Dev. 41 (1994) 1628-1638*

[19] *K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, T. Hashimoto, T. Kobayashi, T. Kure, K. Seki, "Single-electron memory for giga-to-tera bit storage, Proc. IEEE 87 (1999) 633-651*

[20] *N. J. Stone and H. Ahmed, "Silicon single electron memory cell", Appl. Phys. Lett. (1998) 2134-2136*

[21] *H. Mizuta, H.-O. Müller, K. Tsukagoshi, D. Williams, Z. Durrani, A. Irvine, G. Evans, S. Amakawa, K. Nakazoto, and Haroon Ahmed, "Nanoscale coulomb blockade memory and logic devices", Nanotechnology 12 (2001) 155-159*

[22] *M. A. Reed, J. Chen, A. M. Rawlett, D. W. Price, and J. M. Tour, "Molecular random access memory cell", Appl. Phys. Lett. 78 (2001) 3735-3737*

[23] *J. Berg, S. Bengtsson, P. Lundgren, "Can molecular resonant tunneling diodes be used for a local refresh if DRAM memory cells?", Solid-State electronics 44 (2000) 2247-2252*

molecular NEMS, and spin-based molecular devices. Numerical data in Table 43, Emerging Memory Devices, are estimates for potential performance of molecular memory based on calculations and early experimental demonstrations.

LOGIC DEVICES

INTRODUCTION

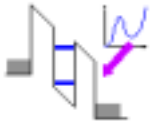
It is likely that the scaling of CMOS device and process technology, as it is known today, will end by the 22 nm node (9 nm physical channel length) by 2016. The grand challenge, then, is to invent and develop one or more new technologies that will extend the scaling of information processing technologies beyond 2016. Such new technologies must meet certain fundamental requirements and possess certain compelling attributes to justify the very substantial investments that will be necessary to build a new infrastructure. First and foremost, any new information processing technology must do the following:

1. Extend microelectronics well beyond (>100×) the domain of CMOS and it should be interfaceable with a CMOS platform. This will require one or more of the following:
   - Functionally scaleable well beyond (>100×) CMOS
   - High information/signal processing rate and throughput
   - Minimum energy per functional operation
   - Minimum scaleable cost per function
2. Provide a means for an energy restorative functional process to sustain steady state operation (e.g., in traditional devices provide a gain mechanism.)

Similar to the memory technologies subsection, the logic technologies tabulated below in Table 44 are a selected, inexhaustive, sample of research concepts to describe many of the alternative approaches currently being pursued by the research community. As with memory concepts, including a particular concept or approach in this subsection does not in any way constitute approval, advocacy, or endorsement. Conversely, not including a particular concept in this subsection does not in any way constitute rejection of that approach. An excellent summary of nanoelectric devices is contained in the *Technology Roadmap for Nanoelectronics*, produced by the European Commission's IST programme (Future and Emerging Technologies).[24]

[24] *"Technology Roadmap for Nanoelectronics," 2nd Edition, Nov. 2000, Ed. R. Compano.*

Table 44 Emerging Logic Devices

						
DEVICE	*RESONANT TUNNELING DIODE – FET*	*SINGLE ELECTRON TRANSISTOR*	*RAPID SINGLE QUANTUM FLUX LOGIC*	*QUANTUM CELLULAR AUTOMATA*	*NANOTUBE DEVICES*	*MOLECULAR DEVICES*
TYPES	**3-terminal**	**3-terminal**	**Josephson Junction +inductance loop**	**-Electronic QCA -Magnetic QCA**	**FET**	**2-terminal and 3-terminal**
ADVANTAGES	**Density, Performance, RF**	**Density, Power, Function**	**High speed, Potentially robust, (insensitive to timing error)**	**High functional density, No interconnect in signal path, Fast and low power**	**Density, Power**	**Identity of individual switches (e.g., size, properties) on sub-nm level. Potential solution to interconnect problem**
CHALLENGES	**Matching of device properties across wafer**	**New device and system, Dimensional control (e.g., room temp operation), Noise (offset charge), Lack of drive current**	**Low temperatures, Fabrication of complex, dense circuitry**	**Limited fan out, Dimensional control (room temperature operation), Architecture, Feedback from devices, Background charge**	**New device and system, Difficult route for fabricating complex circuitry**	**Thermal and environmental stability, Two terminal devices, Need for new architectures**
MATURITY	**Demonstrated**	**Demonstrated**	**Demonstrated**	**Demonstrated**	**Demonstrated**	**Demonstrated**

LOGIC DEVICES –DEFINITION AND DISCUSSION OF TABLE ENTRIES

RTD-FET [25]—The integration of silicon- (or silicon/germanium-) based Resonant Tunneling Diodes (RTDs) with MOSFETs performs certain circuit functions at higher speeds using fewer active devices. Although the RTD is a two-terminal device, integration of RTDs with MOSFETs provides all the advantages of three-terminal devices in circuits that realize the benefits of the RTD's multi-valued operating points of regenerative (negative resistance) I-V characteristics convolved with MOSFET drain I-V characteristics. Applications benefiting most from this technology include SRAMs (one transistor, one stacked capacitor, two RTDs), shift registers and latched comparators. Integration of RTDs with MOSFETs provides high speed operation due to the inherently fast tunneling process, and negative differential resistance regime (NDR) that provides at least two stable operating points (i.e., multiple valued logic) when combined with MOSFETs. A critical challenge is precise control of the MBE-grown layers required to fabricate the RTDs, necessary to control the tunneling current that is exponentially dependent on the tunneling barrier thickness.

