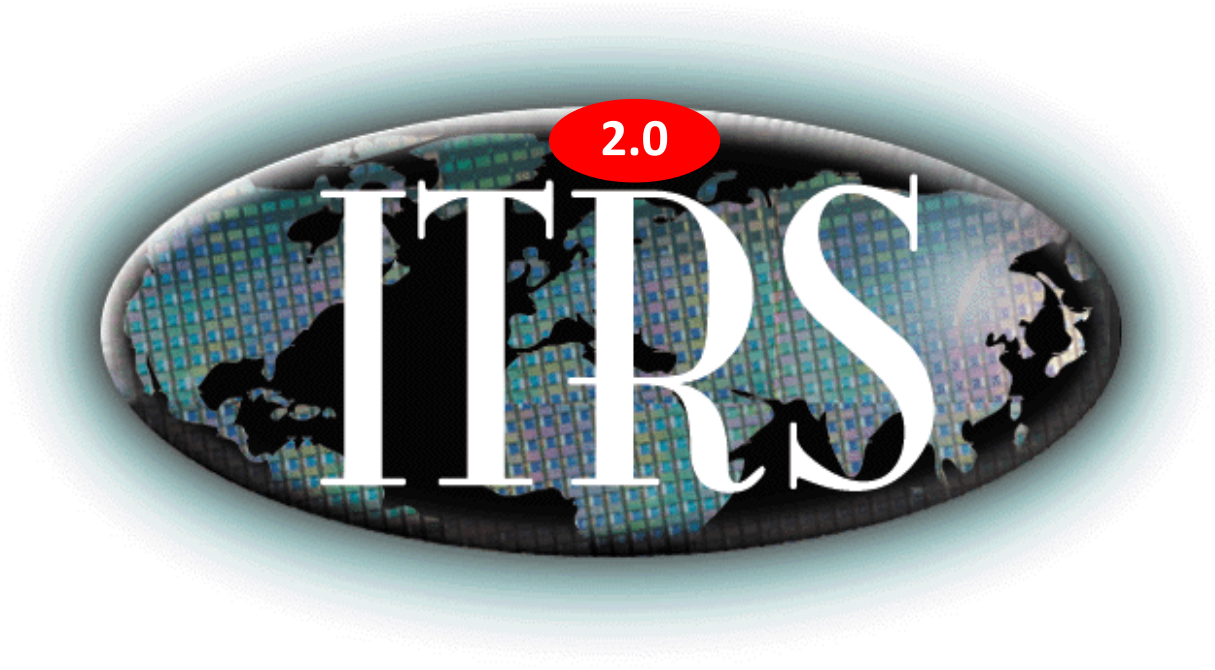


INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

2015 EDITION

MORE MOORE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

MORE MOORE

1 SCOPE AND MISSION

System scaling enabled by Moore's scaling is more and more challenged with the scarcity of resources such as power and interconnect bandwidth. Particularly due to the emergence of cloud, seamless interaction of big-data and instant data have become a necessity (Figure MM1). Instant data generation require ultra-low-power device with "always-on" feature at the same time with high-performance device that can generate the data instantly. Big data require abundant computing and memory resources to generate the service and information that clients need.

Following applications drive requirements of More Moore technologies.

- High-performance computing – targeting more performance (operating frequency) at constant power density (constrained by thermal).
- Mobile computing – targeting more performance (operating frequency) and functionality at constant energy (constrained by battery) and cost
- Autonomous sensing & computing (Internet-of-Things: IoT) – targeting reduced leakage & variability

Figure MM1: Big data and instant data.

These applications dictated the need for More Moore platform to bring the PPAC value for node-to-node scaling (every 2-3 years):

- (P)erformance: >30% more maximum operating frequency at constant energy
- (P)ower: >50% less energy per switching at a given performance
- (A)rea: >50% area reduction
- (C)ost: <25% wafer cost – 35-40% less die cost for scaled die

Battery limits keep the power consumption budget for the application processor. Unfortunately, this power budget does not improve from node-to-node while on the other hand the amount of logic gates placed in a System-on-Chip (SoC) increases by a factor 2 from node-to-node. Increasing amount of components under constant power budget puts a limit to the usage percentage (utilization) of gates in an SoC. Currently the reduction of supply voltage does not suffice to meet the system requirements for the battery and thermal power envelope of the high-performance application. Therefore, it is necessary to aggressively reduce the supply voltage for maximal amount of functions at iso power.

2 MORE MOORE

The main goals of the ITRS include identifying key technical requirements and challenges critical to sustain the historical scaling of CMOS technology per More Moore (MM) and stimulating the needed research and development to meet the key challenges. The objective of listing and discussing potential solutions in this chapter is to provide the best current guidance about approaches that address the key technical challenges. However, the potential solutions listed here are not comprehensive, nor are they necessarily the most optimal ones. Given these limitations, the potential solutions in the ITRS are meant to stimulate but not limit research exploring novel and different approaches. MM mission is listed below:

MM focus team in ITRS provides physical, electrical and reliability requirements for logic and memory technologies to sustain More Moore (PPAC: power, performance, area, cost) scaling for big data, mobility, and cloud (IoT and server) applications and forecast logic and memory technologies (15 years) in main-stream/high-volume manufacturing (HVM).

In the next section of this chapter we will discuss challenges, roadblocks, and potential solutions. The chapter is subdivided into the following major subsections:

- Ground rules: logic and memory
- Logic core device technologies
- Memory technologies
    - DRAM
    - Non-volatile memories: FLASH and NVM
- Interconnect technologies
- Process integration
    - Front-end processes
    - Lithography
    - Metrology
    - Yield
    - Reliability

1.1 LOGIC TECHNOLOGIES

A major portion of semiconductor device production is devoted to digital logic. In this section, both high-performance logic and low-power logic which is typically for mobile applications are included and detailed technology requirements and potential solutions are considered for both types separately. Key considerations are speed, power, density requirements, and goals. One key theme is continued scaling of the MOSFETs for leading-edge logic technology in order to maintain historical trends of improved device performance. This scaling is driving the industry toward a number of major technological innovations, including material and process changes such as higher-K gate dielectrics and strain enhancement, and in the near future, new structures such as gate-all-around (nanowire) and alternate high-mobility channel materials. These innovations are expected to be introduced at a rapid pace, and hence understanding, modeling, and implementation into manufacturing in a timely manner is expected to be a major issue for the industry.

1.2 MEMORY TECHNOLOGIES - DRAM

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

For DRAM, the main goal is to continue to scale the foot-print of the 1T-1C cell, to the practical limit of $4F^2$. The issues are vertical transistor structures, high-κ dielectrics to improve the capacitance density, and meanwhile keeping the leakage low.

1.3 MEMORY TECHNOLOGIES - NON-VOLATILE MEMORY: FLASH AND EMERGING MEMORIES

The NVM discussion in this chapter is limited to devices that can be written and read many times; hence read-only memory (ROM) and one-time-programmable (OTP) memory are not included although many such memories are important both for standalone and embedded applications. The current mainstream NVM is Flash memory. NAND and NOR flash memories are used for quite different applications – data storage for NAND and code storage for NOR flash. There are serious issues with scaling for both NOR and NAND flash memories that are dealt with at some length in the chapter. Other non-charge-storage types of NVM are also considered, including ferroelectric RAM (FeRAM), magnetic RAM (MRAM), and phase-change RAM (PCRAM), all are in volume production. These emerging memories promise to continue NVM scaling beyond Flash memories. However, because NAND Flash and to some extent NOR Flash are still dominating the applications emerging memories have been used in specialty applications and have not yet fulfilled their original promise to become dominating mainstream high-density NVM. Starting in 2013 edition, resistive memory (ReRAM) is added to the More Moore chapter as a potential solution.

1.4 INTERCONNECT TECHNOLOGIES

Mission of interconnect technologies is to address the wiring system that distributes clock and other signals to the various functional blocks of a CMOS integrated circuit, along with providing necessary power and ground connections. The process scope begins at the contact level with the pre-metal dielectric and continues up to the wirebond pads, describing deposition, etch and planarization steps, along with any necessary etches, strips and cleans. It also assesses reliability and performance includes specifications for electromigration and calculations of delay. Expanded treatment of Emerging Interconnects and 3D integration are also considered by the Interconnect TWG.

1.5 PROCESS INTEGRATION

FRONT-END PROCESSES

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

LITHOGRAPHY

The Lithography Technology Working Group's mission is to identify lithographic options that could enable future semiconductor nodes and better semiconductor products and to describe the driving forces for their implementation and the challenges to their implementation.

METROLOGY

The Metrology Technology Working Group's mission is to identify emerging measurement challenges and describe research and development pathways for meeting them, primarily for extending CMOS, accelerating Beyond CMOS technologies, materials characterization and structure function relationships. Metrology also provides the measurement

capability necessary for cost-effective manufacturing. As such, the metrology chapter of the ITRS focuses on difficult measurement needs, possible solutions, metrology tool development, and standards.

RELIABILITY

Reliability is a critical aspect of process integration. Emerging technology generations require the introduction of new materials and processes at a rate that exceeds current capabilities for gathering and generating the required database to ensure product reliability. Consequently, process integration is often performed without the benefit of extended learning, which will make it difficult to maintain current reliability levels. Uncertainties in reliability can lead to performance, cost, and time-to-market penalties. Insufficient reliability margin can lead to field failures that are costly to fix and damaging to reputation. These issues place difficult challenges on testing and reliability modeling. This chapter discusses many reliability issues. The goal is to identify the challenges that are in need of significant research and development.

YIELD

Critical dimensions of devices and corresponding defect densities continue shrinking, posing new challenges for the detection of defects as well as tolerable contamination. The wafer edges and backside were identified to have a significant impact on yield as well as process variations and design. Also multi-patterning is creating additional challenges. Development of defect detection, defect review, and classification technologies showing highest sensitivity at high throughput is crucial for cost efficient manufacturing. Furthermore for efficient manufacturing the monitoring of contamination in the environment and on the wafer surface requires appropriate analytic capabilities. Automated, intelligent analysis and reduction algorithms, which correlate facility, design, process, electrical and virtual metrology results and their correlation to yield, test and work-in-progress data, will have to be developed to enhance root cause analysis and therefore enable rapid yield learning.

2 DIFFICULT CHALLENGES

The goal of the semiconductor industry is to be able to continue to scale the technology in overall performance. The performance of the components and the final chip can be measured in many different ways; higher speed, higher density, lower power, more functionality, etc. Traditionally, dimensional scaling had been adequate to bring about these aforementioned performance merits but it is no longer so. Processing modules, tools, material properties, etc., are presenting difficult challenges to continue scaling. We have identified these difficult challenges and summarized in Table MM1below. These challenges are divided into near-term 2015-2022 and long-term 2023-2030.

Table MM1: Process integration difficult challenges.

Near-Term 2015-2022	*Summary of Issues*
1. Scaling Si CMOS	• Scaling of fully depleted SOI and multi-gate (MG) structures • Implementation of gate-all-around (nanowire) structures • Controlling source/drain series resistance within tolerable limits • Further scaling of EOT with higher K materials (K > 30) • Threshold voltage tuning and control with metal gate and high-K stack • Inducing adequate strain in advanced structures
2. Implementation of high-mobility CMOS channel materials	• Basic issues same as Si devices listed above • High-K gate dielectrics and interface state (D_{it}) control • CMOS (*n*- and *p*-channel) solution with monolithic material integration • Epitaxy of lattice-mismatched materials on Si substrate • Process complexity and compatibility with significant thermal budget limitations
3. Scaling of DRAM and SRAM	• DRAM— • Adequate storage capacitance with reduced feature size; implementing high-κ dielectrics • Low leakage in access transistor and storage capacitor; implementing buried gate type/saddle fin type FET • Low resistance for bit- and word-lines to ensure desired speed • Improve bit density and lower production cost in driving toward $4F^2$ cell size • SRAM— • Maintain adequate noise margin and control key instabilities and soft-error rate • Difficult lithography and etch issues
4. Scaling high-density non-volatile memory	• Endurance, noise margin, and reliability requirements • Multi-level at < 20 nm nodes and 4-bit/cell MLC • Non-scalability of tunnel dielectric and interpoly dielectric in flash memory – difficulty of maintaining high gate coupling ratio for floating-gate flash • Few electron storage and word line breakdown voltage limitations • Cost of multi-patterning lithography • Implement 3-D NAND flash cost effectively • Solve memory latency gap in systems
5. Reliability due to material, process, and structural changes, and novel applications.	• TDDB, NBTI, PBTI, HCI, RTN in scaled and non-planar devices • Gate to contact breakdown • Increasing statistical variation of intrinsic failure mechanisms in scaled and non-planar devices • 3D interconnect reliability challenges • Reduced reliability margins drive need for improved understanding of reliability at circuit level • Reliability of embedded electronics in extreme or critical environments (medical, automotive, grid...)
Long-Term 2023-2030	• *Summary of Issues*
1. Implementation of advanced multi-gate	• Fabrication of advanced non-planar multi-gate and nanowire MOSFETs to below 10 nm gate length

structures	• Control of short-channel effects • Source/drain engineering to control parasitic resistance • Strain enhanced thermal velocity and quasi-ballistic transport
2. Identification and implementation of new memory structures	• Scaling storage capacitor for DRAM • DRAM and SRAM replacement solutions • Cost effective installation of high density 3-D NAND (512 Gb – 4 Tb) with high layer numbers or tight cell pitch • Implementing non-charge-storage type of NVM cost effectively • Low-cost, high-density, low-power, fast-latency memory for large systems
3. Reliability of novel devices, structures, and materials.	• Understand and control the failure mechanisms associated with new materials and structures for both transistor and interconnect • Shift to system level reliability perspective with unreliable devices • Muon-induced soft error rate
4. Power scaling	• V_{dd} scaling while supplying sufficient current drive • Controlling subthreshold current or/and subthreshold slope • Margin issues for low V_{dd}
5. Integration for functional diversification	• Integration of multiple functions onto Si CMOS platform • 3-D integration

2.1 NEAR-TERM 2015-2022

[1] Scaling of Si CMOS—

Implementation of fully depleted SOI and multi-gate (field-effect limited devices) will be challenging. Since such devices will typically have lightly doped channels, the threshold voltage will not be controlled by the channel doping. Among the most critical will be controlling the thickness and its variability for these ultra-thin bodies, and establishing a cost-effective method for reliably setting the threshold voltage. Threshold-voltage tuning and control with metal gate/high-K gate stacks has proven to be challenging, especially for low-threshold-voltages as V_{dd} continues to go down. This issue will be critical in fully depleted channels such as multi-gate and FDSOI, where the effective work-function needs to be in the bandgap (although at different values for *p*-MOSFETs and *n*-MOSFETs), and where the work-function is especially critical in setting the threshold voltage because of the lack of channel doping as a variable. Furthermore, since multiple threshold voltages are sometimes required, an ability to cost effectively tune the work-function over the bandgap would be very useful.

Additionally for multi-gate structures, the channel surface roughness may present problems in carrier transport and reliability. These issues will be more severe in nanowire structures.

Controlling source/drain series resistance within tolerable limits will be significant issues. Due to the increase of current density, the demand for lower resistance with smaller dimensions at the same time poses a great challenge. This problem becomes even more severe with thin bodies in fully depleted SOI and multi-gate structures, and in the extreme case, nanowire structures. It is estimated that in current technologies, series resistance degrades the saturation current by 1/3 from that of ideal case. This proportion will likely become harder to maintain or worst with scaling.

Metal gate/high-K gate stacks have been implemented in the most recent technology generation in order to allow scaling of the EOT, consistent with the overall transistor scaling while keeping gate leakage currents within tolerable limits. Further scaling of EOT with higher-K materials (K>30) becomes increasingly difficult and has diminishing returns. The reduction or elimination of the SiO_2 interfacial layer has been shown to cause interface states and degradation of mobility and reliability. Another challenge is growing gate dielectrics on vertical surfaces in multi-gate structures. A fundamental burden placed on the overall gate capacitance is the non-scalable quantum capacitance in series with the gate dielectric capacitance.

Enhanced channel-carrier low-field mobility and high-field velocity due to internally applied strain is a major contributor to meeting the MOSFET performance requirements. In inducing adequate strain some current process techniques tend to be less effective with scaling. Also, to apply known techniques derived from planar structure to non-planar structures will be facing additional difficulty and complexity. Moreover, transport enhancement is projected to saturate with strain at some point. (For more detail, see Logic Potential Solutions section.)

[2] Implementation of high-mobility CMOS channel materials—

The basic challenges are similar to that of Si CMOS scaling described above. Following presents additional challenges from these new channel materials.

Growing MOSFET quality oxides on III-V materials has long been an industry goal and struggle. Work on the field has been going on for decades, and success has only started to appear only very recently. Nevertheless, there are still much work to be done in the areas of high-K dielectrics, interface quality, yield, variability, and reliability.

Most III-V materials lack good mobility for *p*-type carriers. In order to provide a CMOS solution, Ge is projected to be a good choice, even though it adds complexity to the whole process (see below). A single channel material for both types of channels would be preferable, and materials other than InGaAs are being researched. Ge CMOS is promising for much higher intrinsic mobility for both *n*- and *p*-type carriers compared to Si, but the *n*-channel implementation has been challenging due to source-drain doping and contact problems. Another possibility is to have strained Si for NMOS and while having SiGe or Ge channel for PMOS.

In order to take advantage of the well-established Si platform, it is anticipated that the new high-mobility materials will be epitaxially grown on Si substrate. The lattice mismatch presents a fundamental challenge in terms of material quality and yield, and a practical challenge in cost.

The reason for the requirement of the high-mobility materials to be grown on Si substrate is not only for the established processing steps, but also for the expectation that Si components will be included in the same chips. Examples of these Si based components are embedded DRAM and nonvolatile memories, active analog devices including power devices, analog passives, and large circuit CMOS blocks that do not require high performance but better yield. Integrating these different materials with different process requirements is a huge challenge. Take as an example to integrate Si CMOS with III-V/Ge CMOS. There would be likely three kinds of high-K dielectrics required. Different kinds of metal gates are also required to provide different work functions to yield the necessary threshold voltages. And all processes have to be compatible with one another in terms of thermal budget.

[3] Scaling of DRAM and SRAM—

For DRAM, a key issue is implementation of high-κ dielectric materials in order to get adequate storage capacitance per cell even as the cell size is shrinking. Also important is controlling the total leakage current, including the dielectric leakage, the storage junction leakage, and the access transistor source/drain subthreshold leakage, in order to preserve adequate retention time. The requirement of low leakage currents causes problems in obtaining the desired access transistor performance. Deploying low sheet resistance materials for word- and bit-lines to ensure acceptable speed for scaled DRAMs and to ensure adequate voltage swing on word-line to maintain margin is critically important**.** The need to increase bit density and to lower production cost is driving toward $4F^2$ type cell, which will require high aspect ratio and non-planar FET structures. Revolutionary solution to have a capacitor-less cell would be highly beneficial.

For SRAM scaling, difficulties include maintaining both acceptable noise margins in the presence of increasing random V_T fluctuations and random telegraph noise, and controlling instability, especially hot-electron instability and negative bias temperature instability (NBTI). There are difficult issues with keeping the leakage current within tolerable targets, as well as difficult lithography and etch process issues with scaling. Solving these SRAM challenges is critical to system performance, since SRAM is typically used for fast, on-chip memory.

[4] Scaling high-density non-volatile memory (NVM)—

For floating-gate devices there is a fundamental issue of non-scalability of tunnel oxide and interpoly dielectric (IPD), and high (> 0.6) gate coupling ratio (GCR) must be maintained to control the channel and prevent gate electron injection during erasing. For NAND Flash, these requirements can be slightly relaxed because of page operation and error code correction (ECC), but IPD < 10 nm still seems unachievable. This geometric limitation will severely challenge scaling far below 20 nm half-pitch. In addition, fringing-field effect and floating-gate interference, noise margin, and few-electron statistical fluctuation for V_t all impose deep challenges. Since NAND half-pitch has pulled ahead of DRAM and logic, lithography, etching, and other processing advances are also first tested by NAND technology.

Charge-trapping devices help alleviate the floating-gate interference and GCR issues, and the planar structure relieves lithography and etching challenges slightly. Recently, high-K IPD and metal gate for planar floating gate Flash memory have been successfully developed and products with 1/2 pitch as small as 16nm have been introduced. Scaling far below 16 nm is still a difficult challenge, however, because fringing-field effects and few-electron V_t noise

margin are still not proven and more important, electric breakdown between adjacent word lines may ultimately restrict word line 1/2 pitch to > 10nm.