SET—The principle of operation of the Single Electron Transistor (SET) proposed for logic applications is the same as that of the Single Electron Memory element described above in the Memory Section. Application of single electron transistors in logic devices are addressed in *Single-Electron Devices and Their Applications* [26] and *Nanoscale Coulomb Blockade Memory and Logic Devices.*[27]

RSFQ [28]—Rapid Single Flux Quantum (RSFQ) Logic is a dynamic logic based upon a superconducting quantum effect, in which the storage and transmission of flux quanta (Fluxon) defines the device operation. The basic RSFQ structure is a

[25] *D. J. Paul, B. Coonan, G. Redmond, G. M. Crean, B. Holländer, S. Mantl, I. Zozoulenko, K.-F. Berggren, "Silicon quantum integrated circuits", in:* S. Luryi, J. Xu, and A. Zaslavsky (Eds.), Future Trends in Microelectronics (1999 John Wiley & Sons, Inc), pp. 183-192

[26] *K. K. Likharev, "Single-electron devices and their applications", Proc. IEEE 87 (1999) 606-632*

[27] *H. Mizuta, H.-O. Müller, K. Tsukagoshi, D. Williams, Z. Durrani, A. Irvine, G. Evans, S. Amakawa, K. Nakazoto, and Haroon Ahmed, "Nanoscale coulomb blockade memory and logic devices", Nanotechnology 12 (2001) 155-159*

[28] *K. Block, K. Track and M. Rowell, Superconducting ICs: the 100 GHz second generation, IEEE Spectrum, December 2000, P 40-46*

superconducting ring that contains one Josephson junction (JJ) plus an external resistive shunt. The storage element is the superconducting inductive ring and the switching element is the Josephson junction. RSFQ dynamic logic uses the presence or absence of the flux quanta in the closed superconducting inductive loop to represent a bit as a "1" or "0," respectively. The circuit operates by temporarily closing the Josephson junction, thereby ejecting the stored flux quanta. A quantized voltage pulse is then generated across the Josephson junction. This voltage pulse is propagated down a superconducting transmission line and can be used to trigger other RSFQ structures in various combinations to form complex circuits. As this quantum effect occurs at a macroscopic scale, sub-micron lithography is not a prerequisite. With RSFQ, circuit speeds above 100 GHz, perhaps up to 750 GHz, are possible. RSFQ circuits are currently built on low temperature superconducting Josephson Junctions (~ 5 K). However, high temperature superconductors may eventually be exploited. RSFQ devices need extreme cooling because the device operating temperature is lower than the critical temperature of the bulk superconductor material. The availability of adequate cooling systems, which comply with needed specifications (temperature, size, weight, dimensions, etc.) in the limits of reasonable prices, is one of the most important drawbacks for the market introduction of this technology.

Quantum cellular automata (QCA)—The QCA paradigm is one in which a regular array of cells, each interacting with its neighbors, is employed in a locally interconnected architecture. The coupling between the cells is given by their electrostatic interactions and not by wires. Such arrays are in principle capable of encoding either analog or digital information. Two predominant types of QCA are the following:

- Electronic QCAs (EQCAs) [29,]
- Magnetic QCAs (MQCAs) [30]

Magnetic QCAs are a recent development, and their full potential performance cannot yet be assessed. The archetype QCA is the electronic QCA. A single EQCA cell is made up of 4, 5, or 6 individual "quantum" dots or isolated metallic islands. In a 4-dot cell, the quantum dots occupy the corners of a square cell. It can be shown that the charges will occupy the dots in diagonally opposite corners of the cell and form two bistable states representing + 1 and -1. The 5 and 6-dot systems are more complicated, but are similar in principle. The physical mechanisms of interaction between the nanoelectronic structures or dots are the Coulomb interaction and quantum mechanical tunneling. If the cells are arranged in a regular square grid then long-established cellular automaton theory can be applied, together with its extension, Cellular Non-linear (or Neural) Network (CNN) theory to describe the information processing algorithm. This allows a large body of theory to be applied directly to QCA architectures, which are further described below in the Architectures Section.

Nanotube devices [31]—Nanotube electronic materials are a special and a very important subset of molecular electronic materials described below. A carbon Nanotube is a molecular "tube" or cylinder formed from an atomic "sheet" of carbon atoms. These carbon atoms are bonded together into an array of hexagons, which form a planar sheet, similar to an atomic sheet of graphite (resembling a planar assembly of open hexagons). This graphite-like (graphene) sheet is rolled up to form a carbon nanotube. Carbon nanotubes can have diameters between 1–20 nm and lengths from 100 nm to several microns. The tube diameter and just how the sheet of carbon hexagons is rolled up determine whether a tube is a semiconductor or a metal. If a tube is a semiconductor, the details of rolling also determine the energy bandgap and, therefore, the electronic properties of the tube. These bandgap energies range all the way from zero (like a metal) to values as large as silicon (1.1-eV), with many values in between. The tubes can be doped both p- and n-type making possible p-n junctions. Several groups have demonstrated p-FET device structures in which a gate electrode modulates the conductivity of a conducting channel by a factor of 10^5 or more, similar to silicon MOSFETs.[32] Large arrays of carbon nanotube FETs have been fabricated [33] and, more recently, a voltage inverter or NOT gate circuit using one n-channel and

[29] *C.S Lent. and P.D. Tougaw, "Dynamics of quantum cellular automata", J. Appl. Phys. 80, 4722-4736, 1996*
[30] *R.P. Cowburn and M.E. Welland, "Room temperature magnetic quantum cellular automata", Science 287, 1466-1468, 2000.*
[31] *P. G. Collins and P. Avouris, "Nanotubes for Electronics", Scientific American, Dec. 2000, 62-69*
[32] *R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and P. Avouris, "Single- and multi-wall carbon nanotube field-effect transistors", Appl. Phys. Letters 73, 2447 (1998)*
[33] *P.G. Collins, M. Arnold, and P. Avouris, "Engineering carbon nanotubes and nanotube circuits using electrical breakdown", Science 292, 706 (2001).*

one p-channel FET was demonstrated.[34] Both the p- and n-channel FETs in this inverter were fabricated in a single carbon nanotube.

Molecular devices—Molecular electronic logic devices are based on electron transport properties through a single molecule.[35] Experimental demonstrations to date have been performed using two-terminal molecular devices[36], although three-terminal molecular structures have been proposed[37] and one has been recently demonstrated.[38] Two terminal molecular devices currently being explored consist of thousands of molecules operating in parallel, e.g., as digital switches or as analog diodes. In both cases a voltage applied to a group of parallel molecules results in reconfiguration of the molecular components, or moieties, and a change in the molecule's electrical conduction properties.[37] [39] The mechanism of charge transport in molecules is not understood. One possible model is this change in a molecule's electrical properties is caused by a change in the overlap of the orbitals in the molecule. The correct overlap of the molecular orbitals allows electrons to flow through the molecule. But when this overlap of orbitals is further changed (because the molecule has been twisted or its geometry has been otherwise changed) the flow of electrons is impeded. A near-term opportunity of molecular electronics is in integration of molecular components with sub-100 nm CMOS[40] to form hybrid systems. Potential longer-term opportunities are full-molecular systems. In addition to two-terminal digital switches and analog diodes, several other molecular components have been recently studied both experimentally and theoretically, e.g., bistable switch, molecular NEMS, three-terminal molecular devices (molecular transistor), and spin based molecular devices. A brief summary of the status of exploration (maturity) of molecular components for nanoelectronics is given in Table 45.

Table 45 Status of Molecular Components for Nanoelectronics

MOLECULAR COMPONENT	*MATURITY*
Conductor	**Demonstration** [35]
Insulator	**Demonstration** [35]
Diode	**Demonstration** [35]
Resonant tunnel diode	**Demonstration** [35]
Memory (bistable 2-terminal switch)	**Demonstration** [36]
Transistor (3-terminal switch)	**Demonstration** [37, 38]
Molecular NEMS	**Concept** [39]
Molecular optoelectronic devices	**Concept** [39,40]

ARCHITECTURES

INTRODUCTION

For the past 30 years, relentless scaling of CMOS IC technology to incredibly smaller dimensions has enabled increasingly complex microelectronics systems functions (information processing technologies). The continued scaling of silicon CMOS beyond the end of this roadmap indefinitely, however, is not likely. CMOS scaling may come to an asymptotic end around the 22 nm node by 2016. Driven by this scaling cadence, device and circuit performance issues have emerged, such as leakage current, power dissipation, interconnect latency, etc. In addition, emerging market-driven applications are motivating designers and architects to create new approaches to realizing systems functions using an increasingly stable CMOS platform. Consequently, the pressure to continue performance scaling of silicon-based devices leads to the development of new information processing architectures in a manner similar to development of new devices.

---

[34] V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, "Carbon nanotube inter- and intramolecular logic gates", Nano Letters, 1(9), 453 (2001).

[35] J. C. Ellenbogen and J. C. Love, "Architectures for molecular electronic computers:1. Logic structures and an adder designed from molecular electronic diodes", Proc. IEEE 88 (2000) 386-425

[36] M. A. Reed, J. Chen, A. M. Rawlett, D. W. Price, and J. M. Tour, "Molecular random access memory cell", Appl. Phys. Lett. 78 (2001) 3735-3737

[37] E. Emberly and G. Kirczenov, "Principles for the design and operation of a molecular wire transistor", J. Appl. Phys. 88 (2000) 5280-5282

[38] J.H. Schon, H. Melig, Z. Bao, "Field-effect Modulation of the Conductance of Single Molecules," Sciencexpress, www.sciencexpress.org/8 Nov. 2001, page 2.

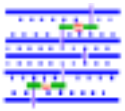
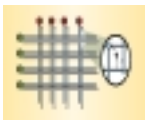
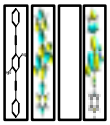
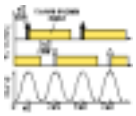
[39] Y. Wada, "Prospects for single molecule information processing devices", Proc. IEEE 89 (2001) 1147-1771

[40] G. F. Cerofolini, G. Ferla, "Toward a hybrid micro-nanoelectronics", J. Nanoparticle Res. (2001)

In the mean time, researchers in other disciplines (for example, physics and computer science) are exploring striking new architectural approaches to information processing. Some of these, such as quantum computing, are driving research on practical materials and device technologies to realize their functions.

Two other forces also are driving research into new circuit design and systems architectures. One is the invention of devices and technologies embodying completely new physical principles of operation beyond the FET (e.g., molecular electronics and quantum cellular automata). These new structures certainly will require new architectures to achieve useful functionality. Development of new devices can be thought of as driving the development of new architectures. The second is new processing technologies combining CMOS functions (such as logic, memory, etc.) with each other and with other functions (RF, analog, optical, MEMS, etc.). Development of advanced wafer bonding techniques has driven consideration of 3D integration of silicon devices and the requirement for new systems architectures. Similarly, the development of optical I/O with broad bandwidth capability may drive a new architecture that would exploit that capability. Table 46 outlines the emerging research architectures.