Endurance reliability and write/read speed for both devices are still difficult challenges for MLC (multi-level cell) high-density applications.

3-D NAND flash is being developed to build high-density NVM beyond 256 Gb. Cost effective implementation of this new technology with MLC and acceptable reliability performance remains a difficult challenge. Contrary to earlier (2011) projection, actual product introduced in 2013 started with larger cell pitch and high layer numbers. Starting with a large layer number will quickly push the layer numbers in the future nodes to > 100 since each new node needs to double the layers. This will cause additional difficult challenges to processing technology to achieve such structures.

[5] Reliability due to material, process, and structural changes, and novel applications—

In order to successfully scale ICs to meet performance, leakage current, and other requirements, it is expected that numerous major processes and material innovations, such as high-κ gate dielectrics, metal gate electrodes, elevated source/drain, advanced annealing and doping techniques, low-κ materials, etc., are needed. Also, it is projected that new MOSFET structures, starting with ultra-thin body FDSOI MOSFETs and moving on to ultra-thin body, multi-gate MOSFETs, will need to be implemented. Understanding and modeling the reliability issues for all these innovations so that their reliability can be ensured in a timely manner is expected to be particularly difficult.

The first near-term reliability challenge concerns failure mechanisms associated with the MOS transistor. The failure could be caused by either breakdown of the gate dielectric or threshold voltage change beyond the acceptable limits. The time to a first breakdown event is decreasing with scaling. This first event is often a "soft" breakdown. However, depending on the circuit it may take more than one soft breakdown to produce an IC failure, or the circuit may function for longer time until the initial "soft" breakdown spot has progressed to a "hard" failure. Threshold voltage related failure is primarily associated with the negative bias temperature instability (NBTI) observed in *p*-channel transistors in the inversion state. It has grown in importance as threshold voltages have been scaled down. Burn-in options to enhance reliability off end-products may be impacted, as it may accelerate NBTI shifts. Introduction of high-κ gate dielectric may impact both the insulator failure modes (e.g., breakdown and instability) as well as the transistor failure modes such as hot carrier effects, positive and negative bias temperature instability. The replacement of polysilicon with metal gates also impacts insulator reliability and raises new thermo-mechanical issues. The simultaneous introduction of high-κ and metal gate makes it even more difficult to determine and model reliability mechanisms. To put this change into perspective, even after decades of study, there are still issues with silicon dioxide reliability that need to be resolved.

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

There are additional reliability challenges associated with advanced packaging for higher performance, higher power integrated circuits. Increasing power, increasing pin count, and increasing environmental regulations (e.g., lead-free) all impact package reliability. The interaction between the package and die will increase, especially with the introduction of low-K intermetallic dielectrics. The move to multi-chip packaging and/or heterogeneous integration makes reliability even more challenging. As currents increase and the size of balls/bumps decreases, there is an increased risk of failures due to electromigration. Cost cutting forces companies to replace gold bond wires to materials like copper, which poses additional requirements in order to make this as reliable as gold.

ICs are used in a variety of different applications. There are some special applications for which reliability is especially challenging. First, there are the applications in which the environment subjects the ICs to stresses much greater than found in typical consumer or office applications. For example, automotive, military, and aerospace applications subject ICs to extremes in temperature and shock. In addition, aviation and space-based applications also have a more severe radiation environment. Furthermore, applications like base stations require ICs to be continuously on for tens of years at elevated temperatures, which make accelerated testing of limited use. Second, there are important applications (e.g., implantable electronics, safety systems) for which the consequences of an IC failure are much greater than in mainstream IC applications.

At the heart of reliability engineering is the fact that there is a distribution of lifetimes for each failure mechanism. With increasing low failure rate requirements we are more and more interested in the early-time range of the failure time distributions. There has been an increase in process variability with scaling (e.g., distribution of dopant atoms, CMP variations, and line-edge roughness). At the same time the size of a critical defect decreases with scaling. These trends will translate into an increased time spread of the failure distributions and, thus, a decreasing time to first failure. We need to develop reliability engineering software tools (e.g., screens, qualification, and reliability-aware design) that can handle the increase in variability of the device physical properties, and to implement rigorous statistical data analysis to quantify the uncertainties in reliability projections. The use of Weibull and log-normal statistics for analysis of breakdown and electromigration reliability data is well established. However, the shrinking reliability margins require more careful attention to statistical confidence bounds in order to quantify risks. This is complicated by the fact that new failure physics may lead to significant and important deviations from the traditional statistical distributions, making error analysis non-straightforward. Statistical analysis of other reliability data such as BTI and hot carrier degradation is not currently standardized in practice, but may be needed for accurate modeling of circuit failure rate.

2.2 LONG-TERM 2023-2030

[1] Implementation of advanced multi-gate structures—

For the long-term years till the end of current roadmap when the transistor gate length is projected to scale below 10 nm, ultra-thin body multi-gate MOSFETs with lightly doped channels are expected to be utilized to effectively scale the device and control short-channel effects. All other material and process requirements mentioned above, such as high-K gate dielectrics, metal gate electrodes, strained silicon channels, elevated source/drain, etc., are expected to be incorporated. Body thicknesses for both fully depleted SOI and MG below 2 nm are projected and the impact of quantum confinement and surface scattering effects on such thin devices are not well understood. The ultra-thin body also adds additional constraint on meeting the source/drain parasitic resistance requirements. Finally, for these advanced, highly scaled MOSFETs, quasi-ballistic operation with enhanced thermal carrier velocity and injection at the source end appears to be necessary for high current drive. But strain enhancement on these non-planar devices is more difficult.

[2] Identification and implementation of new memory structures—

Increasing difficulty is expected in scaling DRAMs, especially in continued demand of scaling down the foot-print of the storage capacitor. Thinner dielectric EOT utilizing ultra-high-κ materials and attaining the very low leakage currents and power dissipation will be required. A DRAM replacement solution getting rid of the capacitor all together would be a great benefit. The current 6-transistor SRAM structure is area-consuming, and a challenge is to seek a revolutionary replacement solution which would be highly rewarding.

Dense, fast, and low-power non-volatile memory will become highly desirable. Ultimate density scaling may require 3-D architecture, such as vertically stackable cell arrays in monolithic integration, with acceptable yield and performance. 3-D NAND flash will require > 100 layers of stacked devices and processing technology to achieve such structures and cost effective implementation are challenging. Cost effective implementation of non-charge-storage type of NVM is a difficult challenge, and its success may hinge on finding an effective isolation (selection) device. Non-charge-storage NVM may also need to be stacked into 3-D structures to reach Tb density. Without a built-in isolation device as flash memory, the stacking of these two-terminal devices is both costly and difficult. Much innovation is needed to continue increasing storage density to 1 Tb and beyond.

See Emerging Research Devices section for more detail.

[3] Reliability of novel devices, structures, and materials—

The long-term reliability difficult challenge concerns novel, disruptive changes in devices, structures, materials, and applications. For example, at some point there will be a need to implement non-copper interconnect (e.g., optical or, carbon nanotube based interconnects), or tunnel-based FETs instead of classical MOSFETs. For such disruptive solutions there is at this moment little, if any, reliability knowledge (as least as far as their application in ICs is concerned). This will require significant efforts to investigate, model (both a statistical model of lifetime distributions and a physical model of how lifetime depends on stress, geometries, and materials), and apply the acquired knowledge (new built-in reliability, designed-in reliability, screens, and tests). It also seems likely that there will be less-than-historic amounts of time and money to develop these new reliability capabilities. Disruptive materials or devices

therefore lead to disruption in reliability capabilities and it will take considerable resources to develop those capabilities.

[4] Power Scaling—

It is well known that V_{dd} is more difficult to scale than other parameters, mainly because of the fundamental limit of the subthreshold slope of ~60 mV/decade. This trend will continue and become more severe when it approaches the regime of 0.6 V. This fact along with the continuing increase of current density (per area) causes the dynamic power density (proportional to V_{dd}^2) to climb with scaling (although power per transistor is dropping), soon to an unacceptable level. Alternate high-mobility channel materials can provide some relief in this area by allowing more aggressive V_{dd} scaling. On the other hand, for supply voltages lower than ~0.6 V, the circuit margin due to process variability on the threshold voltage needs to be considered.

For high-performance logic, in the trend of increasing chip complexity and increasing transistor on-current with scaling, chip static power dissipation is expected to become particularly difficult to control while at the same time meeting aggressive targets for performance scaling. Innovations in circuit design and architecture for performance and power management (e.g., utilization of parallelism as an approach to improve circuit/system performance, aggressive use of power down of inactive transistors, etc.), as well as utilization of multiple types of transistors (high performance with high leakage and low performance with low leakage) on chip, are needed to design chips with both the desired performance and power dissipation. A trade-off of speed performance for low off-current, or low standby power, is the goal of LP technology.

[5] Integration for functional diversification—

The performance of a chip or technology not only can be measured in speed, density, power, noise, reliability, etc, but also in functionality. There has been an industry trend to include more and more functions on the same chip. Examples are; sensors, MEMS, photovoltaic, energy scavenging, RF and mm-wave devices, etc. Naturally to integrate variety of different materials is a huge challenge. Similarly, integration of high-mobility channel CMOS on Si-based CMOS logic and memories present many challenges as mentioned before.

To improve density on the chip, the trend of the industry is 3-D integration. This induces stress, higher temperature of operation, parasitic capacitances, interference, isolation requirement, process requirements and their compatibility with one another, and device reliability.

3 LOGIC CORE DEVICE TECHNOLOGY

3.1 LOGIC CORE DEVICE REQUIREMENTS

CMOS scaling enabled simultaneous system throughput scaling by concurrent delay, power, and area shrinks with thanks to Moore's law. System scaling is getting more difficult with the limitations in interconnect and bandwidth per power as well as the difficulties and cost of monolithic integration. This requires a holistic approach for an optimal balance of performance and power under the limits of technology.

In the early years before the 130 nm node, transistors enjoyed Dennard scaling where oxide thickness (EOT), transistor length (Lg) and transistor width (W) were scaled by a constant factor in order to provide a delay improvement at constant power density. Nowadays there are numerous input parameters that can be varied, and the output parameters are complicated functions of these input parameters, other sets of projected parameter values (i.e., different scaling scenarios) may be found to achieve the same target. In order to maintain the scaling at low voltages, scaling in recent years focused on additional knobs to boost the performance such as the use of introducing strain to channel, stress boosters, high-k metal gate, lowering contact resistance, and improving electrostatics. This was all done in order to compensate the gate drive loss while supply voltage needs to be scaled down for high-performance mobile applications.

The technology requirements address the MOSFET requirements of both high-performance (HP) and low-power (LP) digital ICs. High-performance logic refers to chips of high complexity, high speed, and relatively high power dissipation, such as microprocessor unit (MPU) chips for desktop PCs, servers, etc. Low-power logic mainly refers to chips for mobile systems, where the allowable power dissipation and hence the allowable off-currents are limited by battery life. The LP technology is similar to the LSTP (low standby power) technology of previous years, and the LOP (low operating power or low dynamic power) technology has been eliminated starting from 2013.

Following metrics were used for device targeting:

- I/CV (transistor intrinsic speed) for performance improvement
- CV^2 for the energy improvement metric
- *Cell-level circuit metrics: standard cell and SRAM cell raw area density, FO3 wireloaded stage delay and power (dynamic, IDDQ).*

For off-current I_{off} we consider the subthreshold leakage current while other leakage currents going through the gate and from the drain junction are assumed smaller so they do not add to this value significantly, although their impact on reliability is another consideration.

For generating the entries in the logic technology requirement tables, an MOSFET modeling software MASTAR has been used for many editions until the ITRS 2013 edition [1]-[4]. The software contains detailed analytical MOSFET models that have been verified against literature data. MASTAR is a compact-model based software, different from finite-element, numerical TCAD programs [1]. Its inputs require calibration from measurements or TCAD on external resistance, short-channel parameters such as SS and DIBL, and transport parameters such as mobility and saturation velocity, and the degree of ballistic transport. Starting by the 2013 edition of the ITRS, we used the support from the NanoHub Team of Purdue University using the Nanohub TCAD tools [5]. This allowed us to understand the fundamental limits of standard scaling scenario where Vdd, Lg, EOT, and channel doping (for planar only) are varied. It was shown that the *n*-channel MOSFET saturation drive current, $I_{d,sat}$, is found to increase only for a few years and then starts to drop (Figure MM2). One of the reasons for the drop of current is mainly due to V_{dd} scaling despite the fact that it was kept moderately scaling to maintain enough inversion charge in the channel. There is also significant source-drain tunneling which comes to the picture for channel lengths below 10 nm. This source-drain tunneling makes the device harder to turn off and increases the subthreshold swing (SS). The tunneling current requires the threshold voltage to be higher to maintain the fixed I_{off}, and consequently leads to a reduction in the inversion charge. This then resulted in drop of performance (I/CV) trend (4%/year), particularly after 2018 (Figure MM2).

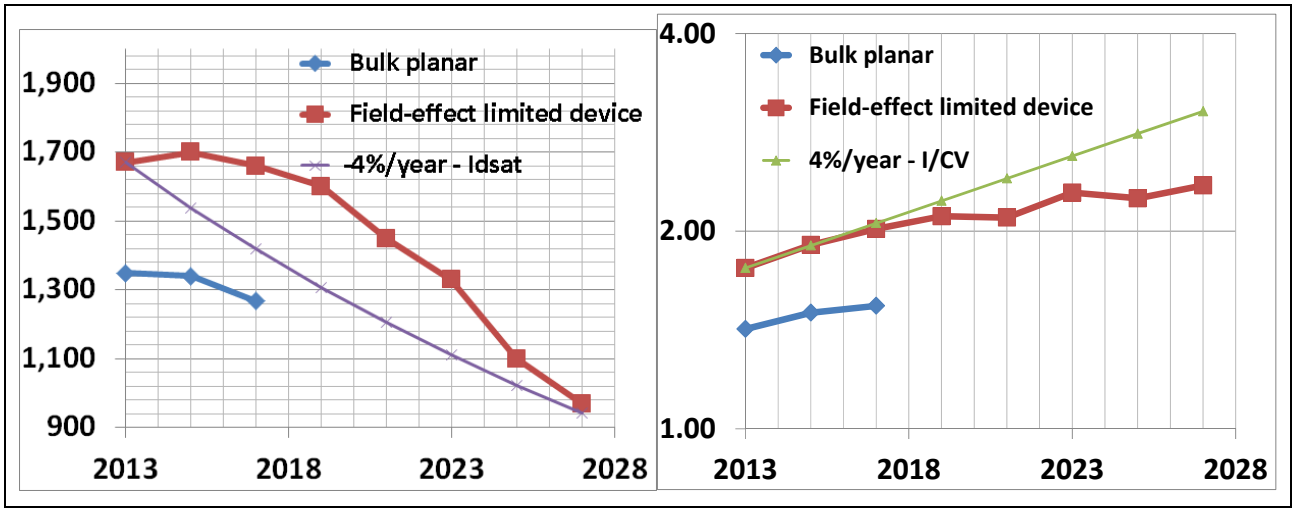


Figure MM2:(left) Idsat (uA/um) and (right) I/CV (ps) scaling trend from the ITRS 2013 edition for bulk planar and field-effect limited devices (e.g. finFET).

NEMO5 is an atomistic quantum transport simulator based on Quantum Transmitting Boundary Method (QTBM) with the nearest-neighbor sp^3d^5s tight-binding (TB) [6] that is used to calculate intrinsic device characteristics in the ballistic regime. The validity of the TB band structure is confirmed via comparison against first-principle electronic structure calculations in ultra-thin body (UTB) silicon for different SOI thicknesses. To capture the scattering effect, the Lundstrom model is used. After calculation of the device ballistic characteristics, backscattering model [7]-[10] is applied using the following equations [7]:

$$T_c = \frac{\lambda}{\lambda + l_{kT}} \quad (2)$$

with high V_{DS}:

$$I_{scattering} = \frac{T_c}{2 - T_c} I_{ballistic} \quad (3)$$

with low V_{DS}:

$$I_{scattering} = T_c \times I_{ballistic} \quad (4)$$

where T_c is the transmission coefficient and l_{kT} is effective on-current channel length, which is the distance between top of the barrier to location of one kT lower [9][11][12]. This value is calculated from ballistic potential profile for each bias point. The mean free path (λ) value is required to include scattering effect by backscattering model. This value is extracted from the experimental reported values [13][14]for different UTB body thicknesses and different charges under the gate. The approach that we used for calculation of mean free path is as below:

$$\lambda = \frac{V_{ds}}{v_{inj}} \mu \quad (5)$$

where V_{ds} is very low, i.e., 5 mV and v_{inj} is calculated by dividing the current by the charge at top-of-the-barrier. μ is mobility which is dependent on the charge under the gate and device body thickness.

For circuit-level transient simulations there is a need for other compact-model based software such as BSIM CMG or open source models such as Virtual Source Model (VSM) from MIT [15]. In the 2015 edition we used VSM models to capture the circuit-level parameters such as delay and power per stage from a ring oscillator. Its inputs were transparently shown in the roadmap table as realizable targets where most were validated in TCAD with the support from the NanoHub Team of Purdue University.

An important speed metric for the transistor is the intrinsic speed *I/CV* where *C* includes the gate capacitance plus the gate fringing capacitances. These fringing capacitances have been found to be larger than the intrinsic capacitance over the channel region. This requires a modeling of parasitic components in the device. As shown in the logic core

technology table, the ratio of total fringing capacitances to the gate capacitance over the channel is assumed to increase with scaling and saturates at 2.0 [16].

3.2 LOGIC CORE DEVICE POTENTIAL SOLUTIONS

The specific set of projected parameter values in each of the tables reflects a particular scaling scenario in which the targeted values for the key outputs are achieved. However, since there are numerous input parameters that can be varied, and the output parameters are complicated functions of these input parameters, other sets of projected parameter values (i.e., different scaling scenarios) may be found to achieve the same target. For example, one technology would scale the EOT more aggressively by introducing higher-K dielectric, while another would achieve equivalent results by optimizing doping or/and strain enhancement. Hence, the scaling scenarios in these tables only constitute a good guidance for the industry but are not meant to be unique solutions, and there will be considerable variance in the actual paths that the various companies will take.

Following knobs are used to scale down Vdd while improving performance and power:

- Reduce parasitics
- Increase drive per footprint by 3D structures
- Improve electrostatics and device isolation
- Transition to new device architectures such as gate-all-around device
- Enhance drive
- Reduce process & material variations

Reducing parasitics: Controlling source/drain series resistance within tolerable limits will become much more difficult. Due to the increase of current density, the demand for lower resistance with smaller dimensions at the same time poses a great challenge. It is estimated that in current technologies, series resistance degrades the saturation current by 40% and more from that of ideal case. This proportion will likely become harder to maintain or worse with the poly pitch scaling and also increasing interconnect resistance by scaling, will all leaving less headroom for the device contact itself. In order to maximize the benefits of high-mobility channels in the drain current, it gets much more important to reduce the contact resistance. Silicide contacts are getting off-stream in maintaining the required reduction of contact resistance with the poly pitch scaling and decreasing channel resistance with improved drive. One promising reduction is achieved by MIS contacts, which utilize an ultra-thin dielectric between the metal and semiconductor interface. This reduces the Fermi level pinning and therefore reduces the Schottky Barrier Height (SBH). This SBH reduction happens by the exponential decay of the metal induced gap states (MIGS) induced charge density in the bandgap of the dielectric.Parasitic capacitance between gate and source/drain terminal of the device is increasing with technology scaling and exceed the channel capacitance as the poly pitch is scaled down. There is a need to focus on low-k spacer materials that still provide good reliability and etch selectivity for S/D contact formation. The intrinsic series resistance of a MOSFET is mainly determined by the S/D doping gradient, S/D sheet resistance, and the metal-semiconductor contact resistivity, all are seldom provided in literature for calibration. So the values for series resistance are not calculated from device structures. Rather, they are given here as a technology target where node-to-node improvement of 15% is required in contact resistivity materials, conformal doping, and contact surface area improvements through elevated source/drain and wrap-around-contact.

Increasing drive per footprint: FinFET and lateral nanowires enable a higher drive at unit footprint (by enabling drive in the third dimension) if fin pitch can be aggressively scaled. This increased drive at unit footprint by scaling the fin pitch comes at a trade-off between fringing capacitance between gate and contact and series resistance.