Table 46 Emerging Research Architectures

						
Architectures	*3-D Integration*	*Quantum Cellular Automata*	*Defect Tolerant Architecture*	*Molecular Architecture*	*Cellular Nonlinear Networks*	*Quantum Computing*
Device Implementation	**CMOS with dissimilar material systems**	**Arrays of quantum dots**	**Intelligently assembles nanodevices**	**Molecular switches and memories**	**Single electron array architectures**	**Spin resonance transistors, NMR devices, Single flux quantum devices**
Advantages	**Less interconnect delay, Enables mixed technology solutions**	**High functional density. No interconnects in signal path**	**Supports hardware with defect densities >50%**	**Supports memory based computing**	**Enables utilization of single electron devices at room temperature**	**Exponential performance scaling, Enables unbreakable cryptography**
Challenges	**Heat removal, No design tools, Difficult test and measurement**	**Limited fan out, Dimensional control (low temperature operation), Sensitive to background charge**	**Requires pre-computing test**	**Limited functionality**	**Subject to background noise, Tight tolerances**	**Extreme application limitation, Extreme technology**
Maturity	**Demonstration**	**Demonstration**	**Demonstration**	**Concept**	**Demonstration**	**Concept**

Architectures –Definition and Discussion of Table Entries

3D heterogeneous integration—Integration of semiconductor devices in 3D arrays is being driven from two distinct directions. The first is associated with the need to integrate dissimilar technologies on a common platform to deliver an optimum information processing solution. It is clear that emerging technologies beyond scaled CMOS offer the potential for greatly improved performance by mixing and matching technologies for particular applications. The combination of technologies requires the 3D integration of various functionally dissimilar technologies beginning with microprocessors, ASICs, DRAMs and extending to RF, analog, optical and MEMS. These dissimilar technologies may later include 3D integration of molecular, plastic, RSFQs and others directly on to silicon platforms. The other principle driver for 3D integration is the need to reduce delays of global interconnects to maximize system performance. It has been shown that 3D superposition of devices will decrease interconnect delays by 60% relative to an equivalent number of transistors arranged in a planar arrangement.[41, 42] Low-temperature wafer or chip-bonding techniques are necessary to achieve 3D

[41] *J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. I. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, and J. D. Meindl, "Interconnect limits on gigascale integration (GSI) in the 21st century", Proc. IEEE 89 (2001) 305-324*

[42] *K. Banerjee, S. J. Souri, P. Kapur, and K. C. Saraswat, "3-D ICs: A novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration", Proc. IEEE 89 (2001) 602-633*

integration. It is very improbable that an integrated process on a common substrate could be developed that would accommodate the various emerging technologies, all of which have superior performance to silicon in specialized niche applications. Bonding of wafers at temperatures less than 200°C is necessary in order not to affect the materials and structures fabricated in previous processing steps.

Quantum cellular automata architecture—The QCA paradigm is one in which a regular array of cells, each interacting with its neighbors, is employed in a locally interconnected architecture. The physical operation of QCA nanostructures is described above in the Logic Devices section. If QCA cells are arranged in a regular square grid then long-established cellular automaton theory can be applied together with its extension Cellular Neural Network (CNN) theory to describe the information processing algorithm. This allows a large body of theory to be applied directly to QCA architectures. Regular EQCA grids have been shown to be capable of solving certain types of diffusion and wave equations.[43] Also, it is theoretically possible that they could be used to solve arbitrary digital logic problems, but such systems would be extraordinarily inefficient in terms of area. However, by departing from the regular grid structure, it is possible to design EQCA structures that could carry out universal computing with moderate efficiency. In addition to a non-uniform layout, EQCAs need a spatially non-uniform "adiabatic clocking field," which controls the switching of the cells from one state to another, and allows them to evolve relatively rapidly to a stable end state. This clock has the additional advantage, that it effectively produces gain, non-linearity and isolation between neighboring parts of a circuit. Recent experimental results[43] have confirmed the original theory. Using clocking adds to the complexity of pure EQCA circuits but greatly extends their possible range of applications. It is possible to construct a complete set of Boolean logic gates with EQCA cells and to design arbitrary computing structures, but current device and circuit analyses indicate that the speed of EQCA circuits will be limited to less than about 10 MHz.[44, 45]

Defect tolerant architecture[46]**—**The defect tolerant computer architecture implies that a system operates satisfactory in the presence of errors made in the hardware during manufacture. Note this is different from fault tolerance, which implies the ability of a machine to recover from errors made during a calculation. The need for defect tolerant hardware emerges from the possibility of fabricating nanometer scale elements that will probably not satisfy tolerance and reliability requirements that are typical for larger scale systems. Systems consisting of molecular-size components are likely to have many imperfections, and a computing system designed on conventional zero-defect basis would not work.

For a conventional integrated circuit, a description of the chip function is first developed, and then the hardware is constructed. The general idea behind defect tolerant architectures is conceptually the opposite. A generic set of wires and switches are fabricated, and then the resources are configured by setting switches linking them together to obtain the desired functionality.[46] A cornerstone of defect tolerant systems is redundancy of hardware resources, thus extra components such as switches, memory cells, and wires are needed. This redundancy in turn implies very high integration density. The fabrication could be very inexpensive (e.g., the limit case would be chemical self-assembly of molecular switches on a 3-dimensional random array of wires). However, to make such a circuit operational a laborious process of testing is needed when the devices are trained to the desired level of proficiency with computer tutors that find the defects and record their locations in on-chip databases.[46] In principle not only information on defects, but all answers to the input questions (such as all logic functions) can be put into memory cells, provided there is adequate amount of fast-access memory. On the other hand, to deal with defected elements, an opportunity for re-routing the computational trajectories in hardware should exist. This implies spare wires to provide a large communications bandwidth. Such a "memory and wires" approach to computation, while very challenging, may be realized in molecular computers.

The main two potential advantages of defect tolerant architectures are as follows:

1. The possibility of building complex systems from inherently imperfect nanoscale components
2. The potential for self-repair from operation-originated defects by reconfiguring the system

An important disadvantage of defect tolerant computing is the need for a laborous post-fabrication learning process. Also, very large amounts of wiring and spare devices may be needed to cope with relatively high defect rates. However, the

---

[43] Porod, W. et al., "Quantum-dot cellular automata: computing with coupled quantum dots", Int. J. Electronics 86, 549-590, 1999.

[44] Bonci L., Iannaccone G. & Macucci M., "Performance assessment of adiabatic quantum cellular automata", J. Appl. Phys. 89, 6435-6443, 2001

[45] Nikolic K., Berzon D. & Forshaw M., "Relative performance of three nanoscale devices - CMOS, RTDs and QCAs - against a standard computing task", Nanotechnology 12, 38-43, 2001

[46] J. R. Heath, P. J. Kuekes, G. S. Snider, R. S. Williams, "A defect-tolerant computer architecture: opportunities for nanotechnology", Science 280 (1998) 1716-1721

defect tolerant architecture is a broad concept, which can be realized with different approaches, for example in memory-based molecular computer or with cellular non-linear networks.

Molecular architecture—The basic requirements for a general-purpose information processing technology are a three-terminal switching device, memory, and a way of connecting arbitrarily large numbers of the devices and memory elements. Also required is a way to interconnect large numbers of devices, particularly over large distances, analogous to global interconnects. To date, researchers have demonstrated two-terminal and, recently, three-terminal single-molecule switches, negative differential resistance diodes, and memory elements. Realistically, construction of complex logic circuits requires a realizable three-terminal device with gain. Even more imposing, a method is needed for connecting huge numbers of the devices to form a systems function. Although no potential solutions to this problem are apparent yet, radically new systems architectures likely will be needed to exploit molecular devices fully to form a system's function.

Cellular non-linear networks (CNN) [47, 48]—Cellular Nonlinear Network is a 2, 3, or N dimensional array of mainly identical dynamical systems, called cells, which satisfy two properties: 1) most interactions are local within a finite radius r, and 2) the state variables are continuous valued signals (not digital). A template specifies the interaction between each cell and all its neighbor cells in terms of their input, state, and output variables. The interaction between the variables of one cell may be either a linear or nonlinear function of the variables associated with its neighbor cells that lie within the radius r. A cloning function defines how the template varies spatially across the grid and determines the dynamical response of the array to boundary values and initial conditions. Since the range of interaction and the connection complexity of each cell are independent of the number of cells, the architecture is extremely scalable, reliable and robust. Programming the array consists of specifying the dynamics of a single cell, the connection template and the cloning function of the templates. This is significantly simpler than traditional VLSI design methodology because of the inherent use of relatively simple functional components. CNNs can be used to realize a complete basis set of the Boolean logical gates as well as more complex gates such as the majority gate, MUX gates, and switches. CNNs can directly solve classical diffusion and convection equations as well as more complex sets of PDEs. They can directly simulate certain functions of the nervous system because of their local connectivity characteristic. The CNN architecture is well suited to hardware implementations that involve imperfect hardware that includes many types of nanoscale or self-assembling devices. They notably include molecular devices, quantum dots, and single electron transistors (in the form of tunneling phase logic) and resonant tunneling diodes.[48]

Quantum information processing—Coherent quantum devices rely on the phase information of the quantum wavefunction to store and manipulate information. The phase information of any quantum state is called a qubit and is extremely sensitive to its external environment. It is easily connected or entangled with the quantum states of particles in the local environment. No physical system can ever be completely isolated from its environment. The same sensitivity however can be used to entangle adjacent qubits in ways that can be controlled by physical gates. The core idea of quantum information processing or quantum computing is that each individual component of an infinite superposition of wavefunctions is manipulated in parallel, thereby achieving a massive speed up relative to conventional computers. The challenge is to manipulate the wavefunctions so that they perform a useful function and then to find a way to read out the result of the calculation.

Essentially three different approaches have been taken to the implementation of quantum computers. These are as follows:

---

[47] *L. O. Chua, "CNN: A paradigm for complexity" (Word Scientific Publishing Co. 1998)*

[48] *In the literature, the abbreviation CNN is* also *used to indicate Cellular* Neural *Networks. Cellular Neural Network architectures also satisfy conditions 1 and 2 above but differ from Cellular* Nonlinear *Networks in that Cellular Neural Networks allow only for* linear *interactions within and between cells. Consequently, Cellular Nonlinear Networks are capable of solving a more general class of problems*

1. Bulk resonance quantum implementations including Nuclear Magnetic Resonance (NMR),[49] linear optics,[50] and cavity quantum electrodynamics (CQED)[51]
2. Atomic quantum implementations including trapped ions[52] and optical lattices[53]
3. Solid state quantum implementations including semiconductors[54] and superconductors[55]

The emphasis of this discussion is centered on solid state quantum computing, with a focus on semiconductors, since this approach is perceived to be most attractive for commercialization. It is conjectured that solid state implementations offer the highest promise to scale the complexity of the quantum computer.