Improving electrostatics and device isolation: FinFET has better electrostatics integrity due to its tall narrow channel that is controlled by a gate from three-sides where this allows relaxing the scaling requirements of fin thickness (i.e. body thickness) compared to UTBB FDSOI. In UTBB FDSOI electrostatic control could be done by using silicon (i.e. body) thickness and BOX thickness where convergent scaling of both silicon thickness and BOX thickness enables electrostatics scaling (DIBL < 100 mV/V) down to gate lengths towards 10 nm. Junction implantation engineering, EOT scaling, and density of interface traps (*Dit*) reduction are potential solutions to

maintain the electrostatics control in the channel. Besides the channel leakage induced by electrostatics, there are potentially other leakage sources such as sub-fin leakage. This leakage current flows through the bottom part of the fin from source to drain. This gets more problematic in Ge channels because of low effective mass of Ge. Ground plane doping and quantum well below the channel will potentially solve this leakage problem; therefore improving the electrostatics.

As scaling continues, the power density of the IC continues to go up with the transistor density, although the power per transistor goes down. An effective solution would be based on transistor actions that do not depend on the Boltzmann distribution which sets a lower limit of subthreshold slope of 60 mV of gate voltage per decade of channel current. One such conduction mechanism is tunneling. A class of transistor based on this effect is called tunneling FET (TFET) [17]. It is basically a *p-n* junction placed under an MOS gate. With a proper design of the heterojunction under the gate, ultra-low V_{dd} operation is the goal.

Another means to achieve sharp subthreshold slope is by incorporating ferroelectric gate dielectrics in an MOSFET [18]. When the transistor is biased towards the on-condition, the electric field moves the charges within the ferroelectric gate oxide, and that polarization further reduces the threshold voltage, resulting in a higher gate over-drive, as if a higher gate voltage was applied. This internal gain, sometimes called negative capacitance, creates an effect of deeper subthreshold slope. The goal also is operation with ultra-low V_{dd} and low power.

Transition to new device architectures: These extensions to the existing device architectures such as FDSOI and finFET will sustain the same device architecture for 2 or 3 nodes until the end of 2020. Beyond 2020 a transition to gate-all-around (GAA) and potentially to vertical nanowires devices will be needed when there will be no room left for the gate length scale down due to the limits of fin width and contact width. GAA is the ultimate structure in terms of electrostatic control to scale to the shortest possible effective channel length.

The transistor structures considered in this chapter are shown in Figure MM3.

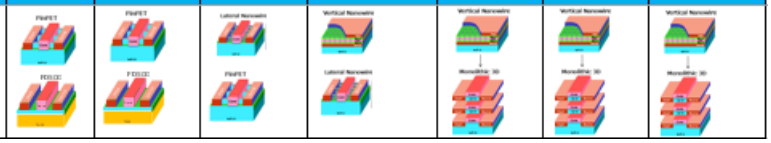
YEAR OF PRODUCTION	2015	2017	2019	2021	2024	2027	2030
Logic device technology naming	P70M56	P48M36	P42M24	P32M20	P24M12G1	P24M12G2	P24M12G3
Logic industry "Node Range" Labeling (nm)	"16/14"	"11/10"	"8/7"	"6/5"	"4/3"	"3/2.5"	"2/1.5"
Logic device structure options	finFET FDSOI	finFET FDSOI	finFET LGAA	finFET LGAA VGAA	VGAA, M3D	VGAA, M3D	VGAA, M3D
							

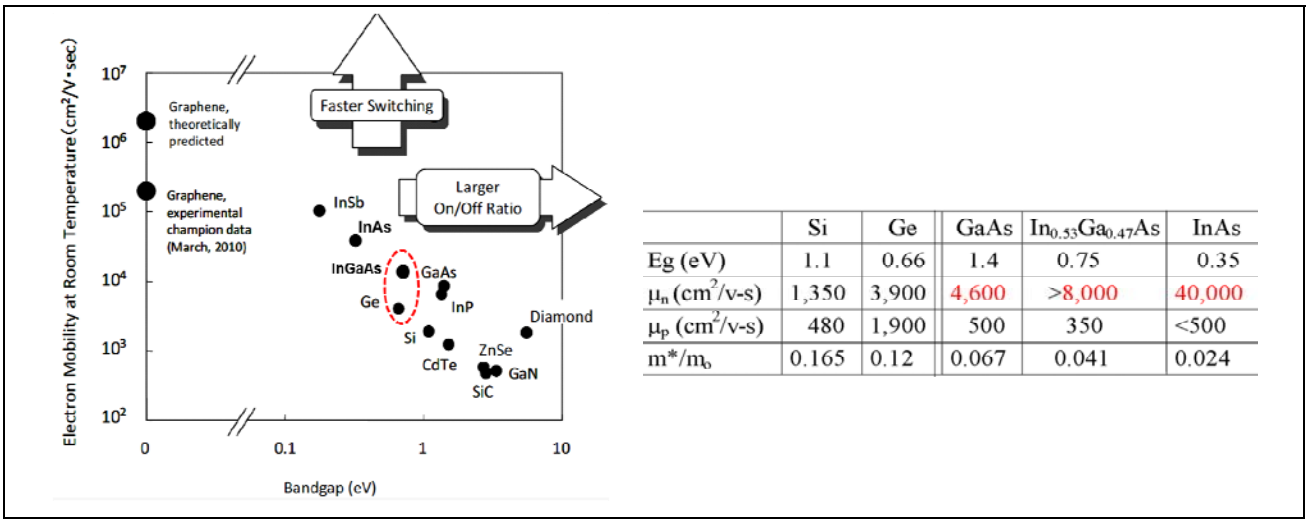
Figure MM3: Transistor structure roadmap: FDSOI, finFET, lateral nanowire, vertical nanowire, and monolithic 3D.

Finally, beyond the roadmap range of this edition (beyond 2030), MOSFET scaling will likely become ineffective and/or very costly. Completely new, non-CMOS type of logic devices and maybe even new circuit architecture are potential solutions (see Emerging Research Devices section for detailed discussions). Such solutions ideally can be integrated onto the Si-based platform to take advantage of the established processing infrastructure, as well as being able to include Si devices such as memories onto the same chip.

Enhancing drive: Eventually later in the roadmap, more forward-looking solutions in the utilization of alternate channel materials to further enhance the transport will be adopted. It is anticipated the first solutions would be III-V (for *n*-channel) and Ge (for *p*-channel) combination, still based on MOSFET operation. High-mobility materials (Ge, IIIV) bring promise in increasing drive current by means of an order of magnitude increase in intrinsic mobility (Figure MM4) [19]. With the scaling in gate length, the impact of mobility of drain current becomes limited because of the velocity saturation. On the other hand whenever gate length further scales down, the carrier transport becomes ballistic. This allows velocity of carriers, which is the so-called injection velocity, scale with the mobility increase. Having drain current mostly ballistic increases the injection velocity because of lower effective mass, therefore results in increase of the drain current. However, low effective mass for the high mobility device can actually bring high tunneling current at higher supply voltage. This may degrade performance of III-V devices at short channel after work

function tuning (e.g. threshold voltage increase) to lower Ioff to compensate the tunneling current. Another consideration for high mobility channel is the lower density of states. The current is proportional to the multiplication of drift velocity and carrier concentration in the channel. This requires correct selection of Lg, Vdd, and device architecture in order to maximize this multiplication, where the selection of those parameters will be different for the type of channel material used [20]. This all needs to be holistically tackled. A shift in the centroid of charge away from the gate potential adds to the equivalent oxide thickness (EOT), reducing the inversion capacitance, particularly in IIIV high-mobility channels. Despite the fact that drive current of IIIV might not be that high, the overall delay merit (CV/I) can result better than the ones of Si and other high-mobility channels (e.g. Ge). Strain engineering is an additional knob to boost mobility on top of high-mobility channels. In fact this knob has been used as one of the most effective knobs in the last decade (Figure MM5) [19]. With the scaling down of contacted poly pitch, SiGe on the S/D EPI contact and strain relaxation buffer (SRB) remain as effective boosters to scale mobility more than double on top of high-mobility channel material [21].

Other possibilities beyond these semiconductors are 2-D crystals. These include grapheme, boron nitride (BN), dichalcogenides such as MoS_2, WS_2, $NbSe_2$, and complex oxides such as $Bi_2Sr_2CaCu_2O_x$.



	Si	Ge	GaAs	$In_{0.53}Ga_{0.47}As$	InAs
Eg (eV)	1.1	0.66	1.4	0.75	0.35
μ_n (cm^2/v-s)	1,350	3,900	4,600	>8,000	40,000
μ_p (cm^2/v-s)	480	1,900	500	350	<500
m^*/m_o	0.165	0.12	0.067	0.041	0.024

Figure MM4: Intrinsic mobility of different materials [19].

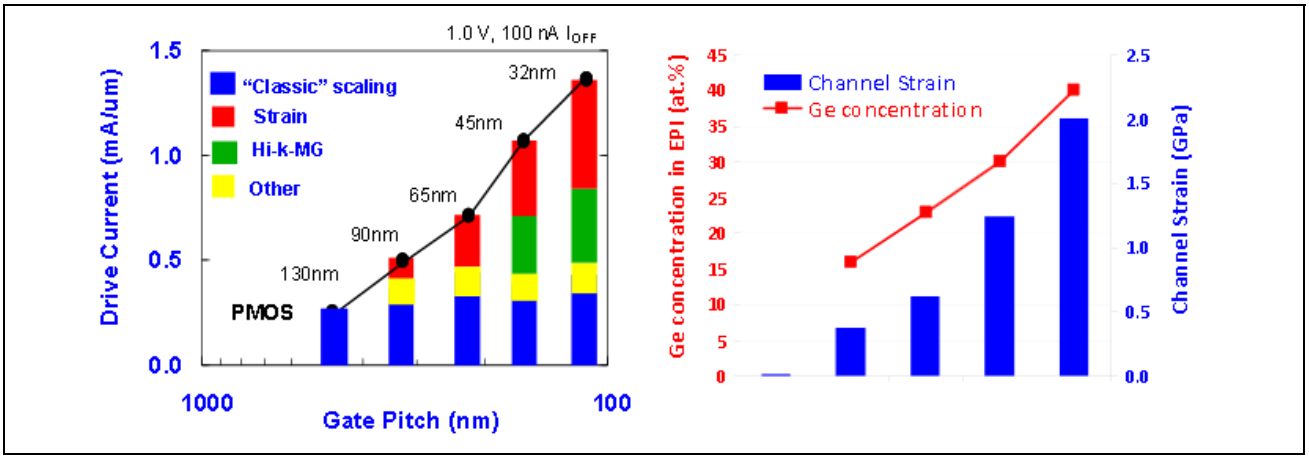


Figure MM5: Impact of strain engineering on device performance [19].

Reducing process and material variations: Reducing variability would further allow Vdd scaling. Controlling channel length and channel thickness are important to maintain the electrostatics in the channel. This would require for instance controlling the profile of the fin and lithography processes to reduce the CD uniformity (CDU), line width roughness (LWR), line edge roughness (LER). Dopant-free channel and low-variability work-function metals would

variations in the threshold voltage. With the introduction of high-mobility materials gate stack passivation is needed to reduce the interface related variations as well as maintaining the electrostatics and mobility.

3.3 REFERENCES

[1] MASTAR tool (*http://www.itrs.net/Links/2011ITRS/MASTAR2011/*) and downloading/installation instructions (*http://www.itrs.net/Links/2011ITRS/MASTAR2011/MASTARDownload.htm*).

[2] T. Skotnicki, G. Merckel, and T. Pedron, "A new punchthrough current model based on the voltage-doping transformation," *IEEE Tr. Electron Devices*, vol. 35, no. 7, pp. 1076–1086, June 1988.

[3] T. Skotnicki, C. Denat, G. Merckel, and B. Hennion, "A new analog/digital CAD model for sub-half micron MOSFETs," *IEEE Int. Electron Devices Meeting*, pp. 165–168, December 1994.

[4] T. Skotnicki and F. Boeuf, "CMOS technology roadmap – Approaching up-hill specials," in *Proc. 9th Int. Symp. On Silicon Materials Science and Tech.*, Editors H.R. Huff, L. Fabry, S. Kishino, pp. 720–734, ECS. Vol. 2002-2.

[5] Nanohub website (*http://nanohub.org*) and ITRS tools on Nanohub (*https://nanohub.org/tools/itrs/*).

[6] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, "Atomistic simulation of nanowires in the $sp^3d^5s^*$ tight-binding formalism: from boundary conditions to strain calculations," *Physical Review B*, vol. 74, no. 20, p. 205323, 2006.

[7] R. Kim and M. S. Lundstrom, "Physics of carrier backscattering in one-and two-dimensional nanotransistors," *IEEE Tr. Electron Devices*, vol. 56, no. 1, pp. 132-139, 2009.

[8] C. Jeong, D. A. Antoniadis, and M. S. Lundstrom, "On backscattering and mobility in nanoscale silicon mosfets," *IEEE Tr. Electron Devices*, vol. 56, no. 11, pp. 2762-2769, 2009.

[9] M. Lundstrom, "Elementary scattering theory of the Si mosfet," *IEEE Electron Device Letters*, vol. 18, no. 7, pp. 361-363, 1997.

[10] P. Palestri, D. Esseni, S. Eminente, C. Fiegna, E. Sangiorgi, and L. Selmi, "Understanding quasi-ballistic transport in nano-mosfets: part i-scattering in the channel and in the drain," *IEEE Tr. Electron Devices*, vol. 52, no. 12, pp. 2727-2735, 2005.

[11] K. Natori, "Ballistic mosfet reproduces current-voltage characteristics of an experimental device," *IEEE Electron Device Letters,* vol. 23, no. 11, pp. 655-657, 2002.

[12] P. Palestri, R. Clerc, D. Esseni, L. Lucci, and L. Selmi, "Multi-subband-montecarlo investigation of the mean free path and of the kt layer in degenerated quasi ballistic nano MOSFETs," *IEEE Int. Electron Devices Meeting,* pp. 1-4, 2006.

[13] A. Khakirooz, K. Cheng, A. Reznicek, T. Adam, N. Loubet, H. He, J. Kuss, J. Li, P. Kulkarni, S. Ponoth, et al., "Scalability of extremely thin soi (etsoi) mosfets to sub-20-nm gate length," *IEEE Electron Device Letters*, vol. 33, no. 2, pp. 149-151, 2012.

[14] T. Hiramoto, G. Tsutsui, K. Shimizu, and M. Kobayashi, "Transport in ultrathin-body soi and silicon nanowire mosfets," in *IEEE International Semiconductor Device Research Symp.,* pp. 1-2, 2007.

[15] A. Khakirooz, O.M. Nayfeh, and D. Antoniadis, "A Simple Semi-empirical short-channel MOSFET current–voltage model continuous across all regions of operation and employing only physical parameters," *IEEE Tr. Electron Devices*, Vol. 56, No. 8, pp. 1674-1680, July 2009.

[16] J. Lacord, G. Ghibaudo, and F. Boeuf, "Comprehensive and accurate parasitic capacitance models for two- and three-dimensional CMOS device structures," *IEEE Tr. Electron Devices*, V. 59, No. 5, p. 1332, 2012.

[17] U. Avci and I. Young, "Heterojunction TFET scaling and resonant-TFET for steep subthreshold slope at sub-9nm gate-length," *IEEE Int. Electron Devices Meeting,* pp. 4.3.1-4.3.4, December 2013.

[18] A. Khan, C. Yeung, C. Hu, and S. Salahuddin, "Ferroelectric negative capacitance MOSFET: capacitance tuning and anti-ferroelectric operation," *IEEE Int. Electron Devices Meeting*, pp. 255-258, December 2011.

[19] K. Kuhn, A. Murthy, R. Kotlyar, and M. Kuhn. "Past, present and future: SiGe and CMOS transistor scaling," *Electrochemical Society Tr.*, Vol. 33, No. 6, pp. 13-17, 2010.

[20] T. Skotnicki, F. Boeuf, "How can high mobility channel materials boost or degrade performance in advanced CMOS," *Digest of Technical Papers VLSI Tech. Symp.*, pp. 153-154, June 2010.

[21] G. Eneman, G. Hellings, A. De Keersgieter, N. Collaert, and A. Thean, "Quantum barriers and ground-plane isolation: a path for scaling bulk-finFET technologies to the 7nm node and beyond," *IEEE Int. Electron Devices Meeting*, pp. 12.3.1-12.3.4, December 2013.

4 MEMORY TECHNOLOGIES - DRAM

4.1 DRAM TECHNOLOGY REQUIREMENTS

In general, technical requirements for DRAMs become more difficult with scaling. In the past several of years, DRAM was introduced with many new technologies (e.g. 193 nm argon fluoride (ArF) immersion high-NA lithography with double patterning technology, improved cell FET technology including fin type transistor [1]-[3], buried word line/cell FET technology [4] and so on). Due to new technologies, DRAM will continue to scale with 2-3 year cycle and 20 nm HP (minimum feature size) DRAM will be available by 2017.

Of course, there are still plenty of technical challenges and also the issue of process step increase to sustain the cost scaling. Fundamentally, there exist several significant process flow issues from a production standpoint, such as process steps of capacitor formation, or high aspect ratio contact etches requiring photoresists with hard mask pattern transferring layer that can stand up for a prolonged etch time. Furthermore, continuous improvements in lithography/hard mask and etch will be needed. Also lower WL/BL resistance is necessary for getting the same or better performance.

Although 3-D type cell FETs like saddle-fin FETs are introduced and have revolutionized the one transistor-one capacitor (1T-1C) cell, it is getting more 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. To optimize these operation windows, fully depleted type FET device (like a surrounded gate)will be needed to increase Cell Ion current and also low-k filling or air-gap filling will be needed to reduce the cross coupling capacitance (i.e. BL(Bitline)-BL, BL-SNC(Storage node contact)).[6] Another challenge is a highly reliable gate insulator. A highly boosted gate voltage is required to drive higher drain current with the relatively high threshold voltage adopted for the cell FET to suppress the subthreshold leakage current. The scaling of the DRAM cell FET dielectric, maximum word-line (WL) level, and the electric field in the cell FET dielectric are critical points for gate insulator reliability concern. To keep the electric field to a sustainable level in the dielectric with scaling, process requirements for DRAMs such as front-end isolation, recess-FET formation, conformal oxidation process, gate filling process, and damageless recess process are all needed for future high-density DRAMs.

4.2 DRAM POTENTIAL SOLUTIONS

Since the DRAM storage capacitor gets physically smaller with scaling, the EOT must scale down sharply to maintain adequate storage capacitance. To scale the EOT, dielectric materials having high relative dielectric constant (κ) will be needed. Therefore MIM (metal-insulator-metal) capacitors have been adopted using high κ ($ZrO2/Al_2O/ZrO2$) [5] as the capacitor of 40-30's nm half-pitch DRAM. And this material evolution and improvement are continued until 20 nm HP and ultra high-K (perovskite $\kappa > 50 \sim 100$) material will be released in 2016. Also, the physical thickness of the high-κ insulator should be scaled down to fit the minimum feature size. Due to that, capacitor 3-D structure will be changed from cylinder to pillar shape.

On the other hand, with the scaling of peripheral CMOS devices, a low-temperature process flow is required for process steps after formation of these devices. This is a challenge for DRAM cell processes which are typically constructed after the CMOS devices are formed, and therefore are limited to low-temperature processing. DRAM peripheral device requirement can relax I_{off} but demands more I_{on} of LSTP device. But, in the future, high-κ metal gate will be needed for sustaining the performance [7].

The other big topic is $4F^2$ cell migration. As the half-pitch scaling become very difficult, it is impossible to sustain the cost trend. The most promising way to keep the cost trend and increasing the total bit output by generation is changing the cell size factor (a) scaling (where a = [DRAM cell size]/[DRAM half pitch]^2). Currently $6F^2$ (a = 6) is the majority. To migrate $6F^2$ to $4F^2$ cell is very challenging. For example, vertical cell transistor must be needed but still a couple of challenges are remaining.

All in all, maintaining sufficient storage capacitance and adequate cell transistor performance are required to keep the retention time characteristic in the future. And their difficult requirements are increasing to continue the scaling of DRAM devices and to obtain the bigger product size (i.e. >16 Gb). In addition to that, if efficiency of cost scaling

become tremendously low in comparison with introducing the new technology, DRAM scaling will be stopped and go to 3D cell stacking structure like as 3D-NAND. Or new DRAM concept will be adopted. 3D cell stacking and new concept DRAM are discussed but there is no clear path for further scaling beyond the 2D DRAM.

4.3 REFERENCES

[1] J. Y. Kim et al., "The breakthrough in data retention time of DRAM using Recess-Channel-Array Transistor(RCAT) for 88 nm feature size and beyond", *Symp. VLSI Technology Digest of Technical Papers,* p.11, 2003.