As stated above, a fundamental notion in quantum computing is the "qubit," a concept that parallels the "bit" in conventional computation, but carrying with it a much broader set of representations. Rather than a finite dimensional binary representation for information, the qubit is a member of a two-dimensional Hilbert space containing a continuum of elements. Thus quantum computers operate in a much richer space than binary computers. Researchers have defined many sets of elementary quantum gates based on the qubit concept that perform mappings from the set of input quantum registers to a set of output quantum registers. A single gate can entangle the qubits stored in two adjacent quantum registers and combinations of gates can be used to perform more complex computations. It can be shown that, just as in Boolean computation, there exist minimal sets of quantum gates that are complete with respect to the set of computable functions. Considerable research has been conducted to define the capabilities of quantum computers. Theoretically quantum computers are not inferior to standard computers of similar complexity and speed of operation. More interesting is the fact that for some important classes of problems, the quantum computer is superior to its standard counterpoint. In particular, it was shown that the two prime factors of a number can be determined by a quantum computer in time proportional to a polynomial in the number of digits in the number.[56] This truly remarkable result showed that for this particular class of problems, the quantum computer is at least exponentially better than a standard computer. The key to this result is the capability of a quantum computer to efficiently compute the quantum Fourier Transform. This result has immediate application in cryptography since it would allow the quick determination of keys to codes such as RSA. It is estimated that few thousand quantum gates would be sufficient to solve a representative RSA code containing of on the order of one hundred digits. There are several other applications that are variants of the factorization problem.[57]

---

[49] M. Steffen, L. M. K. Vandersypen, and I. L. Chuang, "Toward Quantum Computation: A five Qubit Quantum Processor," IEEE Micro 21(2) (2001) 24-34

[50] Shigeki Takeuchi, "Experimental demonstration of a three-qubit quantum computation algorithm using a single photon and linear optics", Phys. Rev. 62 (2000)

[51] Grangier P, Reymond G, Schlosser N, "Implementations of quantum computing using cavity quantum electrodynamics", Fortschr. Phys.-Progress Of Physics 48 (2000) 859-874 schemes

[52] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland, "Demonstration of a fundamental quantum logic gate", Phys. Rev. Lett. 75 (1995) 4714-4717

[53] Milburn GJ, "Quantum computing using a neutral atom optical lattice: An appraisal" Fortschr. Phys.-Progress Of Physics 48 (2000) 957-964

[54] B. E. Kane, "A silicon-based nuclear spin quantum computer", Nature 393 (1998) 133-137

[55] D. V. Averin, "Quantum computation and quantum coherence in mesoscopic Josephson junctions", J. Low Temp. Phys. 118 (2000) 781-793

[56] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring", Proc. 35nd Annual Symposium on Foundations of Computer Science, IEEE Computer Society Press (1994), 124-134

[57] C. P. Williams and S. H. Clearwater, "Explorations in Quantum Computing" (1998 Springer-Verlag, New York, Inc)

EMERGING TECHNOLOGY PARAMETERIZATION

EMERGING RESEARCH TECHNOLOGIES

INTRODUCTION

This section relates some of the radically new information processing technologies to each other and to scaled CMOS. These emerging technologies in general are not a direct replacement for silicon and may require radically new architectures to achieve their functionality. For this reason, it is difficult but not impossible to find one set of parameters that that can characterize and relate the strengths and weaknesses of the emerging technologies in a meaningful way.

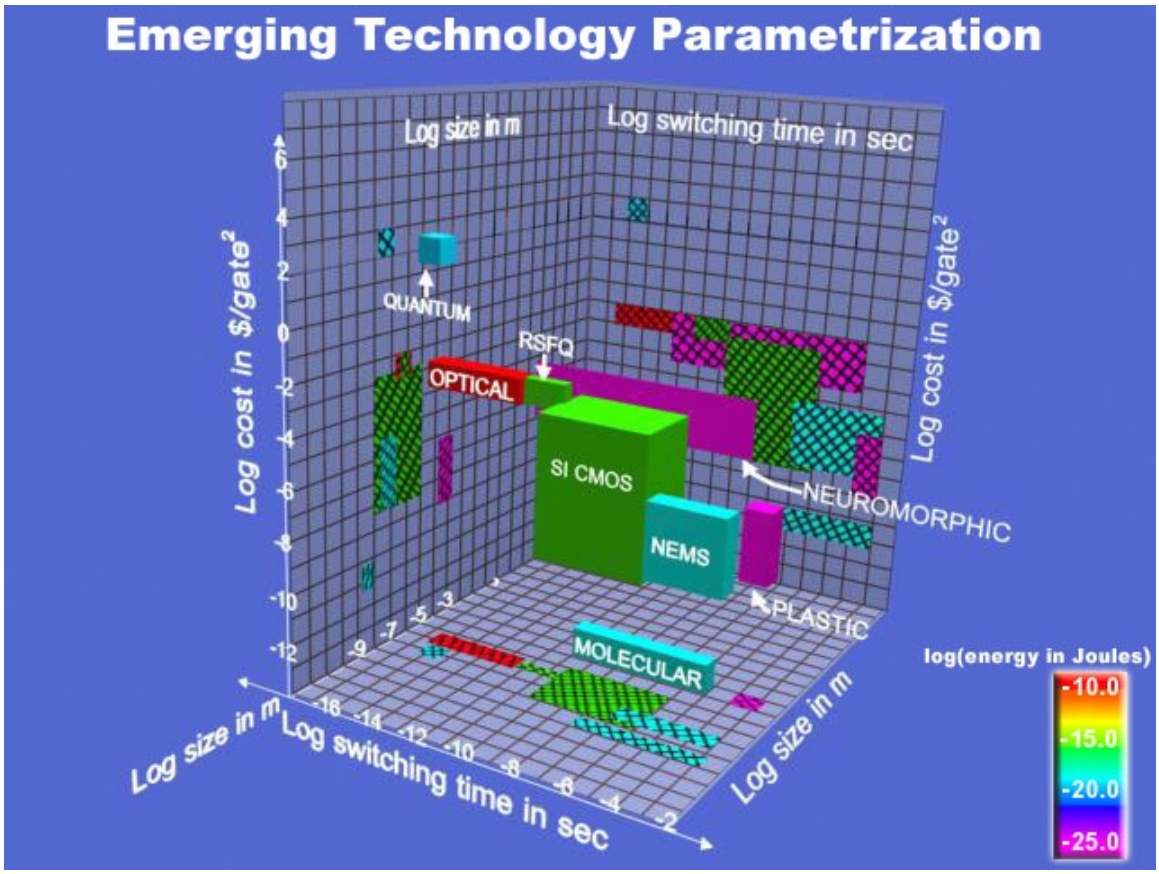


Figure 25 Parameterization of Emerging Technologies and CMOS—Speed, Size, Cost, Power