[2] J. Y. Kim et al., "S-RCAT (sphere-shaped-recess-channel-array transistor) technology for 70nm DRAM feature size and beyond", *Symp. VLSI Technology Digest of Technical Papers,* p.34, 2005.

[3] Sung-Woong Chung et al., "Highly Scalable Saddle-Fin (S-Fin) Transistor for Sub-50 nm DRAM Technology", *Symp. VLSI Technology Digest of Technical Papers,* p.32, 2006.

[4] T. Schloesser et al., "$6F^2$ buried wordline DRAM cell for 40 nm and beyond", *IEDM Technical Digest*, p. 809, 2008.

[5] Deok-Sin Kil et al., "Development of New TiN/ZrO2/Al2O3/ZrO2/TiN Capacitors Extendable to 45nm Generation DRAMs Replacing HfO2 Based Dielectrics", *Symp. VLSI Technology Digest of Technical Papers,* p.38, 2006.

[6] J.M. Park et al., "20nm DRAM: A new beginning of another revolution", IEDM Technical Digest, p. 676, 2015.

[7] Minchul Sung et al, "Gate-first High-k/Metal Gate DRAM Technology for Low Power and High Performance Products", IEDM Technical Digest, p. 680, 2015.

5 MEMORY TECHNOLOGIES – NON-VOLATIVE MEMORY (NVM)

5.1 NON-VOLATIVE MEMORY TECHNOLOGY REQUIREMENTS

Non-volatile memory (NVM) consists of several intersecting technologies that share one common trait – non-volatility. The requirements and challenges differ according to the applications, ranging from RFIDs that only require Kb of storage to high-density storage of hundreds of Gb in a chip. Nonvolatile memory may be divided into two large categories—Flash memories (NAND Flash and NOR Flash), and non-charge-storage memories. Nonvolatile memories are essentially ubiquitous, and a lot of applications use embedded memories that typically do not require leading edge technology nodes. The More Moore nonvolatile memory tables only track memory challenges and potential solutions for leading edge standalone parts. Embedded memories are discussed in a separate section.

Flash memories are based on simple one transistor (1T) cells, where a transistor serves both as the access (or cell selection) device and the storage node. Up to now Flash memory serves more than 99% of applications. Since Flash memories are reaching its scaling limit, several non-conventional non-volatile memories that are not based on charge storage (Ferroelectric or FeRAM, Magnetic or MRAM, phase-change or PCRAM, and resistive or ReRAM) form the category of often called "emerging" memories. These memory elements (the storage node) usually have a two-terminal structure (e.g. resistor or capacitor) thus are difficult to also serve as the cell selection device. The memory cell generally combines a separate access device in the form of 1T-1C, 1T-1R, or 1D-1R.

Information on each technology is organized into three categories. The trend tabulation for each technology first treats the issue of packing density. The applicable feature size "F" is identified and the expected area factor "a" is given (cell size in terms of the number of F^2 units required). Second, non-geometrical means to increase packing density (thus lower cost), such as multilevel cell (MLC) or memory layer stacking are addressed. Third, the endurance (erase-write cycle or read-write cycle) ratings and the retention ratings are presented. Endurance and retention are requirements unique to NVM technologies and determine whether the device has adequate utility to be of interest to an end customer.

Roadmap tables show technology trends for NAND Flash and non-charge-storage memories for 2015 through 2030. Conventional 2D NAND Flash is facing scaling limitations, including the number of electrons per logic level and breakdown voltage between neighboring word lines, thus the 1/2 pitch (F) scaling falls flat after around 2018. Stacking of NAND devices in 3D is one way to breakthrough the 2D scaling bottleneck, however although relaxing the x-y footprint can improve reliability the larger "cell size" requires many 3D layers to continue the cost-per-bit scaling trend.

NOR Flash continues to serve embedded systems applications, but higher density parts are rapidly eroded by NAND Flash and consequently essentially stopped further scaling. Therefore, NOR Flash trend is no longer tracked.

Ferroelectric memory (FeRAM) serves limited specialty low-power market and is slowly evolving, therefore requires no significant updating in 2015.

Phase change memory (PCRAM) has gone through several NOR-Flash-like products in the last few years but failed to become widely adopted, thus is no longer tracked as a NOR/NAND Flash application. However, recently a new 3D cross point memory (3D XP) that is speculated based on phase change material threshold switching (Ovonic threshold switching, OTS) has emerged. This new 3D XP memory combines DRAM-like performance and lower than DRAM cost thus serves a storage-class-memory (SCM) function in the memory hierarchy that improves system performance at relatively low cost. Although the array selector is speculated an OTS device, the storage element may be either phase change memory (PCM) or resistive memory (ReRAM). In this chapter, instead of PCRAM, 3D XP trends are tracked.

Field-switching magnetic memory (MRAM) serves a limited specialty market, and there is consensus that the technology cannot scale below about 65nm node, thus is no longer tracked. Spin-transfer-torque memory (STT-MRAM) has seen much focused effort in the last few years, but has developed no standalone product using advanced node technology. The focus on STT-MRAM has shifted to embedded memory application, thus is no longer tracked in this chapter.

5.2 NON-VOLATILE MEMORY POTENTIAL SOLUTIONS

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

5.2.1 NAND FLASH MEMORY

2D FLOATING GATE NAND FLASH

Floating gate Flash devices achieve non-volatility by storing and sensing the charge stored "in" (on the surface of) a floating gate. The conventional memory transistor vertical stack consists of a refractory polysilicon or metal control gate, an interpoly dielectric (IPD) that usually consists of triple oxide-nitride-oxide (ONO) layers, a polysilicon floating gate, a tunnel dielectric, and the silicon substrate. The tunnel dielectric must be thin enough to allow charge transfer to the floating gate at reasonable voltage levels and thick enough to avoid charge loss when in read or off modes. The gate coupling ratio (GCR), defined as the capacitance ratio of the control gate to floating gate capacitor to the total floating gate capacitance (control gate to floating gate + floating gate to substrate), is a critical parameter for proper function (to ensure sufficient percent of voltage drops across the tunnel oxide during program and erase operations) of the device, and must be ≥ 0.6. In most structures, to achieve a GCR ≥ 0.6, the control gate (word line) needs to wrap around the sidewall of the floating gate to provide extra capacitance.

A NAND Flash cell consists of a single MOS transistor, serving both as the cell selection and as the storage device. The NAND array consists of bit line strings of now 64 devices or more with a selection device at each end. This architecture requires no direct bit line contact to the cell, thus allows the smallest cell size (4F2, or four features square). During programming or reading, the unselected cells in the selected bit line string must be turned on and serve as "pass" devices, thus the data stored in each device cannot be accessed randomly. Data input/output are structured in "page" mode where a page (on the Word line) is of several KB (now 8KB – 16KB) in size. Both programming and erasing are by Fowler-Nordheim tunneling of electrons into and out of the floating gate through the tunneling oxide. The low Fowler-Nordheim tunneling current allows the simultaneous programming of many bits (page), thus gives high programming throughput, suitable for handling large amount of data.

NAND Flash scaling faces several difficult challenges. (1) At < 20nm node the gap between neighboring floating gates (~ 20nm) becomes too narrow to accommodate ONO (~ 11nm x2) and still leaves some room for the control gate to wrap around the floating gate to produce high enough GCR. This may be helped by making thinner floating gates, and may stretch scaling down to 16nm – 15nm nodes. Beyond that, planar cells using high-K dielectric instead of SiN and metal gate must be used to achieve the GCR. (2) FG to FG cross talk is a difficult challenge for MLC applications. This may be partially relieved by using air gap insulation. (3) WL to WL breakdown is an even stronger limitation for scaling and can only be partially mitigated by reducing the program/erase voltage through using HK/MG structures. (4) Too few storage electrons for MLC – when each logic level is defined by < 10 electrons, statistical fluctuation blurs the logic levels. Consequently, 2D NAND Flash is expected to not scale below 10nm node, and the current roadmap shows scaling stops at ~ 12nm after around 2018.

3D NAND FLASH

When the number of stored electrons reaches statistical limits, even if devices can be further scaled and smaller cells achieved, the threshold voltage distribution of all devices in the memory array will become uncontrollable and logic states unpredictable. Thus memory density cannot be increased indefinitely by continued scaling of charge-based devices. However, density increase may continue by stacking memory layers vertically.

However, the economy of stacking by completing one device layer then another and so forth is questionable. As depicted in Figure MM6 [1], the cost per bit starts to rise after stacking several layers of devices. Furthermore, the decrease in array efficiency due to increased interconnection and yield loss from complex processing may further reduce the cost-per-bit benefit of this type of 3D stacking.

In 2007, a "punch and plug" approach is proposed to fabricate the bit line string vertically to simplify the processing steps dramatically [1]. This approach makes 3D stacked devices in a few steps and not through repetitive processing, thus promises a new low cost scaling path to NAND flash. Figure MM7 illustrate one such approach. Originally coined BiCS, or Bit Cost Scalable, this architecture turns the NAND string by 90 degrees from a horizontal position to vertical. The word line (WL) remains in the horizontal planes. As depicted in Fig. 1, this type of 3D approach is much more economical than the stacking of complete devices, and the cost benefit does not saturate up to quite high number of layers.

Various architectures for low cost 3D NAND have been proposed since BiCS, all employing the same principle of making all devices in a few simple operations [2]-[6]. These approaches may be put into three large categories: vertical channel, vertical gate, and floating gate. In August of 2013 the first 3D NAND product using one of the low cost approaches is introduced. All major NAND Flash suppliers have announced plans to introduce 3D NAND products using various 3D architectures.

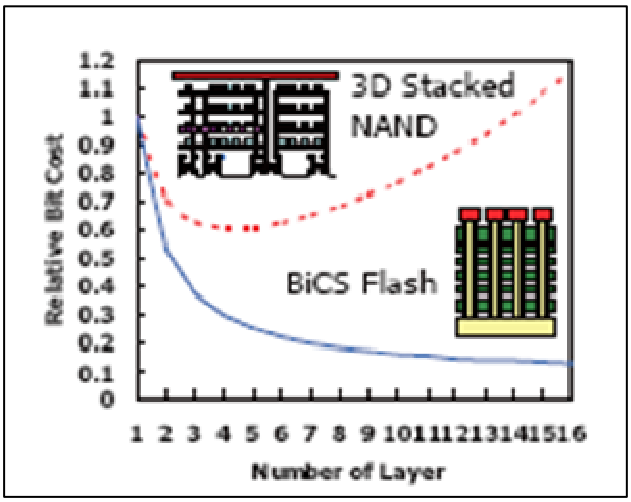


Figure MM6: Comparison of bit cost between stacking of layers of completed NAND devices and making all devices in every layer at once [1].

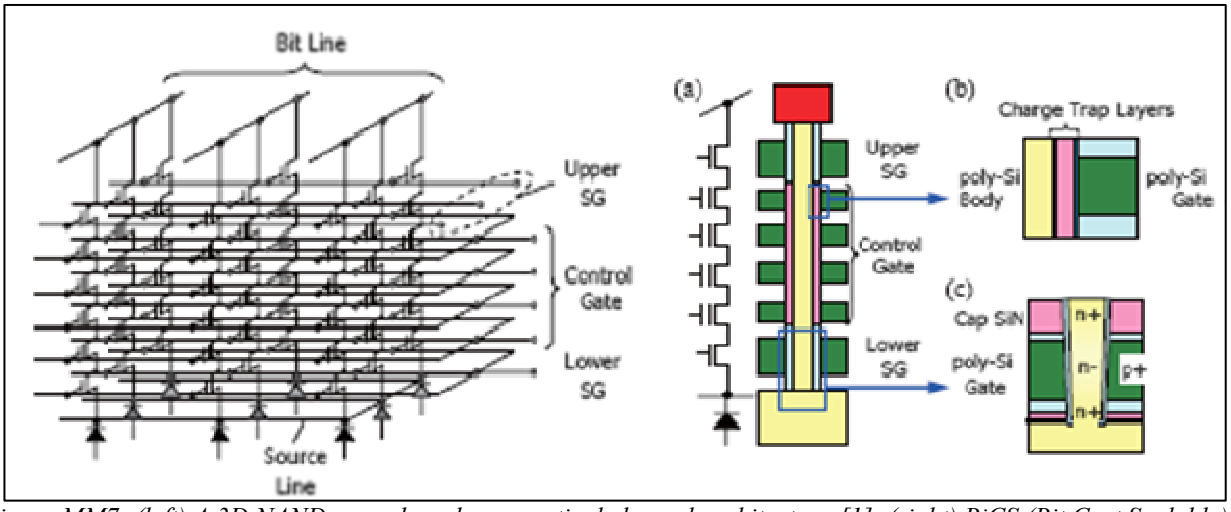


Figure MM7: (left) A 3D NAND array based on a vertical channel architecture [1]. (right) BiCS (Bit Cost Scalable) – a 3D NAND structure using a punch and plug process [1].

2D NAND TO 3D NAND TRANSITION

3D NAND is a logical extension of equivalent scaling after 2D NAND reaches its scaling limit, somewhere near 10nm node. This path, however, might not be followed by every device manufacturer for various reasons.

3D structures achieve high density by increasing the layers and thus circumvent the few-electrons and word line breakdown limitations, thus the 1/2 pitch is not aggressively scaled. In fact, 3D NAND structure faces even worse geometrical limitation 2D floating gate device suffers. In the BiCS approach, the channel hole must be large enough to accommodate twice the thickness of ONO and yet leave enough room for the Si channel material. In addition each layer must be contacted separately thus incurs overhead and additional processing cost. Current 3D products use 4X-5X larger x-y pitch of leading edge 2D NAND products, thus requiring high layer numbers (32+) to compensate for the larger x-y footprint. Consequently, bit cost for 3D NAND has not achieved parity with 2D NAND yet.

Manufacturing of 2D NAND below 20nm nodes requires intensive investment in manufacturing tools for quadruple patterning. Much of this investment is not used when switching to 3D NAND later on since the pitch for 3D is considerably larger. Meanwhile, DRAM technology 1/2 pitch seems also stuck at ~ 20nm, thus investment in fine line patterning capacity may only have limited returns. 3D NAND, on the other hand, requires many layer deposition tools and high aspect ratio etching tools, and tight pitch metal tools that are not compatible with 2D NAND. Therefore, the choice between further scaling 2D and early introduction of 3D is not straightforward. However, since 2D NAND has reached its scaling limit the transition to 3D NAND is inevitable.

5.2.3 NON-CHARGE-BASED NON-VOLATILE MEMORIES

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

One disadvantage of this category of devices is that the storage element itself cannot also serve as the memory selection (access) device because they are mostly two-terminal devices. Even if the On/Off ratio is high a memory cannot be constructed entirely by two terminal devices since the devices in the "On" state would form leakage paths. Therefore, these devices use 1T-1C (FeRAM), 1T-1R (MRAM, PCRAM and ReRAM) or 1D-1R (PCRAM and ReRAM) structures. It is thus challenging to achieve small (4F2) cell size without innovative access device. In addition, because of the more complex cell structure that must include a separate access device, it is more difficult to design 3D arrays that can be fabricated using just a few additional masks like those proposed for 3D NAND.

FeRAM

FeRAM devices achieve non-volatility by switching and sensing the polarization state of a ferroelectric capacitor. To read the memory state the hysteresis loop of the ferroelectric capacitor must be traced and the stored datum is destroyed and must be written back after reading (destructive read, like DRAM). Because of this "destructive read," it is a challenge to find ferroelectric and electrode materials that provide both adequate change in polarization and the necessary stability over extended operating cycles. The ferroelectric materials are foreign to the normal complement of CMOS fabrication materials, and can be degraded by conventional CMOS processing conditions.

FeRAM is fast, low power, and low voltage and thus is suitable for RFID, smart card, ID card, and other embedded applications. Processing difficulty and high cost (compared to FLASH memories) limit wider adoption.

MRAM

MRAM devices employ a magnetic tunnel junction (MTJ) as the memory element. An MTJ cell consists of two ferromagnetic materials separated by a thin insulating layer that acts as a tunnel barrier. When the magnetic moment of

one layer is switched to align with the other layer (or to oppose the direction of the other layer) the effective resistance to current flow through the MTJ changes. The magnitude of the tunneling current can be read to indicate whether a ONE or a ZERO is stored. Field switching MRAM probably is the closest to an ideal "universal memory" since it is non-volatile and fast and can be cycled indefinitely, thus may be used as NVM as well as SRAM and DRAM. However, producing magnetic field in an IC circuit is both difficult and inefficient. Nevertheless, field switching MTJ MRAM has successfully been made into products. The required magnetic field for switching, however, increases when the storage element scales while electromigration limits the current density that can be used to produce higher H field. Therefore, it is expected that field switch MTJ MRAM is unlikely to scale beyond 65nm node and this device is no longer tracked in the NVM table

Recent advances in "spin-transfer torque (STT)" approach where a spin-polarized current transfers its angular momentum to the free magnetic layer and thus reverses its polarity without resorting to an external magnetic field offer a new potential solution. During the spin transfer process, substantial current passes through the MTJ tunnel layer and this stressing may reduce the writing endurance. Upon further scaling the stability of the storage element is subject to thermal noise, thus perpendicular magnetization materials are projected to be needed at 32nm and below. New materials for perpendicular magnetization are still being researched, and are discussed in the ERM chapter. Perpendicular magnetization has been recently demonstrated.

With rapid progress of NAND Flash and the recent introduction of 3D NAND that promises to continue the equivalent scaling, the hope of STT-MRAM to replace NAND seems remote. However, its SRAM-like performance and much smaller footprint than the conventional 6T-SRAM have gained much interest in that application, especially in mobile devices which do not require high cycling endurance as in computation. Therefore, STT-MRAM is now mostly considered not as a standalone memory but an embedded memory, and is not tracked in the standalone NVM table.

PCRAM and Cross Point Memory

PCRAM devices use the resistivity difference between the amorphous and the crystalline states of chalcogenide glass (the most commonly used compound is $Ge_2Sb_2Te_5$, or GST) to store the logic levels. The device consists of a top electrode, the chalcogenide phase change layer, and a bottom electrode. The leakage path is cut off by an access transistor (or diode) in series with the phase change element. The phase change write operation consists of: (1) RESET, for which the chalcogenide glass is momentarily melted by a short electric pulse and then quickly quenched into amorphous solid with high resistivity, and (2) SET, for which a lower amplitude but longer pulse (usually >100ns) anneals the amorphous phase into low resistance crystalline state. The 1T-1R (or 1D-1R) cell is larger or smaller than NOR Flash, depending on whether MOSFET or BJT (or diode) is used, and the device may be programmed to any final state without erasing the previous state, thus provides substantially faster programming throughput. The simple resistor structure and the low voltage operation also make PCRAM attractive for embedded NVM applications. The major challenges for PCRAM are the high current (fraction of mA) required to reset the phase change element, and the relatively long set time. Since the volume of phase change material decreases rapidly with each technology generation, there is hope both above issues become easier with scaling. Interaction of phase change material with electrodes may pose long-term reliability issues and limit the cycling endurance and is a major challenge for DRAM-like applications. Like DRAM PCRAM is a true random access, bit alterable memory.

The scalability of PCRAM device to < 5nm has been demonstrated using carbon nanotubes as electrodes [7], and the reset current followed the extrapolation line from larger devices. In at least one case, cycling endurance of 1E11 was demonstrated [8].

Phase change memory has been used in feature phones to replace NOR Flash since 2011, and has been in volume production at ~ 45nm node since 2012, but no new product has been introduced since. Recently, a 3D Cross Point memory is reported [9]. Details are still lacking but it is speculated that the threshold switching (Ovonic threshold switching, OTS) property of chalcogenide based phase change material constitutes the core of the selector device responsible for the cross point cell, which was first reported in 2009 [10]. This is the first commercial realization of the widely published storage class memory (SCM) [11][12] computer systems badly needed to improve I/O throughput and reduce power and cost, and is promising to change the entire memory hierarchy not only for high-end computation but for mobile systems as well. In addition, since the memory including the selector device is completely fabricated in the BEOL process it is relatively inexpensive to stack multiple layers to reduce bit cost. Due to the importance of this new technology, 3D XP trend will be tracked instead of the original PCRAM.

3D cross point memory (3D XP) consists of a selector element made of OTS (or an equivalent device) in series with a storage element. The selector device has a high ON/OFF ratio and is at OFF state at all times except when briefly turned on during writing or reading. The storage element is programmed to various logic states. Since the selector is always off thus with high resistance the memory array has no leakage issue even if all storage elements are at low resistance state. During write or read operation the selector is temporarily turned on (by applying a voltage higher than its threshold voltage) and the OTS characteristic suddenly reduces its resistance to very low, allowing reading (or programming) current to be dominated by the resistance of the storage element.

The storage element may be a phase change material and in this case the memory cell is a PCRAM switched by OTS. The storage element may also be a resistive memory material. Although bipolar operation makes the circuitry and operation more complicated but the array structure is very similar to that using PCRAM.

PCRAM has the advantage of being unipolar in operation, more product proven, and high cycling endurance. ReRAM, on the other hand, promises higher temperature operation and in some cases faster switching. At this time, high-density ReRAM is still in a development stage. Once developed, there seems little barrier prohibiting it from achieving 3D XP structure. Therefore, both trends are currently tracked with ReRAM 3D XP starting about three years (one node) behind PCRAM but after that essentially maintains a parallel path.

Resistive Memory - ReRAM

Beyond FeRAM, MRAM and PCRAM a large category of two-terminal resistive devices are being studied for memory applications. Many of these resistive memories are still in research stage and are discussed in more detail in the emerging device (beyond CMOS) chapter. Because of their promise to scale below 10nm and the focused R&D efforts in many industrial labs make this technology widely considered a potential successor to NAND (including 3D NAND).

Being only a resistor and requires bipolar operation high-density ReRAM development has been limited by the lack of a good selector device, since simple diodes cannot be used. Recent advances in 3D XP memory, however, seem to have solved this bottleneck and ReRAM could make rapid progress if other technical issues such as erratic bits are solved.

ReRAM trends are shown in several tabulation forms. In addition to 3D XP array similar to PCRAM-based 3D XP memory high-density ReRAM products may be fabricated using a 2D array and small WL/BL 1/2 pitch. Furthermore, if eventually OTS type of selector device is adopted it seems feasible to fabricate BiCS type 3D ReRAM using a transistor in the bottom and OTS selector for each ReRAM device in the 3D array, as depicted in Figure MM8 [13]. Since up to now no high-density ReRAM product has been introduced these trends seem speculative. Yet since the bottleneck of bipolar selector device seems solved by the introduction of 3D XP memory, progress in ReRAM should be expected thus these speculative trends are included in the potential solutions.

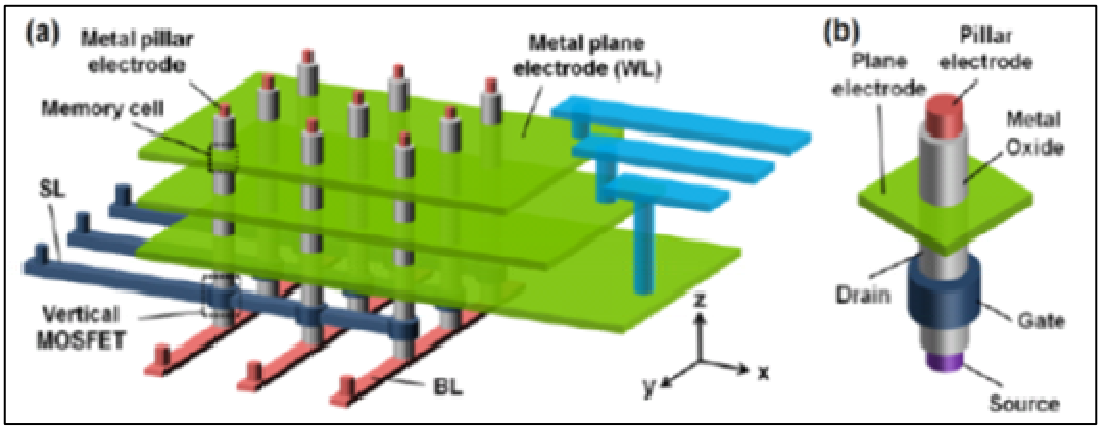


Figure MM8: Schematic view of (a) 3D cross-point architecture using a vertical RRAM cell and (b) a vertical MOSFET transistor as the bit-line selector to enable the random access capability of individual cells in the array [13].

5.3 REFERENCES

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

[2] R. Katsumata, M. Kito, Y. Fukuzumi, M. Kido, H. Tanaka, Y. Komori, M. Ishiduki, J. Matsunami, T. Fujiwara, Y. Nagata, L. Zhang, Y. Iwata, R. Kirisawa, H. Aochi and A. Nitayama, "Pipe-shaped BiCS Flash Memory with 16 Stacked Layers and Multi-Level-Cell Operation for Ultra High Density Storage Devices," *Digest of Technical Papers, 2009 Symposium on VLSI Technology,* pp. 136-137, 2009.

[3] J. Jang, H.S. Kim, W. Cho, H. Cho, J. Kim, S.I. Shim, Y. Jang, J.H. Jeong, B.K. Son, D.W. Kim, K. Kim, J.J. Shim, J.S. Lim, K.H. Kim, S.Y. Yi, J.Y. Lim, D. Chung, H.C. Moon, S. Hwang, J.W. Lee, Y.H. Son, U.I. Chung, and W.S. Lee, "Vertical Cell Array using TCAT (Terabit Cell Array Transistor) Technology for Ultra High Density NAND Flash Memory," *Digest of Technical Papers, 2009 Symposium on VLSI Technology,* pp. 192-193, 2009.

[4] J. Kim, A.J. Hong, S. M. Kim, E.B. Song, J.H. Park, J. Han, S. Choi, D. Jang, J.T. Moon, and K.L. Wang, "Novel Vertical-Stacked-Array-Transistor (VSAT) for Ultra-high-density and Cost-effective NAND Flash Memory Devices and SSD (Solid State Drive)," *Digest of Technical Papers, 2009 Symposium on VLSI Technology,* pp. 186-187, 2009.

[5] W. Kim, S. Choi, J. Sung, T. Lee, C. Park, H. Ko, J. Jung, I. Yoo, and Y. Park, "Multi-layered Vertical Gate NAND Flash Overcoming Stacking Limit for Terabit Density Storage," *Digest of Technical Papers, 2009 Symposium on VLSI Technology,* pp. 188-189, 2009.

[6] S.J. Whang, K.H. Lee, D.C. Shin, B.Y. Kim, M.S. Kim, J.H. Bin, J.H. Han, S.J. Kim, B.M. Lee, Y.K. Jung, S.Y. Cho, C.H. Shin, H.S. Yoo, S.M. Choi, K. Hong, S. Aritome, S.K. Park, and S.J. Hong, "Novel 3-dimentional Dual Control-Gate with Surrounding Floating-Gate (DC-SF) NAND Flash Cell for 1Tb File Storage Application", *Tech. Digest 2010 International Electron Devices Meeting,* pp. 668-671, 2010.

[7] J. Liang, R.G.D. Jeyasingh, H-Y. Chen and H-S. P. Wong, "A 1.4uA Reset Current Phase Change Memory Cell with Integrated Carbon Nanotube Electrodes for Cross-Point Memory Application", *Digest of Technical papers, 2011 Symposium on VLSI Technology,* 5B-4, 2011.

[8] I.S. Kim, S.L. Cho, D.H. Im, E.H. Cho, D.H. Kim, G.H. Oh, D.H. Ahn, S.O. Park, S.W. Nam, J.T. Moon, and C.H. Chung, "High Performance PRAM Cell Scalable to Sub-20nm Technology with below 4F2 Cell Size, Extendable to DRAM Applications", *Digest of Technical papers, 2010 Symposium on VLSI Technology,* 19-3, 2010.

[9] H. Castro, "Accessing memory cells in parallel in a cross-point array", Publication 2015/0074326 A1 US Patent Office, March 12, 2015.

[10] DC Kau, S. Tang, I.V. Karpov, R. Dodge, B. Klehn, J.A. Kalb, J. Strand, A. Diaz, N. Leung, J. Wu, S. Lee, T. Langtry, K-W Chang, C. Papagianni, JW Lee, J. Hirst, S. Erra, E. Flores, N. Righos, H. Castro and G. Spadini, "A Stackable Cross Point Phase Change Memory", *Tech. Digest 2009 International Electron Devices Meeting,* pp. 617-620, 2009.

[11] R. Freitas and W. Wilcke, "Storage Class Memory, the Next Storage System Technology", **52**(4/5), 439, IBM Journal of Research and Development, 2008.

[12] G.W. Burr, B. Kurdi, J.C. Scott, C.H. Lam, K. Gopalakrishnan, and R. Shenoy, "An overview of candidate device technologies for storage class memory", **52**(4/5), 449, IBM Journal of Research and Development , 2008.

[13] H.Y. Chen, S.M. Yu, B. Gao, P. Huang, J.F. Kang, H.-S.P. Wong, "HfOx Based Vertical Resistive Random Access Memory for Cost-Effective 3D Cross-Point Architecture without Cell Selector", *Tech. Digest 2012 International Electron Devices Meeting,* pp. 497-500, (20.7.1-20.7.4), 2012.

6 INTERCONNECT TECHNOLOGIES

Table MM2 highlights and differentiates the top key challenges. The most difficult challenge for interconnects is the introduction of new materials that meet the wire conductivity requirements and reduce dielectric permittivity. As for the conductivity, the impact of size effects on interconnect structures must be mitigated. Future effective κ requirements preclude the use of a trench etch stop for dual damascene structures. Dimensional control is a key challenge for present and future interconnect technology generations and the resulting difficult challenge for etch is to form precise trench and via structures in low-κ dielectric material to reduce variability in RC. The dominant architecture, damascene, requires tight control of pattern, etch and planarization. To extract maximum performance, interconnect structures cannot tolerate variability in profiles without producing undesirable RC degradation. These dimensional control requirements place new demands on high throughput imaging metrology for measurement of high aspect ratio structures. New metrology techniques are also needed for in-line monitoring of adhesion and defects. Larger wafers and the need to limit test wafers will drive the adoption of more *in situ* process control techniques.

For more details please refer to the Interconnect Technology Workgroup (TWG) Chapter.

Table MM2: Interconnect difficult challenges.

Critical Challenges	Summary of Issues
Materials Mitigate impact of size effects in interconnect structures	Line and via sidewall roughness, intersection of porous low-κ voids with sidewall, barrier roughness, and copper surface roughness will all adversely affect electron scattering in copper lines and cause increases in resistivity.
Metrology Three-dimensional control of interconnect features (with its associated metrology) will be required	Line edge roughness, trench depth and profile, via shape, etch bias, thinning due to cleaning, CMP effects. The multiplicity of levels, combined with new materials, reduced feature size and pattern dependent processes, use of alternative memories, optical and RF interconnect, continues to challenge.
Process Patterning, cleaning, and filling at nano dimensions	As features shrink, etching, cleaning, and filling high aspect ratio structures will be challenging, especially for low-κ dual damascene metal structures and DRAM at nano-dimensions.
Complexity in Integration Integration of new processes and structures, including interconnects for emerging devices	Combinations of materials and processes used to fabricate new structures create integration complexity. The increased number of levels exacerbate thermomechanical effects. Novel/active devices may be incorporated into the interconnect.
Practical Approach for 3D Identify solutions which address 3D interconnect structures and other packaging issues	Three-dimensional chip stacking circumvents the deficiencies of traditional interconnect scaling by providing enhanced functional diversity. Engineering manufacturable solutions that meet cost targets for this technology is a key interconnect challenge.

ROADMAP AND POTENTIAL SOLUTIONS

Table MM3: Interconnect roadmap for More Moore scaling.

YEAR OF PRODUCTION	2015	2017	2019	2021	2024	2027	2030
Logic device technology naming	P70M56	P48M36	P42M24	P32M20	P24M12G1	P24M12G2	P24M12G3
Logic industry "Node Range" Labeling (nm)	"16/14"	"11/10"	"8/7"	"6/5"	"4/3"	"3/2.5"	"2/1.5"
Logic device structure options	finFET FDSOI	finFET FDSOI	finFET LGAA	finFET LGAA VGAA	VGAA, M3D	VGAA, M3D	VGAA, M3D
INTERCONNECT TECHNOLOGY							
Conductor	Cu	Cu	Cu	Cu Silicides Carbon Collective Excitations	Cu Silicides Carbon Collective Excitations	Cu Silicides Carbon Collective Excitations	Cu Silicides Carbon Collective Excitations
Number of wiring layers	13	14	15	16	18	22	30
Barrier metal - intermediate wire (tight pitch)	Ta(N)	Ta(N), Mn(N)	Ta(N), Mn(N)	Ta(N),Mn(N), SAM	Ta(N), Mn(N),SAM	Ta(N), Mn(N),SAM	Ta(N), Mn(N),SAM
Barrier thickness - intermediate wire							
Inter-metal dielectrics (IMD) and k value - intermediate wire	SiCOH (2.55)	SiCOH (2.40-2.55) Airgap (1.0)	SiCOH (2.20-2.55) Airgap (1.0)	SiCOH (2.20-2.55) Airgap (1.0) MOF, COF	SiCOH (2.00-2.55) Airgap (1.0) MOF, COF	SiCOH (2.00-2.55) Airgap (1.0) MOF, COF	SiCOH (2.00-2.55) Airgap (1.0) MOF, COF
Di-electric Young Modulus							
Mx - tight-pitch interconnect resistance [Ohms/um]	40	130	500	900	3000	3000	3000
Mx - tight-pitch interconnect capacitance [aF/um]	200	190	180	180	180	180	180
Vx - tight-pitch interconnect via resistance [Ohms/via]	10	15	20	35	35	35	35
ARx -tight-pitch interconnect aspect ratio	2.0	2.0	2.0	2.0	2.0	2.0	2.0
TDDB Emax - tight-pitch interconnect (MV/cm)							
Jmax - tight-pitch interconnect (MA/cm2 at 105°C)							
MP80 - 80nm pitch interconnect resistance [Ohms/um]	13	13	13	13	13	13	13
MP80 - 80nm pitch interconnect capacitance [aF/um]	210	210	210	210	210	210	210
VP80 - 80nm pitch interconnect via resistance [Ohms/via]	10	10	10	10	10	10	10
ARP80 - 80nm pitch interconnect aspect ratio	2.0	2.0	2.0	2.0	2.0	2.0	2.0

Conductor: Cu will be the preferred solution for the M1 and Mx levels. Although a resistivity increase due to electron scattering is already apparent, a hierarchical wiring approach such as scaling of line length along with the fact that of the width still can overcome the problem. As the alternative materials, two directions are proposed. One is the usage of the metals with less size effect e.g. silicides and the other is the introduction of materials that have different conductance mechanism e.g. carbon and collective excitations. The latter materials are still in R&D phase to implement to the semiconductor.

Barrier Metal: Cu wiring barrier materials must prevent Cu diffusion into the adjacent dielectric but also must form a suitable, high quality interface with Cu to limit vacancy diffusion and achieve acceptable electromigration lifetimes. Ta(N) is a well-known industry solution. Although the scaling of Ta(N) deposited by PVD is limited, other nitrides such as Mn(N) which can be deposited by CVD or ALD have recently attracted attention. As for the emerging materials, SAM (Self-Assembled Monolayers) are researched as the candidates for future generation.

IMD (Inter-metal Dielectrics): Reduction of the ILD κ value is slowing down because of problems with manufacturability. The poor mechanical strength and adhesion properties of lower-κ materials are obstructing their incorporation. Delamination and damage during CMP are major problems at early stages of development, but for mass production, the hardness and adhesion properties needed to sustain the stress imposed during assembly and packaging must also be achieved. The difficulties associated with the integration of highly porous ultra-low-κ ($\kappa \leq 2$) materials are becoming clearer, and air-gap technologies is the alternative path to lower the inter-layer capacitance. As the emerging materials, MOF (Metal Organic Framework) and COF (Carbon Organic Framework) are advocated.

Reliability-EM (Electromigration): An effective scaling model has been established assuming that the void is located at the cathode end of the interconnect wire containing a single via with a drift velocity dominated by interfacial diffusion as shown in Figure MM9. The model predicts that τ scales with w*h/j, where w is the linewidth (or the via diameter), h the interconnect thickness, and j the current density. Whereas the geometrical model predicts that the lifetime decreases by half for each new generation, it can also be affected by small process variations of the interconnect dimensions. J_{max} (The maximum equivalent dc current density) and J_{EM} (The maximum current density) are limited by the interconnect geometry scaling. J_{max} increases with scaling due to reduction in the interconnect cross-section and increase in the maximum operating frequency. The practical solutions to overcome the lifetime decrease in the narrow linewidths are discussed actively over the past years. Recent studies show an increasingly important role of grain structure in contributing to the drift velocity and thus the EM reliability beyond the 45nm node. Process options with

Cu alloys seed layer (e.g., Al or Mn) have shown to be an optimum approach to increase the lifetime. Other approaches are the insertion of a thin metal layer (e.g CoWP or CVD-Co) between the Cu trench and the dielectric SiCN barrier and the usage of the short length effect. The short length effect has effectively been used to extend the current carrying capability of conductor lines and has dominated the current density design rule for interconnects.

Reliability - TDDB (Time Dependent Dielectric Breakdown): Basically, the dielectric reliability can be categorized according to the failure paths and mechanisms as shown in Figure MM9-Figure MM11. While a large number of factors and mechanisms have already been identified, the physical understanding is far from complete.

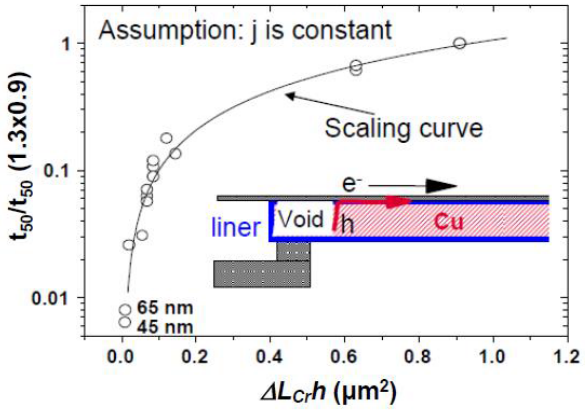


Figure MM9: Experiment and model of lifetime scaling versus interconnect geometry.

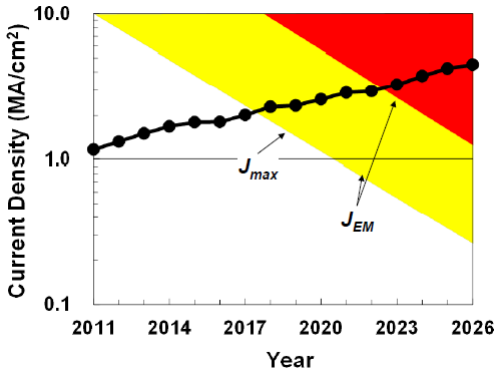


Figure MM10: Evolution of Jmax (from device performance) and J_{EM} (from targeted lifetime).

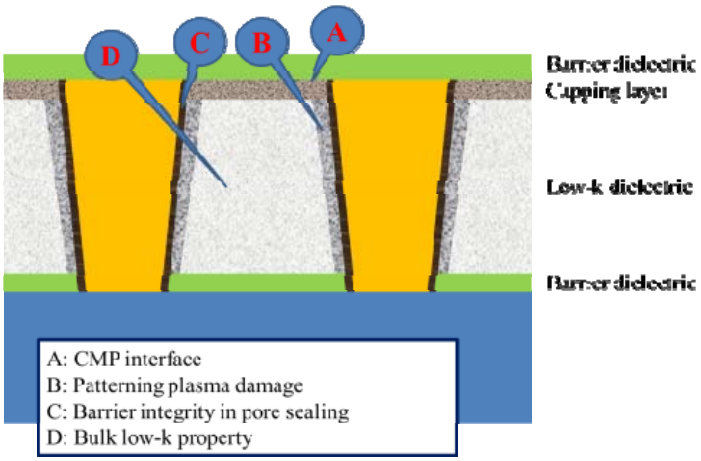


Figure MM11: Degradation paths in low-k damascene structure.

7 PROCESS INTEGRATION

7.1 FRONT-END PROCESSES

The purpose of Front-End Processes TWG is to define comprehensive future requirements and potential solutions for the key front end wafer fabrication process technologies and the materials. Scope of this TWG is shown in Figure MM14. For more details please refer to the Front-End Processes Technology Workgroup (TWG) Chapter.