Figure 25 shows a parameterization of a selected set of emerging technologies and CMOS in terms of speed, size, cost and power.[58] Four of the technologies, including CMOS, are introduced in the CMOS and the Logic sections and four others (Plastic, Optical, NEMS, and Neuromorphic) are described in this section. (Neuromorphic, like CMOS, is plotted for the purpose of reference comparison.) The first three parameters in this figure are used to define a 3-dimensional space and the fourth parameter, switching energy, is displayed as color code shown in the legend in the lower right hand corner of the graph. All the scales are logarithmic and cover many orders of magnitude as shown in the graph. Each of the technologies displaces a certain volume in this parameter space and is color coded in a solid color representing the energy required for a single gate operation. Each of the volumes is then projected onto the bounding 2D planes so that quantitative values can be determined. The projections of the volume corresponding to a given technology are shown as crosshatched rectangles filled with the same color as the corresponding volume.

[58] *Research Triangle Institute, Technology Assisted Learning Division, [Ms. P. J. Woodard, Mr. Sam Field, and Mr. Dale Rowe] is gratefully acknowledged for providing technical support in the preparation of Figure 25.*

In the absence of firm measured data, a number of assumptions were made to estimate the parameters for the emerging technologies. The parameters used for each technology are listed in Table 47. If an emerging technology is in the concept stage with no measured data, the parametric assumptions are based on the underlying physical principles. If some measured data exists, the assumptions involve an estimate on how far the technology can be scaled. In this case, the scaling arguments are based on physical principles.

Several of the technologies listed are strongly tied to a single application area or niche where the technology is particularly effective. For example, quantum computing can find prime factors very efficiently by means of Shore's algorithm[56] but is much less efficient on other applications. In this case, we define an "effective" time per operation as the time required by a classical device in a classical architecture using a classical algorithm to do the calculation. Therefore the "effective" operation time of an N qubit quantum computer factoring a large number is very much faster than the operation time of an N gate classical computer because of the inherent parallelism associated with quantum computing. Similar arguments can be made for neuromorphic and optical computing.

This figure conveys meaningful information about the relative positions of the emerging technologies in this application space. It shows that few of the new technologies are directly competitive with scaled CMOS and most are highly complimentary. It also shows very clearly the benefit to be derived from heterogeneous integration of the emerging technologies with silicon to expand its overall application space. Figure 25 and Table 47 represent initial estimates for the comparison of these very disparate technologies. In addition to this comparison, the further intent of this figure and table is to stimulate substantive discussion of the basis and means for making this comparison.

Table 47 Estimated Parameters for Emerging Research Devices and Technologies in the year 2016

Technology	T_{min} sec	T_{max} sec	CD_{min} m	CD_{max} m	Energy J/op	Cost min $/gate	Cost max $/gate
Si CMOS	3E-11[59]	1E-6	8E-9	5E-6	4E-18	4E-9	3E-3
RSFQ	1E-12	5E-11	3E-7	1E-6	2E-18	1E-3	1E-2
Molecular	1E-8	1E-3	1E-9	5E-9	1E-20	1E-11	1E-10
Plastic	1E-4	1E-3	1E-4	1E-3	1E-24	1E-9	1E-6
Optical (digital)	1E-16	1E-12	2E-7	2E-6	1E-12	1E-3	1E-2
NEMS (conservative)	1E-7	1E-3	1E-8	1E-7	1E-21	1E-8 [60]	1E-5
Neuromorphic	1E-13	1E-4	6E-6	6E-6	3E-25	5E-5	1E-2
Quantum	1E-16	1E-15	1E-8	1E-7	1E-21	1E2	1E3

In this table T stands for system cycle time (switching time), CD stands for critical dimension (e.g., physical gate length), Energy is the intrinsic operational energy of one device, and Cost is defined as $ per gate.

EMERGING RESEARCH TECHNOLOGIES –DEFINITION AND DISCUSSION OF TABLE ENTRIES

Plastic Transistors—Plastic transistors are defined to be Thin Film Transistor (TFT) devices fabricated on plastic substrates. The active layer of the TFT can be amorphous or poly-Si as well as organic semiconductors. Often, the TFTs are combined with Organic Light Emitting Diodes (OLEDs) to form intelligent, flexible display devices than can be bent, folded, worn or conformally mapped on to arbitrarily shaped surfaces. All-plastic chips based entirely on organic materials have already been demonstrated whose mechanical flexibility offers totally new perspectives to, for example, the rapidly growing market of identification and product tagging as well as for pixel drivers for flexible displays.[61] Typical devices have a supply voltage of 10 V and critical dimensions of 100-μm with reasonable electron mobilities and I-V characteristics. Pentacene-based plastic transistors with on/off current ratio $>10^5$ at operating voltage ranges as low as 5 volts were reported.[62] Recently, analog and digital circuits using organic (pentacene) transistors on polyester substrates were fabricated and characterized. The highest operation frequency reported to date for organic circuits on plastic

[59] T_{min} for silicon CMOS is based on the local clock rate for the 22 nm node (physical gate length < 9 nm), and not upon CV/I intrinsic switching time.
[60] Estimated on the principle of reasonable cost and assumed two-dimensional architecture of NEMS computer
[61] F. Würthner, "Plastic transistors reach maturity for mass applications in microelectronics", Angew. Chem. Int. Ed. 40 (2001) 1037-1039
[62] C. D. Dimitrakopoulos, S. Purushothaman, J. Kymissis, A. Callegari, J. M. Shaw, "Low-voltage organic transistors on plastic comprising high-dielectric constant gate insulators", Science 283 (1999) 822-824

substrates is 1.7 kHz.[63] Plastic transistors have the potential to provide very cheap, rugged large area electronics that have many potential applications.[64, 65] A process technology consisting just of printing operations on paper based substrates would have an intrinsic cost structure similar to color inkjet printing today. It could support disposable devices such as periodicals and dynamic bar codes.