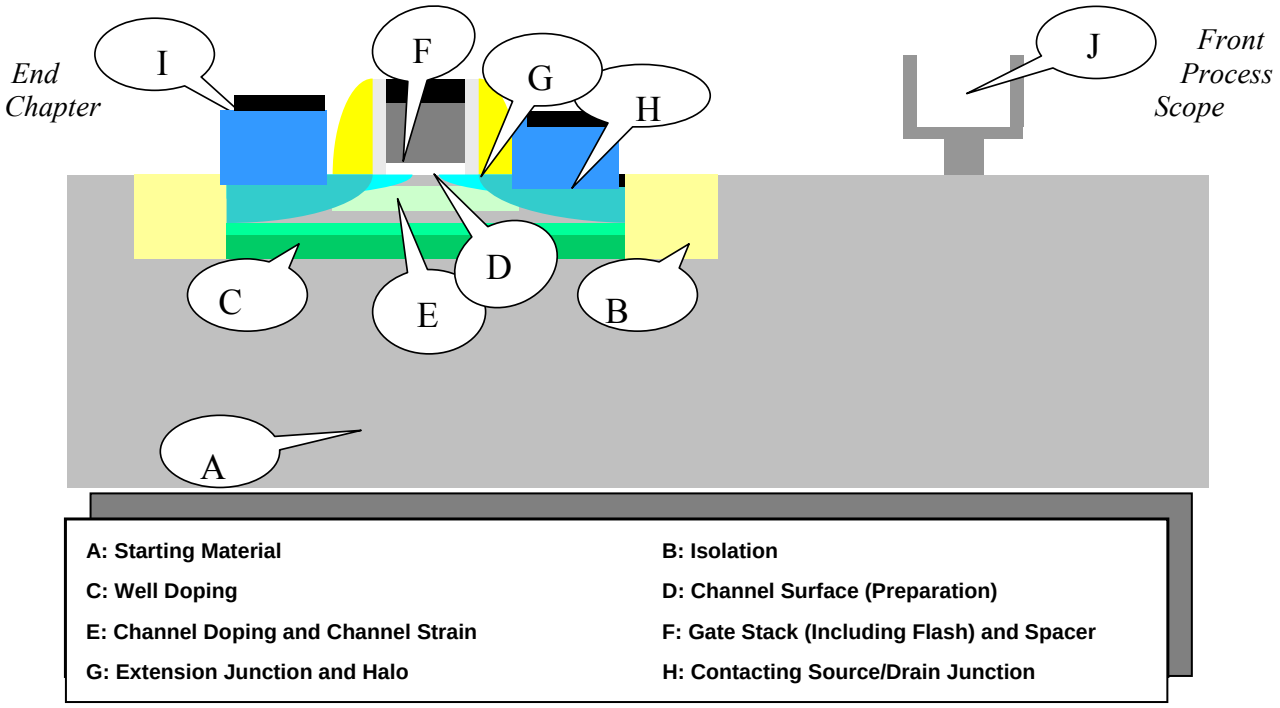


Figure MM12: Front-end processes scope.

DIFFICULT CHALLENGES

A transition to new device architectures such as GAA is needed to continue the scaling, particularly beyond 2020. Table MM4 lists the associated challenges of manufacturing GAA devices.

Table MM4: GAA manufacturing difficult challenges.

Difficult challenges	Opportunities and issues
Transition from fin to GAA	• Extension of fin processes • Need to deal with higher aspect ratio starting topography
Strain engineering for GAA devices	• Continued use of embedded epitaxy for channel mobility boost • GAA mobility enhancements • Integration of dual channel materials • Usage of sSOI substrates and stress conversion through condensation techniques • In vertical GAA architectures new techniques are needed to induce channel stress (e.g. reintroduction of stressed liners)
Junction engineering	• reducing junction concentration and achieving dopant redistribution • Conformal doping solutions are needed, increased importance for GAA or NW architectures
High mobility	• different materials needed for NFET and PFET which leads to significant challenges to co integrate the materials

Difficult challenges	Opportunities and issues
material integration	• process solutions need to be compatible with material requirements such as low temperatures needed for post processing steps and new requirements on limiting the material losses in the subsequent processing steps • the materials used need to have low defectivity requirements (no killer defects in the channel)
Starting substrates	• GAA architectures can potentially be easier integrated and with less parasitics on SOI, sSOI, thin SOI substrates • Substrates for high mobility solutions – cost and defectivity are issues that will need to be addressed
Etch	• high aspect ratio – deep trenches, high pillars • high selectivity requirements compatible with the aggressive ground rules • GAA architectures and the need to eliminate parasitics drive the requirements for directional etching • Improved LER, etch bias and loading • Gate recess control and uniformity – driven by the aggressive ground rules and yield requirements
Material deposition	• Very high conformality processes are needed to wrap-around gate materials around the wires • High aspect ratio fill capability is needed (taller and thinner structures) • Gate fill solutions for sub 20 nm gates with acceptable gate resistance • Contact metal deposition solutions for increased aspect ratio contacts (vertical devices) • TDDB requirements (smaller distances needed between gate and contacts drive the requirements for the insulator materials used)
Cleans	• Good clean or removal without any residual defects or material removal
CMP	• Control and uniformity for gate and contact • Higher starting topography • Both non selective and highly selective slurries

7.2 LITHOGRAPHY

The Lithography Technology Working Group's mission is to identify lithographic options that could enable future semiconductor nodes and better semiconductor products and to describe the driving forces for their implementation and the challenges to their implementation.

For more details please refer to the Lithography Technology Workgroup (TWG) Chapter.

The overall 2015 roadmap shows a progression of new types of logic devices being introduced and many new types of memory devices as possible options in the future. This change in device structure drives what sort of patterning challenges will come in the future. In the past, flash memory and the fins in finFET devices drove the industry to implement self-aligned quadruple patterning (SAQP). But in the future, planar flash memory will stop shrinking due to electrical reasons and fin structures will be replaced by gate all around structures and SAQP will continue to be usable to pattern these levels. Now it is the critical levels for DRAM and the metal levels for logic that will drive improved line and space patterning instead. And hole type patterns will continue to be a challenge. Tighter pitch metal levels drive tighter pitch vias, and new types of devices drive tighter pitch contacts. In addition, the projected introduction of vertical gate all around structures will drive patterning to produce even smaller hole type structures.

Lithographic patterning methods have critical dimension "cliffs" where a given patterning method can no longer be used to produce features with a half pitch below that value. For example, the theoretical limit for ArF immersion lithography is about 36nm half pitch, but if one actually imaged this pitch with ArF, the process window would be unacceptably low and it would be impossible to pattern any sort of two dimensional feature, such as a line end or jog. So the practical limit of ArF immersion lithography for lines and spaces is about 40nm half pitch. The edge of a "cliff" is not one specific CD but is actually a small range of CDs, where the range depends on how much patterning process window is acceptable and how complex the desired pattern is. An example of this is seen in a published 22nm node logic chip design [1]. The smallest pitch metal layers in this design have a half pitch of 40nm, representing the line and space cliff for single exposure ArF immersion lithography; but the metal one level has a half pitch of 42nm, reflecting a more complicated design for this level to match the transistor structure just below it. Chip makers will tend

to design layers on the larger side of cliffs where possible and to introduce layers at sizes below the current cliffs only when a new node is introduced and significant product improvement is gained. Printing smaller features than a cliff requires new tooling or a different process and almost always increases cost and/or process complexity.

Since patterning cliffs have importance in selecting patterning methods and designing chips, our possible options tables are structured around these cliffs. Roughly 40nm is the cliff for ArF immersion single exposure line and space patterning, so 20nm is the cliff for pattern doubling and 10nm is the cliff for pattern quadrupling using ArF immersion patterning. Since pattern doubling is in widespread manufacturing use and pattern quadrupling is already in use for layers that can be quadrupled through self-aligned quadruple patterning (SAQP), this is our baseline process to compare possible new line and space patterning methods to.

Multiple patterning is also the baseline process for critical level hole type patterning. Hole type patterning is needed for contact, via and cut type levels. Double patterning of dense contact holes shrinks the pitch by approximately 30% and quadruple patterning shrinks the pitch by approximately 50%. This means that quadruple patterning of dense contact holes matches the shrink achieved with just double patterning of lines and spaces. So even though the designed pitches of the smallest hole type patterns in a chip design is larger than that of the smallest lines and spaces, already multiple patterning is in use for these levels, and the introduction of smaller pitch devices provides greater challenges for hole type patterns than for line and space type patterns. These types of patterns can be self-aligned, in the sense that the actual sides of a patterned contact hole are created by the features on either side of it rather than the dimension of the lithographically printed hole. In this case, the hole CD per se is not as critical as the edge placement error, or EPE, which is a combination of the effects of CD and overlay variations on the position of the outer edges of features.

7.2.1 POSSIBLE OPTIONS

Possible patterning options for different pitch ranges and their time frames for implementation are shown in Figure MM13 and Figure MM14, metal levels and for hole type patterns, respectively. The ranges in the left column reflect the litho "cliffs" described above. For Figure MM13, the range down to 14nm half pitch is shown twice. The first instance shows that manufacturing processes for pitches in this range are already committed for 2017 using quadruple patterning (QP). The second instance relates to pitches in this range projected for production starting in 2019, where other patterning options are still possibilities.

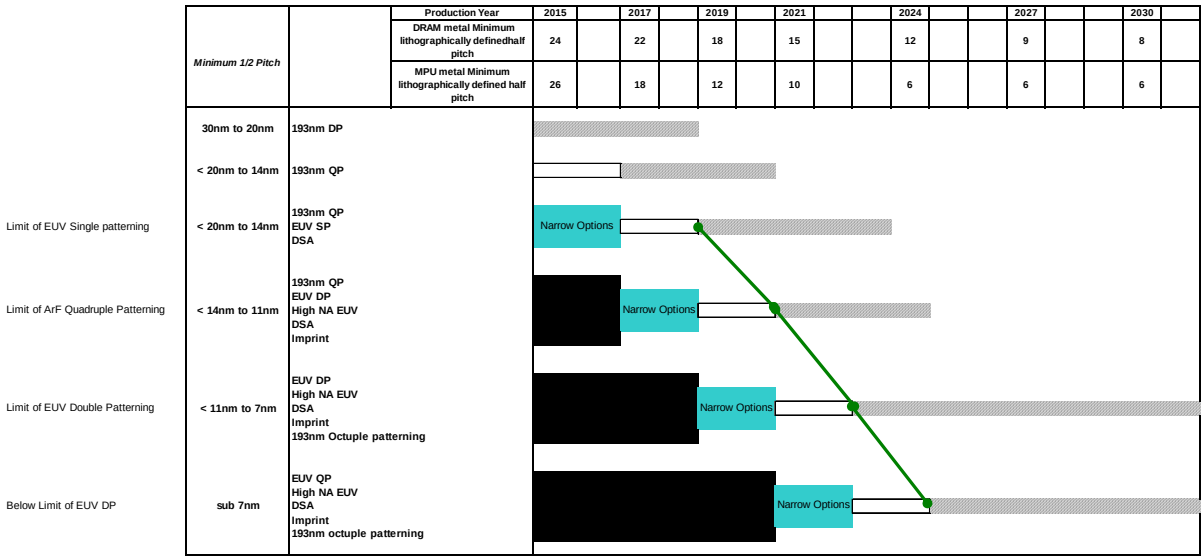


Figure MM13: DRAM critical and MPU metal level potential solutions.

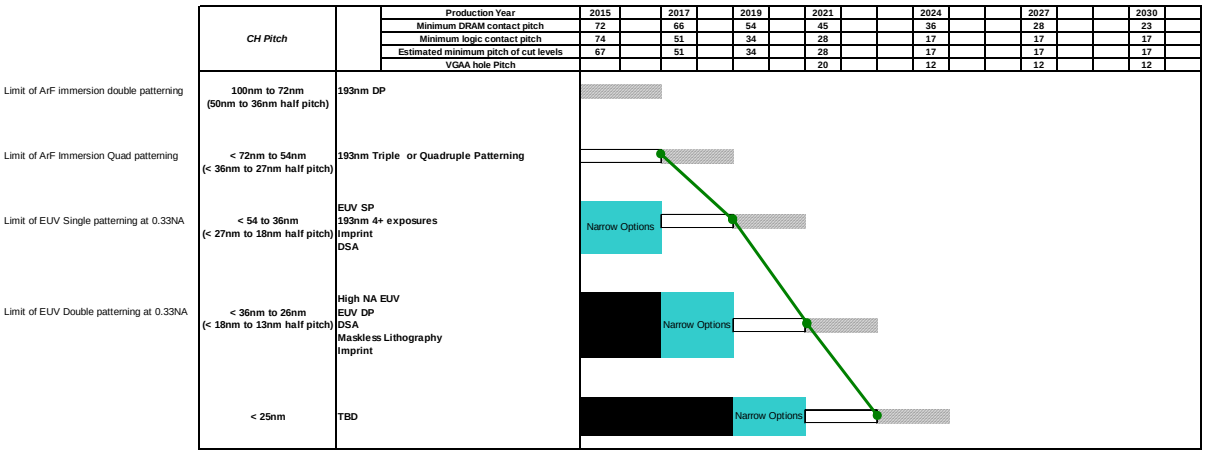


Figure MM14: Hole type pattern potential solutions.

For lines and spaces, quadruple patterning can be extended at least through 2018 and possible through 2022, depending on the complexity and tolerance control required of the projected 10nm half pitch metal levels projected for manufacturing in 2019 and 2021. The alternative technologies to multiple patterning shown in Figure MM11will have to show some substantial cost or design benefit over multiple pattering in order to be adopted before 2024.

For hole type patterns, quadruple patterning is also already in use. However, more than quadruple patterning is projected to be needed for early manufacturing production in 2019. The opportunity for implementing alternatives to multiple patterning is thus much sooner and more beneficial for the industry for hole type patterns than it is for line and space patterns.

Each possible alternative patterning option with current R&D for future devices is discussed in detail in the sections below.

7.2.2 MULTIPLE PATTERNING

Metal levels with a half pitch under 20nm using SAQP should start initial production lots some time in 2016 and will be accompanied by contact levels that require triple or quadruple patterning. Because quadruple patterning is already in use, our concern in this roadmap is whether multiple patterning will be extended beyond quadruple patterning, rather than can it work in manufacturing. Key challenges for extending multiple lithography are cost, process complexity, and design issues, especially for random logic. The challenges relate more to managing the design and managing all the tolerance variations the many patterning steps entail, rather than in doing extra processing steps per se.

Specific challenges are somewhat dependent on the sort of multiple patterning being extended. One possible strategy for patterning metal levels is complementary lithography and SAQP, complementary lithography and self-aligned octuple patterning. This strategy puts pressure on the patterning of cut levels that segment the lines into usable circuit elements. ArF lithography tools can be used for the cut levels, but many cut masks and printing steps are needed, leading to high cost and process complexity. The other strategy for patterning metal levels is to use a litho/etch, litho/etch type of multiple patterning process. This enables more complicated designs, which are needed for some metal levels. In this case it is the additional lithography steps for the lines themselves rather than the cuts that impose more cost and complexity to the process.

More than quadruple patterning will be needed in manufacturing for line and spaces patterns sometime between 2021 and 2024, depending on whether 10nm critical CDs can be quadruple patterned with adequate tolerance control or not. But more than quadruple patterning will be needed for hole type patterns as early at 2019. This suggests that hole type patterns are likely to be the first driver for the introduction of alternatives to multiple patterning.

Possible alternative technologies to multiple patterning with active development efforts underway are discussed individually below.

7.2.2 EUV LITHOGRAPHY

EUV is an exposure technique that uses light with a wavelength of 13.5nm and all reflective optics. The biggest issue with EUV has been source power. As of the 2013 ITRS roadmap, there was insufficient demonstrated source power to do much other than characterize the tools and some of the imaging. EUV exposure tools have made significant progress since that roadmap. Now customers are reported to have achieved throughputs of several hundred wafers a day over a period of several weeks. This is enough throughput to do chip process development work, but sufficient uptime is also needed. The current average availability of 55% to 60% reported by ASML, the tool manufacturer, is projected by them to improve to something over 70% in the beginning of 2016. Reaching this availability target and then improving throughput is the first key challenge for EUV. If the current power and availability roadmaps are met, production throughput capable tools are projected to be available in time for use in manufacturing the foundry "7nm" logic node production starting in 2018.

Other key challenges for EUV are defects and resist resolution, sensitivity and throughput. Defects relate both to mask defects and to particles generated during tool operation. EUV Pellicles are strongly desired to improve overall process defectivity. There is progress being reported on pellicles. Targeting hole type layers such as contacts and cut levels that have less bright area on the mask will make mask defects easier to manage. Per the possible options charts discussed above, this is the type of layer that requires patterning innovation the soonest; and thus is a likely first target for EUV use. Overall defectivity has to be demonstrated at close to manufacturing levels in 2016 for EUV to be committed for production in 2018.

EUV resist line edge roughness (LER), line width roughness (LWR) and contact hole critical dimension uniformity are also challenges. Reported resolution on a commercial scanner with commercial chemically amplified resist has reached 14nm half pitch with reasonable process windows, but the best reported results for LWR do not meet the numbers in the requirements table for litho, and the photospeed is worse than needed for scanners to reach their specified throughput.

Since lines and spaces produced with SADP and SAQP have very good roughness, EUV is unlikely to be used for levels that are suitable for these processes. Line and space patterns with a design that requires litho/etch-litho/etch type multiple patterning are more promising targets for EUV implementation. The most promising target for implementation is hole type levels that already require multiple patterning using litho/etch type processes.

Progress has also been made on extendibility. ASML has stated that they will produce a 0.55NA EUV scanner that uses different magnification in the x and y directions and given an overall description of what such a tool would be like. We estimate the earliest they could have such a tool available would be 2020, suggesting it could be used in manufacturing in 2021. There has also been progress in novel metal based resists, which have the potential for imaging in thin films with adequate absorbance of EUV light.

7.2.3 DIRECTED SELF ASSEMBLY (DSA)

DSA is a chemical based process for creating patterns by using special polymers that separate into domains of different composition upon annealing. Lithographically printed guide features are used to direct the phase separation into producing useful patterns. Additional levels, such as "cut" levels are often used to trim the phase separated features after they form. DSA can be used for pitch multiplication, both of lines and spaces and of hole type patterns. DSA can also be used to "rectify" hole patterns. That is, larger holes can be made smaller and more uniform. Patterns with larger CDs require larger polymers and larger polymers are harder to anneal, so features that are relatively large are not suitable for DSA. Critical CDs for chips are just now reaching the range appropriate for DSA use. Guide structures appropriate for such CDs can be printed with ArF immersion scanners but could also be printed with other patterning methods.

DSA progress has clearly moved into a development phase at many fabs. DSA materials are now available in gallon quantities from several vendors and there are enough publications to show that both memory and logic producers are working on DSA. The most likely first application of DSA is contact holes and/or cut levels, either for rectification or limited pitch multiplication. Limited pitch multiplication is using guide features to create pairs or other small numbers of contacts in close proximity to each other. For line/space pitch multiplication, DSA has to compete with SAQP and so far SAQP is reported to be superior in LER/LWR [2]. While the LER/LWR of DSA produced lines and spaces is not as good at that of SAQP, it is still considerably better that that of EUV. Contact hole CD uniformity is also considerably better than EUV. Assessments of these and other parameters are contained in an industry survey conducted by the ITRS and included in the tables accompanying this document.

The key challenges of DSA are pattern placement, defectivity and pattern inspection, and DSA compatible chip design and layout. Pattern placement is an issue for DSA features that are in between guide structures. Reported defectivity levels have improved in the past two year, but are not yet reported to be at manufacturing levels. Buried defects are a particular problem. With current DSA processes, it is desirable to do the defect inspection before etch. But there can be buried three dimensional defects in phase separation. These sorts of defects are a particular challenge to detect. Since DSA makes only very simple patterns so far, chip designs have to be remade to work with structures available by DSA and layout engines need to incorporate DSA related process windows in designing particular chips.

7.2.4 NANOIMPRINT

Nanoimprint is a technique in which three dimensional molds stamp a pattern into a liquid resist and the resist is cured into a solid pattern while still molded. Then the mold is peeled off and reused. Because it's a contact technique defectivity is a concern and because it has no magnification factor (that is, the templates are 1X), making the templates is a challenge, especially at leading edge semiconductor dimensions. Flash memory, which is more defect tolerant than logic or DRAM, is the first target of nanoimprint. In 2015, new results were published indicating that throughput and overlay have improved significantly. Tools were announced that meet manufacturing throughput needs and are expected to be available for a possible first implementation in 2018. The overlay of these tools is coming close to the requirements for flash memory. Defectivity, template infrastructure and template inspection are still key challenges. These challenges depend significantly on the critical dimension of the pattern being printed. There is enough momentum in nanoimprint to make it possible to make a go/no-go decision for manufacturing sometime in 2016.

7.2.5 MASKLESS LITHOGRAPHY

Maskless lithography refers to the technique of direct writing circuit patterns in resist on a wafer using some sort of programmable writing tool. Maskless lithography tools currently under development use e beam lithography. Direct write is an intrinsically slow technique, and gets slower as scaled to smaller features. So current tool implementation under development uses tools that have multiple ebeams writing at once, either using multiple small ebeam columns or using one column and a specially designed chip that is a source of multiple ebeams, all of which go through the same column. The use of enough multiple beams at once is projected to give adequate throughput for chip writing. This technique is very appealing for low volume chip designs, where the costs of the masks can't be amortized over very many chips. These are typically logic chips. It is also potentially appealing for chip patterns that are very sparse in writing area, for example, hole type patterns such as contacts or cut levels, although the speed of writing such patterns will depend significantly on the tool architecture.

The principle of writing with multiple beams is well established and tool development for writing mask patterns with multiple beams is progressing well. But the mask tools still are expected to take many hours to write a mask. Writing wafers requires smaller features and faster throughput than the mask tools under development can come close to. Actually demonstrating a pilot type tool which integrates the writing of wafers with positional control of the images and some sort of throughput indicative of progress has not happened yet. So the key challenge for this technology is to actually build a working demo tool for writing whole wafers with chip like patterns and overlay control. Work is underway on this task, but the lack of recent published progress suggests that the earliest insertion point for such technology would be 2021 and the earliest insertion target would be the "5nm" logic node. Maskless lithography is shown in Figure MM14 for this node because of the technique's potential for hole type patterns. The lack of relevant published performance targets for the tools under development makes it difficult to list maskless lithography as a possible option anywhere else in Figure MM13 or Figure MM14.

7.2.6 SUMMARY

Table MM5 shows a comparison of the key challenges of each of the possible patterning options discussed above, along with dates and product types for their earliest possible insertion.

Table MM5: Lithography difficult challenges.

Next Generation Technology	First Possible Use in Mfg.	Feature Type	Device Type	Key Challenges	Required Date for Decision making
Multiple Patterning Extension to >4X patterning	2020 to 2023	10nm or less hp metal for logic MPUs 10nm hp for LGAA structures	"5nm" node logic	-Extension to random logic -Printing and overlay of cut levels -Design to cost tradeoff	2018
EUV	2018	22 to 24nm hp CH/Cut Levels 18nm hp LS	"7nm" node logic 18nm DRAM	-Availability & Throughput -Mask Defects -Resist sensitivity and roughness -High NA field size	2016
NanoImprint	2017	14nm hp LS 20nm hp bit lines	2D Flash Memory 3D Flash Memory	-Defectivity -Overlay -Master Template writing and inspection <20nm -Template replication <20nm	2016
DSA (for pitch multiplication)	2018	Contact holes/cut levels	1x DRAM "7nm" node logic	-Pattern Placement -Defectivity and defect inspection -Design -3D Metrology	2016
Maskless Lithography (ML)	2021	Cut levels -- possibly 20nm on 40nm pitch (estimated)	"5nm" node logic (estimated)	-Concept demonstration -Functioning tool	2019

While the implementation dates for multiple patterning extension shown in the table above is later than the dates shown for new pattering technologies currently under development; that is because multiple patterning has already been demonstrated to support critical CDs needed earlier than 2021 to 2024. The alternative technologies can't just demonstrate patterning capability; they have to be better than multiple patterning for at least some levels to be selected for implementation. Each of the techniques contending with multiple patterning has its own strengths and weaknesses. They are shown in spider charts in Figure MM5 below. Successful implementation of any of them requires improvement in some aspect of that technique. These strengths and weaknesses are relative to the current ITRS roadmap targets. These targets are harder than they were two years ago, so technologies can get worse compared to requirements despite making technological progress. Compared to 2013, EUV is closer to manufacturing readiness, but may have issues with resolution if there are further delays in implementation. Nanoimprint has improved its readiness significantly. DSA has improved its readiness slightly and maskless lithography has fallen farther behind requirements. We look forward to seeing progress in all these technologies during the next two years.

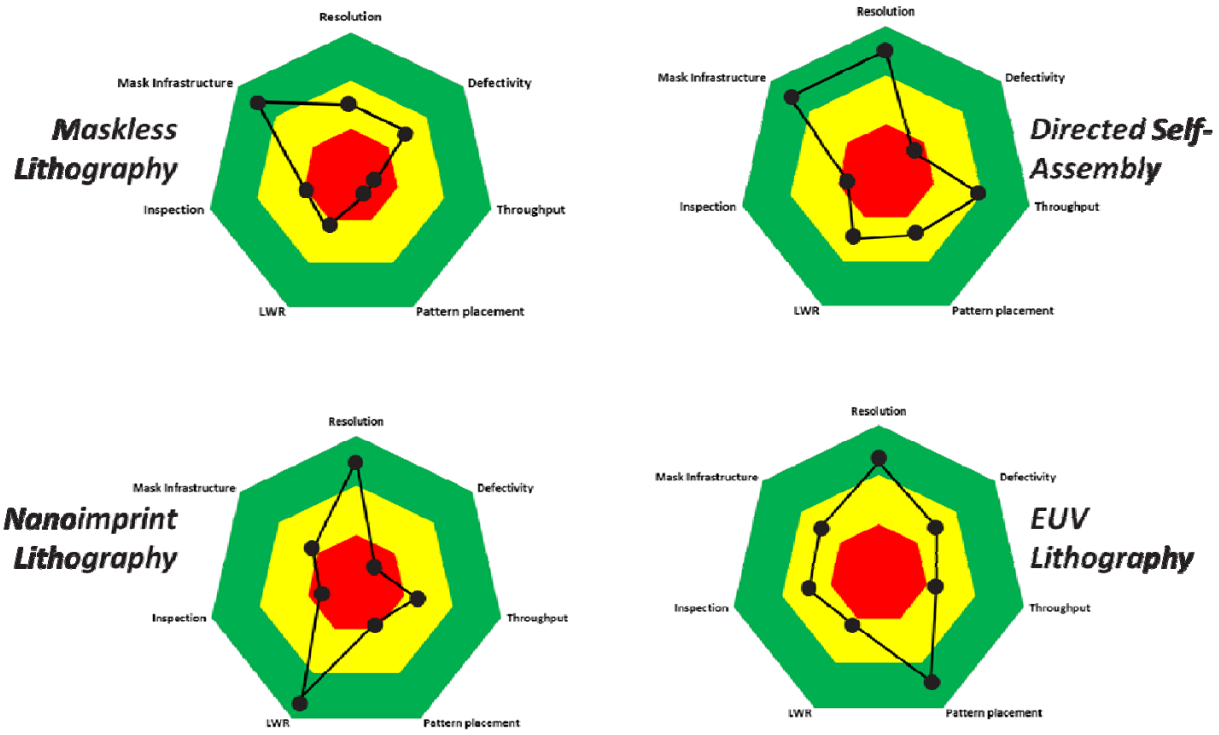


Figure MM15: Radar charts comparing possible alternative to multiple patterning.

7.2.7 REFERENCES

[1] C.-H. Jan et al., "A 22nm SoC platform technology featuring 3-D tri-Gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications,", *IEEE Int. Electron Devices Meeting (IEDM)*, pp. 3.1.1-3.1.4, 2012.

[2] D. Millward et al., "Graphoepitaxial and chemoepitaxial methods for creating line-space patterns at 33nm pitch: comparison to a HVM process", *Proc. SPIE*, Vol. 9423, 2015.

7.3 METROLOGY

The Metrology Technology Working Group's mission is to identify emerging measurement challenges and describe research and development pathways for meeting them, primarily for extending CMOS, accelerating Beyond CMOS technologies, materials characterization and structure function relationships. Metrology also provides the measurement capability necessary for cost-effective manufacturing. As such, the metrology chapter of the ITRS focuses on difficult measurement needs, possible solutions, metrology tool development, and standards.

Broadly speaking, scaling is expected to reach its physical limits within the next few years. As such, feature size reduction will no longer be the main technology driver for the industry. Applications such as Mobile Communication and Information, Smart Automotive, Big Data, Green Energy, and Medical and Health Technologies among others are expected to be the main drivers. These new system-level drivers will greatly reshape semiconductor technology road mapping activities going forward. The new drivers will likely have implications for metrology with respect to the pace and timing by which new technologies are adopted. Over the past ten years, device and integrated circuit technology has rapidly evolved toward the use of complex 3D structures fabricated using new materials and processes with ever decreasing dimensions. The 3D nanoscale nature of these structures provides considerable challenges for all areas of metrology. Several examples of new process technology help describe the new challenges facing metrology. Research into new patterning processes covers the use of directed self-assembly of block copolymers, extreme ultra violet (EUV) lithography, and 3X and 4X multiple patterning. All of these methods result in different challenges for

measurement of critical dimensions (CD), overlay, and defectivity. FinFET transistors are now the dominant microprocessor device architecture, and the challenges associated with 3D nature of all measurements are amplified by shrinking dimensions. Adding to the challenges facing Front End Processes (FEP) metrology are control of the fabrication processes for memory structures which are among the most complex 3D device structures. On-chip and off-chip interconnect materials continue to evolve and interconnect metrology challenges continues to include process control for 3D interconnect. Research into a replacement for copper interconnect is an example of a new material that could change metrology methods for interconnect. The metrology challenges found in Emerging Research Materials include characterization of 2D materials such as graphene and other new materials as well as electrical characterization at the nano-level.

Although new system drivers could affect the pace and timing of technology adoption for the industry as a whole, feature size and complexity will continue to play a major role in the timeline for metrology solutions for new materials, process, and structures. Metrology methods must routinely measure near and at atomic scale dimensions, which require a thorough understanding of nano-scale materials properties and of the physics involved in making the measurement. Metrology development must be done in the context of these issues.

Moreover, it is entirely possible that different materials will be used by different manufacturers at a given technology generation, potentially requiring different metrologies. In the near term, advances in electrical and physical metrology for high- and low-κ dielectric films must continue. The requirement for technology for measurement of devices on ultra-thin and possibly strained silicon on insulator comes from the best available information that is discussed in the Front End Processes Roadmap. The increasing emphasis on active area measurements instead of test structures in scribe (kerf) lines places new demands on metrology. Measurement of relevant properties, such as stress or strain, in a nano-sized, buried area such as the channel of a small dimension gate is a difficult task. Often, one must measure a film or structure property at the surface and use modeling to determine the resultant property of a buried layer.

The relationship between metrology and process technology development needs fundamental restructuring. In the past the challenge has been to develop metrology ahead of target process technology. Today we face major uncertainty from unresolved choices of fundamentally new materials and radically different device designs. Understanding the interaction between metrology data and information and optimum feedback, feed forward, and real-time process control are key to restructuring the relationship between metrology and process technology.

The metrology topics covered in the 2015 Metrology roadmap are microscopy; critical dimension (CD) and overlay; film thickness and profile; materials and contamination analysis; 3D metrology; emerging research materials and devices; and reference materials. These topics are reported in the following sections in this chapter: Microscopy; Lithography Metrology; Front End Processes Metrology; 3D Interconnect Metrology; Traditional Interconnect Metrology; Materials and Contamination Characterization; Metrology for Emerging Research Materials Devices; Reference Materials; 3D Nanometrology needs and challenges; and Reference Measurement Systems. The section on Lithography Metrology includes an introduction to small angle X-ray scattering (SAX). Each topic is covered at length in the main Metrology chapter.

For more details please refer to the Metrology Technology Workgroup (TWG) Chapter.

7.3.1 METROLOGY DIFFICULT CHALLENGES AND POTENTIAL SOLUTIONS

Metrology requirements continue to be driven by advanced lithographic and multi-patterning processes, new materials, and Beyond CMOS materials, structures, and devices. Many short-term metrology challenges listed below will continue beyond the 12 nm ½ pitch. Metrology needs after 2019 will be affected by unknown new materials, device design and processes. Thus, it is difficult to identify all future metrology needs. Shrinking feature sizes, tighter control of device electrical parameters, such as threshold voltage and leakage current, and new interconnect technology such as 3D interconnect will provide the main challenges for physical metrology methods. To achieve desired device scaling, metrology tools must be capable of measurement of properties on atomic distances. Table MM6 presents twelve major challenges for metrology.

Table MM6: Metrology difficult challenges.

Difficult Challenges ≥ 10 nm	*Summary of Issues*
Factory level and companywide metrology integration for real-time *in situ*, integrated, and inline metrology tools.	Standards for process controllers and data management must be agreed upon. Conversion of massive quantities of raw data to information useful for enhancing the yield of a semiconductor manufacturing process. Need for continued development of robust sensors and process controllers; Data management that allows integration of add-on sensors. Reduction of scrap, increased product quality and cycle time.
Starting materials metrology and associated manufacturing metrology	Existing capabilities will not meet Roadmap specifications. Very small particles must be detected and properly sized. Capability for SOI, III-V, GeOI wafers needs enhancement. Challenges come from the extra optical reflection in SOI and the surface quality. CD, film thickness, and defect detection are impacted by thin SOI optical properties and charging by electron and ion beams. Impurity detection (especially particles) at levels of interest for starting materials and reduced edge exclusion for metrology tools.
Control of new process technology such as Directed Self Assembly (DSA) Lithography, multi-patterning, complicated 3D structures such as FinFET & MuGFET transistors, capacitors and contacts for memory, and 3D Interconnect are not ready for their rapid introduction.	Increased adoption of FinFET transistor technology has placed renewed emphasis on the near term need for in-line metrology for dimensional, compositional, and doping measurements. The materials properties of block co-polymers for DSA result in new challenges for lithography metrology. The increased use of multi-patterning techniques introduce the need to independently solve a large set of metrics to fully characterize a multi-patterning process. 3D Interconnect comprises a number of different approaches. New process control needs are not yet established. For example, 3D (critical dimension (CD) and depth) measurements will be required for trench structures including capacitors, devices, and contacts. Traditional metrology instruments do not have the range and resolution required for accurate TSV measurement.
Measurement of complex material stacks and interfacial properties including physical and electrical properties.	Reference materials and standard measurement methodology for new high-κ gate and capacitor dielectrics with engineered thin films and interface layers as well as interconnect barrier and low-κ dielectric layers, and other process needs. Optical measurement of gate and capacitor dielectric averages over too large an area and needs to characterize interfacial layers. Carrier mobility characterization will be needed for stacks with strained silicon and SOI, III-V, GeOI, and other substrates, or for measurement of barrier layers. Metal gate work function characterization is another pressing need.
Measurement test structures and reference materials.	The area available for test structures is being reduced, especially in the scribe lines. Measurements on test structures located in scribe lines may not correlate with in-die performance. Overlay and other test structures are sensitive to process variation, and test structure design must be improved to ensure correlation between measurements in the scribe line and on chip properties. Standards institutions need rapid access to state of the art development and manufacturing capability to fabricate relevant reference materials.
Difficult Challenges < 10 nm	
Nondestructive, production worthy wafer and mask-level metrology for CD measurement for 3D structures, overlay, defect detection, and analysis	Surface charging and contamination interfere with electron beam imaging. CD measurements must account for overall feature profile. Metrology tool imaging resolution must improve to be able to discern 3D information. It is important to have both imaging and scattering techniques available for any given process control situation. Focus, exposure, and etch bias control will require better precision and 3D capability.
New strategy for in-die metrology must reflect across chip and across wafer variation.	Correlation of test structure variations with in-die properties is becoming more difficult as devices shrink. Sampling plan optimization is key to solving these issues.
Statistical limits of sub-12 nm process control	Controlling processes where the natural stochastic variation limits metrology will be difficult. Examples are low-dose implant, thin-gate dielectrics, surface, sidewall and edge roughness of very small structures. Complementary, and hybrid metrology combined with state of the art statistical analyses would be required to reduce the measurement uncertainty.
Structural and elemental analysis at device dimensions and measurements for *beyond CMOS, and emerging materials and devices.*	Materials characterization and metrology methods are needed for control of interfacial layers, dopant positions, defects, and atomic concentrations relative to device dimensions. One example is 3D dopant profiling. Measurements for self-assembling processes are also required.
Determination of manufacturing metrology when device and interconnect technology remain undefined.	The replacement devices for the transistor and structure and materials replacement for copper interconnect are being researched.

Mask Defects	Mask defects, especially for EUV will continue to be a challenge. These include non-visible defects, film thickness non-uniformity, phase separation, and reflectivity.
DSA	Key measurands such as size, location, and alignment need to be better defined. Some of the measurands are material and system dependent. Many of the materials are similar enough that identifying a property with the required contrast may be difficult. A key question seems to be if we can detect low densities of surface and buried defects.

Potential solutions for some of the difficult challenges listed in Table MM6 are highlighted throughout the main Metrology chapter. Table MM7 summarizes potential solutions that cut across different areas, with some examples. The information in Table MM7 cover two broad categories: (1) Combining different measurement technologies in ways that increase the number and/or enhance the usefulness of specific parameters. (2) Increased range and resolution for available measurement technologies. These include not only possible new technologies, but also improving or adapting existing technologies to other uses.

Table MM7: Metrology potential solutions.

Potential Solutions	Examples
Improved Resolution	Better resolution for current technologies such as CD-SEM, CD-AFM, and optical CD among others. Increased range for high resolution instruments and vice –versa (This will greatly increase the ability to measure features such as TSVs) • *Introduction of aberration-corrected low energy SEM column* • *Utilization of high energy SEM* • *Enhanced CD-AFM tip technology* • *Reduced spot size and uniform intensity for optical instruments* • *Use of multi-column electron beam instrument for defect inspection.* • *Increased use of data fusion or image stitching to increase range.*
Improved X-ray metrology for CD and films characterization	Higher brightness sources. • *An X-ray source with >100x brightness of conventional rotating anode sources can enable new techniques such as CD-SAXS and X-ray tomography* • *Improved throughput and increased utilization of already-mainstream X-ray metrology solutions such as HR-XRD, XRF, TXRF, XRR, XPS among others.*
3D Metrology	Non raster capabilities for scanning instruments. • *This would allow extraction of information from non-orthogonal axes.* Multi head/column for scanning instruments: • *Each head could extract different type of information (dimensional, material ...etc),* • *Faster measurements of large areas of the wafer.* Hybrid Metrology: • *Increased use of a combination of instruments to achieve the desired resolution, speed, or low levels of uncertainty needed to characterize different aspects of a feature.*

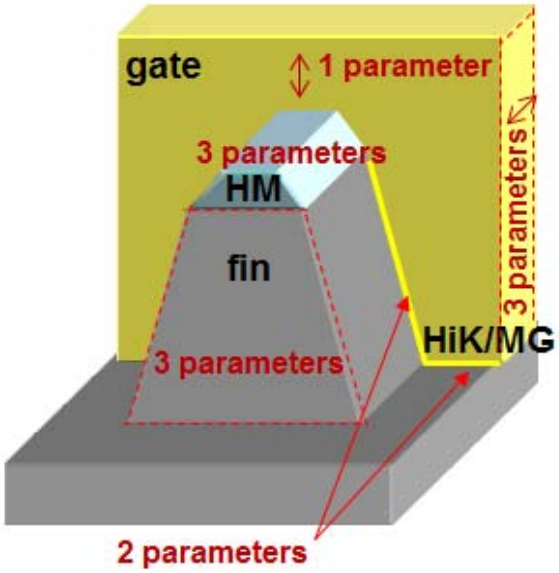


Figure MM16: Complex structures such as finFETs require 3D metrology. To obtain a full 3D profile of the feature above, measurements such as fin CD, height, sidewall angle and roughness, and film thickness are needed. Figure courtesy of Benjamin Bunday, SEMATECH.

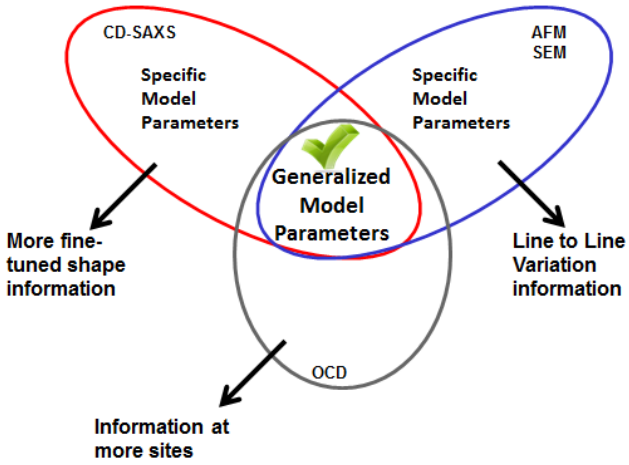


Figure MM17: Conceptual diagram of hybrid metrology. Different instruments provide specific model parameters that are used in a generalized model of the measurement. The arrows indicate specific information from different instruments. Figure courtesy of Richard Silver – NIST.