Optical[66]—Optical computing is based on using light transmission and interaction with solids for information processing. The potential advantages of digital optical computers relate to the following properties of light as a carrier of information:

- Optical beams do not interact with each other
- Optical information processing functions can be performed in parallel (performing a Fourier transform, for example)
- Ultimately high speed of signal propagation (speed of light)

It should be noted that what is called the all-optical computer still contains electronic components, such as lasers and nonlinear elements in which a material's optical properties are affected by charge carriers or atoms interacting with light. Some disadvantages of digital optical computing include:

- The relatively large size of components (e.g. optical switch) arising from diffraction limitation
- Potential of high speed computation can be realized only at the expense of dissipated power. For example, in an optically controlled phase change material (switch or memory), faster rearrangement of atoms in a cell requires larger energy supply. In a practical device "computing at the speed of light" is unlikely since it would require an huge operational energy.

Near term opportunities in optoelectronics are in integration of photonic components with sub-100 nm CMOS. Another opportunity arises from using optically controlled phase-change materials, such as e.g., OUM described in the Memory section. Another direction is perfection of existing analog optical computers, which perform e.g. Fourier processing much faster than electronics. Analog optical computers are fast and operate with continuous data, while their accuracy is not comparable to that of digital computers.

Nano-electro-mechanical systems (NEMS)—In the concept of the nanomechanical computer, mechanical digital signals are represented by displacements of solid rods, and the speed of signal propagation is limited to the speed of the sound (e.g., 1.7×10^4 m/s in diamond). Optimistic estimates predict NEMS logic gates that switch in 0.1 ns and dissipate less than 10^{-21} J and computers that perform 10^{16} instructions per Watt (compared to 5×10^{12} instruction per Watt in human brain operation). This estimated switching energy is below the thermodynamic limit of $kT\ln 2$ for irreversible computation. It is believed [67] that this low dissipation is possible because this NEMS computation is logically reversible. More conservative estimates of characteristics of the NEMS computers can be made based on recent demonstrations of VLSI-NEMS chip for parallel data storage (e.g., IBM's Milliped concept [68]). Reported storage densities are 500 Gbit/in^2. The highest data rates achieved so far are 6 Mbit/s. A summary of the conservative estimates of parameters of the NEMS computers is given in Table 47.

Neuromorphic—The human brain is defined to be the archetypal Neuromorphic information processing device and is included here to provide a basis of comparison with silicon-based information processing systems. The scale length of individual neurons is estimated from the volume of the brain and the estimated number of neurons. It is possible to derive an "effective operation time" of Neuromorphic computing as explained in the overview of this section. In that case, the reference operation is vision processing where there is a great deal of information relating to technological systems. The effective times defined in this way are very much faster that the synaptic speed and reflects that the interconnect density of the human brain is very much greater than any silicon-based system. The speed quoted in Table 47 above for T_{min} is based on the estimated information processing rate of 1E13 bits per second[69] related to vision processing. Similarly, the speed quoted in Table 47 for T_{max} is the experimentally observed time scale for opening and closing of synapses. Each

---

[63] *M. G. Kane, J. Campi, M. S. Hammond, F. P. Cuomo, B. Greening, C. D. Sheraw, J. A. Nichols, D. J. Gundlach, J. R. Huang, C. C. Kuo, L. Jia, H. Klauk, T. N. Jackson, "Analog and digital circuits using organic thin-film transistors on polyester substrates", IEEE Electron. Dev. Lett. 21 (2000) 534-536*

[64] *J. M. Xu, "Plastic electronics and futurte trends in microelectronics", Synthetic Metals 115 (2000) 1-3*

[65] *S. Forrest, P. Burrows, and M. Thompsan, "The dawn of organic electronics", IEEE Spectrum, Aug. 2000 p. 29-34*

[66] *H. J. Caulfield, "Perspectives in Optical Computing", Computer, Feb. 1998, p. 22-25*

[67] *K. Eric Drexler, Nanosystems: molecular machinery, manufacturing and computation (John Wiley & Sons, Inc. 1992)*

[68] *M. Despont, J. Brugger, U. Drechsler, U. Düring, W. Haberle, M. Lutwyche, H. Rothuizen, R. Stutz, R. Widmer, G. Binnig, H. Rohrer, P. Vettiger, "VLSI-NEMS chip for parallel AFM data storage", Sensors and Actuators 80 (2000) 100-107*

[69] *R. U. Ayres, Information, Entropy, and Progress (AIP Press 1994)*

neuron will connect to between 100 and 10,000 synapses and this is one of the primary ways in which the architecture of the human brain differs from silicon-based systems.

The fundamental parameters of the human brain are estimated to be

- Number of neurons—1E11–1E12[70]
- A single neuron can make 100 to 10,000 synaptic connections[71]
- Mass—1.5 kg
- Volume—1.5 l
- Energy consumption—2.5 W
- Information stored—1E14 bits[71]
- Information processed—1E13 bits/s[71]

The set of secondary parameters shown in Table 47 are based on the fundamental parameters above.

[70] *J. E. Dowling, The retina: an approachable part of the brain (The Belknap Press of Harvard University Press 1987)*
[71] *R. U. Ayres, Information, Entropy, and Progress (AIP Press 1994)*