7.4 YIELD

The sensitivity requirements of the inline defect detection and defect review are described in the tables. For the difficult challenges as well as details on the topics "Wafer Environment Contamination Control" please refer to the Yield Chapter.

7.5 RELIABILITY

Reliability is an important requirement for almost all users of integrated circuits. The challenge of realizing the required levels of reliability is increasing due to (1) scaling, the introduction of (2) new materials and devices, (3)

more demanding mission profiles (higher temperatures, extreme lifetimes, high currents), and (4) increasing constraints of time and money.

1. Scaling produces ICs with more transistors and more interconnections, both on-chip and in the package. This leads to an increasing number of potential failure sites. Failure mechanisms are also impacted by scaling. For example, the time dependent dielectric breakdown (TDDB) of silicon oxy-nitride gate insulators has changed from electric-field-driven to voltage-driven as the insulator thickness has been scaled below 5 nm. In addition, negative bias temperature instability (NBTI) in p-channel devices, which used to be a minor effect when threshold voltages were larger, is now a great concern at the smaller threshold voltages of state-of-the-art devices. When the size of the transistor becomes comparable to or smaller than the values of the fundamental parameters such as mean-free-path of phonons and electrons, and de Broglie wavelength, familiar degradation mechanism may change and new ones may appear. For example, random telegraph noise (RTN) has emerged as a serious reliability issue due to the smallness of the transistor. Another scaling induced new issue is gate to contact breakdown.

Increase in variability is expected as a result of scaling. Reliability mechanisms that are sensitive to device parameters will couple with the variability and be magnified, making reliability projection with limited number of measurements extremely difficult. For example, Vth variation can change the tunneling current and therefore affect all modes that are sensitive to gate current (BTI, TDDB, etc…). Channel length variation can affect lateral field dependent degradation such as HCI. Both channel length and gate overdrive affect the magnitude of RTN, and therefore their variation will also have effects. Large initial variation will make the observation of variability amplified degradation spread difficult without greatly increase the number of device tested.

Scaling may also lead to an effective increase of the stress factors. First, the current density is increasing and this increase impacts interconnect reliability. Second, voltages are often scaled down more slowly than dimensions, leading to increased electric fields that impact insulator reliability. Third, scaling has led to increasing power dissipation that result in higher chip temperatures, larger temperature cycles, and increased thermal gradients, all of which impact multiple failure mechanisms. The temperature effects are further aggravated by the reduced thermal conductivity that accompanies the reduction in the dielectric constant of the dielectrics between metal lines.

2. There are even more profound reliability challenges associated with revolutionary changes associated with new materials and new devices. Recognized failure mechanisms can change. New materials, such as high/low-K dielectrics or metal gates, and new device architectures, such as multi gate or FINFETs, can introduce new failure mechanisms or change the behavior of well-known failure mechanisms such as TDDB or BTI. Reliability evaluation is further complicated by the interaction between the materials in the gate stack being strongly affected by the process details (deposition techniques, thermal budget, etc.). Such complex multi-component gate stack structures may give rise to novel process-specific degradation mechanisms, both intrinsic and extrinsic. For example, with the transition from oxynitride/poly-Si gates to high-k/metal gates, positive bias temperature instability (PBTI) in n-channel devices appears presenting a more serious problem to the device stability. In addition, the nature of TDDB changes from progressive or multiple breakdowns, observed in poly-Si gate MOSFETs, to a more abrupt breakdown. The poor mechanical and thermal properties of low-k intermetallic dielectrics can lead to mechanical failure mechanisms not seen in silicon dioxide intermetallic dielectrics.

One of the routes to continue to the increase functionality of an IC is to integrate sensors and actuators on top of the CMOS platform. Such kind of "more than Moore" approach will greatly increase the complexity of reliability assurance. It is highly likely that such technology will come on line before the end of the road map and we must prepare for it. The likelihood that each sensor/actuator brings along a unique set of reliability problem is high and will present a whole new challenge to the reliability effort.

3. Mission profiles tend to be stretched further. For instance in sensor applications in automotive where temperatures exceeding 200C will be required, and in applications like base stations and solar cells, where (almost) continuous use during tens of years is required

4. Almost needless to say, but the ever increasing constraints of time and money in combination with possible major technology changes poses a real challenge for reliability engineering to keep in sync. Moreover the speed of introduction of these new materials and devices is exceeding our capability to build up learning on new failure mechanisms and physics, whereas the failure rate requirements are become more and more demanding. The impact of an unrecognized failure mechanism that makes it into end products would be significant.

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

7.5.1 RELIABILITY DIFFICULT CHALLENGES

Table MM8 indicates the top near-term reliability challenges. It expands on the More Moore overall Difficult Challenge, "Timely assurance for the reliability of multiple and rapid material, process, and structural changes," described at the beginning of this chapter.

The first near-term reliability challenge concerns failure mechanisms associated with the MOS transistor. The failure could be caused by either breakdown of the gate dielectric or threshold voltage change beyond the acceptable limits. The time to a first breakdown event is decreasing with scaling. This first event is often a "soft" breakdown. However, depending on the circuit it may take more than one soft breakdown to produce an IC failure, or the circuit may function for longer time until the initial "soft" breakdown spot has progressed to a "hard" failure. Threshold voltage related failure is primarily associated with the negative bias temperature instability observed in p channel transistors in the inversion state. It has grown in importance as threshold voltages have been scaled down and as silicon oxy-nitride has replaced silicon dioxide as the gate insulator. Burn-in options to enhance reliability offend-products may be impacted, as it may accelerate NBTI shifts. Introduction of high-κ gate dielectric may impact both the insulator failure modes (e.g., breakdown and instability) as well as the transistor failure modes such as hot carrier effects, positive and negative bias temperature instability. The replacement of polysilicon with metal gates also impacts insulator reliability and raises new thermo-mechanical issues. The simultaneous introduction of high-κ and metal gate makes it even more difficult to determine reliability mechanisms. To put this change into perspective, even after decades of study, there are still issues with silicon dioxide reliability that need to be resolved.

As mentioned above, the move to copper and low κ has impacted front end reliability due to poorer thermal conductivity of low-κ dielectrics, leading to higher on-chip temperatures and higher localized thermal gradients.

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

At the heart of reliability engineering is the fact that there is a distribution of lifetimes for each failure mechanism. With low failure rate requirements we are interested in the early-time range of the failure time distributions. There has been an increase in process variability with scaling (e.g., distribution of dopant atoms, CMP variations, and line-edge roughness). At the same time the size of a critical defect decreases with scaling. These trends will translate into an increased time spread of the failure distributions and, thus, a decreasing time to first failure. We need to develop reliability engineering software tools (e.g., screens, qualification, and reliability-aware design) that can handle the increase in variability of the device physical properties, and to implement rigorous statistical data analysis to quantify the uncertainties in reliability projections. The use of Weibull and log-normal statistics for analysis of breakdown reliability data is well established, however, the shrinking reliability margins require a more careful attention to statistical confidence bounds in order to quantify risk. This is complicated by the fact that new failure physics may lead to significant and important deviations from the traditional statistical distributions, making error analysis non-straightforward. Statistical analysis of other reliability data such as BTI and hot carrier degradation is not currently standardized in practice, but may be needed for accurate modeling of circuit failure rate.

Table MM8: Reliability difficult challenges.

Near-Term 2015-2022	***Summary of issues***
Reliability due to material, process, and structural changes, and novel applications.	TDDB, NBTI, PBTI, HCI, RTN in scaled and non-planar devices. Gate to contact breakdown. Increasing statistical variation of intrinsic failure mechanisms in scaled and non-planar devices. 3D interconnect reliability challenges. Reduced reliability margins drive need for improved understanding of reliability at circuit level. Reliability of embedded electronics in extreme or critical environments (medical, automotive, grid...).
Long-Term 2023-2030	***Summary of issues***
Reliability of novel devices, structures, and materials.	Understand and control the failure mechanisms associated with new materials and structures for both transistor and interconnect. Shift to system level reliability perspective with unreliable devices. Muon induced soft error rate.

The single long-term Reliability Difficult Challenge concerns novel, disruptive changes in devices, structures, materials, and applications. For example, at some point there will be a need to implement non-Copper interconnect (e.g., optical or, Carbon nanotube based interconnects), or tunnel-based FET's instead of classical MOSFET's. For such disruptive solutions there is at this moment little, if any, reliability knowledge (as least as far as their application in ICs is concerned). This will require significant efforts to investigate, model (both a statistical model of lifetime distributions and a physical model of how lifetime depends on stress, geometries, and materials), and apply the acquired knowledge (new building-in reliability, designing-in reliability, screens, and tests). It also seems likely that there will be less-than-historic amounts of time and money to develop these new reliability capabilities. Disruptive material or devices therefore lead to disruption in reliability capabilities and it will take considerable resources to develop those capabilities.

7.5.2 RELIABILITY REQUIREMENTS

Reliability requirements are highly application dependent. For most customers, current overall chip reliability levels (including packaging reliability) need to be maintained over the next fifteen years in spite of the reliability risk inherent in massive technology changes. However, there are also niche markets that require reliability levels to improve. Applications that require higher reliability levels, harsher environments, and/or longer lifetimes are more difficult than the mainstream office and mobile applications. Note that a constant overall chip reliability levels requires a continuous improvement in the reliability per transistor because of scaling. Meeting reliability specifications is a critical customer requirement and failure to meet reliability requirements can be catastrophic.

These customer requirements flow down into requirements for manufacturers that rely on an in-depth knowledge of the physics of all the relevant failure modes and a powerful reliability engineering capability in design-for-reliability, building-in-reliability, and reliability qualification, defect screening and safe-launch methodologies to meet them. 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.

Reliability qualification always involves some risk. There is a risk of qualifying a technology that does not, in fact, meet reliability requirements or a risk of rejecting a technology that does, in fact, meet requirements. At any point in time a qualification can be attempted on a new technology. However, the risk associated with that qualification can be large. The level of risk is directly related to the quality of the reliability physics and reliability engineering knowledge base and capabilities. To mitigate this risk, the concept of robustness validation needs to be exploited further. The combination of thorough failure mode knowledge, modeling and mission profile assessment is meant to minimize the

probability of releasing technologies that have inherent wear-out issues, when properly employed it will lead to shorter qualification times, and lower risks.

The other challenge is that already in the product development and qualification phase, a low PPM level in the early part of the bathtub curve needs to be guaranteed. Samples sizes typically used for qualifications will never be able to supply enough statistics to support such guarantee.

The color-coding of the Reliability technology requirements is meant to represent the reliability risk associated with incomplete knowledge and tools for new materials and devices. The assumption is that there is no problem for the current year (2015) and that the next year is largely ready). The progression from yellow to striped indicates a growing reliability risk. The requirements first turn to yellow (Manufacturing Solutions are Known) in 2014 indicating a relatively small risk associated with scaling, increased power. The requirements then turn to striped (Interim Solutions Known) in 2015 for nBTI and 2017 for pBTI. This date is approximate. The calculated Vmax is extracted from one published paper 0. The result may or may not represent intrinsic values. Technology and therefore the gate stack quality vary from company to company, so would the Vmax. Thus when the calculated Vmax is smaller than the Vdd, the color is not red which represents no known solution to the problem. Only when the difference is significant red color is used. The risk assessment is, naturally, not very reliable for there are a number of known reliability issues that are still poorly understood. A case in point is the strong acceleration of NBTI in the presence of a drain bias, particularly for highly scaled devices. The assessment of moderate risk is a reflection of the awareness level of the problems. Solving these problems requires considerable effort and resources.

Also not included is the point in time where novel devices or materials are introduced (e.g., optical interconnect or a non-CMOS transistor or memory). As mentioned above these changes present a considerable reliability risk and require a considerable lead time to develop the needed capabilities in reliability physics and reliability engineering. Since we do not know exactly what these disruptive technologies will be and when they will be introduced, we have no way of knowing in advance the reliability risk. Solid red reflects the combination of increase variability, unknown reliability behavior from new materials and new structures, and the interaction between them. It signifies the greatly increased unknown rather than known issues that do not have known solution. The poorer the quality of our reliability knowledge is, the greater the reliability risks.

7.5.3 RELIABILITY POTENTIAL SOLUTIONS

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

It is commonly thought that the ultimate nanoscale device will have high degree of variation and high percentage of non-functional devices right from the start. This is viewed as an intrinsic nature of devices at the molecular scale. As a result it will not be possible any longer for designer to take into account a ‘worst case’ design window, because this would jeopardize the performance of the circuits too much. To deal with it, a complete paradigm change in circuit and system design will therefore be needed. While we are not there yet, the increase in variability is clearly already a reliability problem that is taxing the ability of most manufacturers. This is because variability degrades the accuracy of lifetime projection, forcing a dramatic increase in the number of devices tested. The coupling between variability and reliability is squeezing out the benefit of scaling. At some point, perhaps before the end of the roadmap, the cost of ensuring each and every one of the transistors in a large integrated circuit to function within specification may become too high to be practical. As a result, the fundamental philosophy of how to achieve product reliability may need to be changed. This concept is known as resilience, the ability to cope with stress and catastrophe. One potential solution would be to integrate so-called knobs and monitors in the circuits that are sensing circuit parts that are running out of performance and then during runtime can change the biasing of the circuits. Such solutions needs to be further explored and developed. Ultimately, circuits that can dynamically reconfigure itself to avoid failing and failed devices (or to change/improve functionality) will be needed.

Growing complexity of a reliability assessment due to proliferation of new materials, gate stack compositions tuned to a variety of specific applications, as well as shorter cycle for process development, may be alleviated to some degree by greater use of the physics-based microscopic reliability models, which are linked to material structure simulations and consider degradation processes on atomic level. Such models, a need for which is slowly getting wider

recognition, will reduce our reliance on statistical approach, which is both expensive and time consuming, as discussed above. These models can provide additional advantage due to the fact that they can be incorporated in compact modeling tools with a relative ease and required only a limited calibration prior to being applied to a specific product.

Some small changes may already be underway quietly. A first step may be simply to fine-tune the reliability requirements to trim out the excess margin. Perhaps even have product specific reliability specifications. More sophisticated approaches involve fault-tolerant design, fault-tolerant architecture, and fault-tolerant systems. Research in this direction has increased substantially. However, the gap between device reliability and system reliability is very large. There is a strong need for device reliability investigation to address the impact on circuits. Recent increase in using circuits such as SRAM and ring oscillator to look at many of the known device reliability issue is a good sign, as it addresses both the issues of circuit sensitivity as well as variability. More device reliability research is needed to address the circuit and perhaps system aspects. For example, most of the device reliability studies are based on quasi-DC measurements. There is no substantial research on the impact of degradation on devices at circuit operation speed. This gap in measurement speed make modeling the impact of device degradation on circuit performance difficult and risky.

In the meantime, we must meet the conventional reliability requirements. That means an in-depth understanding of the physics of each failure mechanism and the development of powerful and practical reliability engineering tools. Historically, it has taken many years (typically a decade) before the start of production for a new technology generation to develop the needed 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.

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 the pre-existing and adequate reliability knowledge base.

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

The needs are clearly many, but a specific one is the optimal reliability evaluation methodology, which would deliver relevant long-term degradation assessment while preventing excessive accelerated testing which may produce misleading results. This need is driven by the decreasing process margin and increasing variability, which greatly degrades the accuracy of lifetime projection from a standard sample size. The ability to stress a large number of devices simultaneously is highly desirable, particularly for long term reliability characterization. Doing it at manageable cost is a challenge that is very difficult to meet and becoming more so as we migrate to more advanced technology nodes. A break-through in testing technology is badly needed to address this problem.

7.5.4 REFERENCES

[1] Linder, B. P., E. Cartier, S. Krishnan, and E. Wu, "Improving and optimizing reliability in future technologies with high-κ dielectrics", *Int. Symp. VLSI Technology*, Systems, and Applications (VLSI-TSA), 2013.

8 ACKNOWLEDGMENTS

TECHNOLOGY WORKING GROUP KEY CONTRIBUTORS

Cross TWG Study Group (More Moore)

Alan Allan, Mustafa Badaroglu, Joel Barnett, Yuzo Fukuzaki, Paolo Gargini, Christina Gottschalk, Hirofumi Inoue, Atsunobu Isobayashi, Andrew Kahng, SungGeun Kim, Rich Liu, Ichiro Mizushima, Gerhard Klimeck, Dan Mocuta, Yannick Le Tiec, Jurgen Lorenz, Mark Neisser, Ndubuisi George Orji, Kwok Ng, Mehdi Salmani-Jelodar, Prasad Sarangapani, Thomas Skotnicki, Ines Thurner, Hitoshi Wakabayashi, Yanzhong Xu.

ITRS Roadmap Process Integration and Devices & Structures (PIDS) Working Group

Yasushi Akasaka, Hiroaki Ammo, Dimitri Antoniadis, Mustafa Badaroglu, Rock-Hyun Baek, Gennadi Bersuker, Azeez Bhavnagarwala, Frederic Boeuf, Joe Brewer, Alex Burenkov, Jeff Butterbaugh, Chorng-Ping Chang, Kin (Charles) Cheung, Kristin DeMeyer, James Fonseca, Yuzo Fukuzaki, Bill Gallagher, Yasushi Gohou, Michael Haond, Christopher Henderson, Olivier Faynot, Herve Jaouen, Toshiro Hiramoto, Digh Hisamoto, Jim Hutchby, Jiro Ida, Hirofumi Inoue, Kunihiko Iwamoro, Herve Jaouen, Moon-Young Jeong, Malgorzata Jurczak, Naoki Kasai, SungGeun Kim, Gerhard Klimeck, Fred Kuper, Hajime Kurata, Chung Lam, Robert Lander, Laurent Le-Pailleur, Rich Liu, Witek Maszara, Saumitra Mehrotra, Toru Mogami, Kwok Ng, Tak Ning, Masaaki Niwa, Tony Oates, Hidekazu Oda, Sang Hyun Oh, Tatsuya Ohguro, Sam Pan, Jongwoo Park, Thierry Poiroux, Siddharth Potbhare, Kirk Prall, Xiang Qi, Bhagawan Sahu, Mehdi Salmani-Jelodar, Prasad Sarangapani, Thomas Schulz, Saurabh Sinha, Thomas Skotnicki, James Stathis, Eric Snyder, Toshihiro Sugii, Shinichi Takagi, Akira Tanabe, Tetsu Tanaka, Cheng-Tzung Tsai, Wilman Tsai, Hitoshi Wakabayashi, Philip Wong, Qi Xiang, Yanzhong Xu, Geoffrey Yeap, Makoto Yoshimi, Scott Yu, Peter Zeitzoff.

ITRS Roadmap Interconnect Working Group

Nobuo Aoi, Lucile Arnaud, Koji Ban, Hans-Joachim Barth, Eric Beyne, Boyan Boyanov, Jon Candelaria, Chung Liang Chang, Hsien Wei Chen, Wen Chih Chiou, Gilheyun Choi, Jinn P. Chu, Mike Corbett, Alexis Farcy, Paul Feeney, Takashi Hayakawa, Paul Ho, Cheng-Chieh Hsieh, Masayoshi Imai, Atsunobu Isobayashi, Raymond Jao, Shin Puu Jeng, Morihiro Kada, Sibum Kim, Nobuyoshi Kobayashi, Kaushik Kumar, Nohjung Kwak, Hyeon Deok Lee, Anderson Liu, Didier Louis, James Lu, Toshiro Maekawa, David Maloney, Akira Matsumoto, Azad Naeemi, Mehul Naik, Tomoji Nakamura, Yuichi Nakao, Akira Ouchi, Sesh Ramaswami, Hideki Shibata, Michele Stucchi, Zsolt Tokei, Thomas Toms, Manabu Tsujimura, Kazuyoshi Ueno, Osamu Yamazaki, Paul Zimmerman.

ITRS Roadmap Frond-End Processes (FEP) Working Group

Mauro Alessandri, Sehgal Aksey, John Arnold, Souvik Banerjee, Joel Barnett, Twan Bearda, Meridith Beebe, Steve Benner, Ivan (Skip) Berry, Chris Borst, John Boyd; Ian Brown, Arifin Budihardio, George Celler, Luke Chang, Victor Chia, Ho Jin Cho, Phil Clark, Lee Cook, Jeffrey Cross, Adrien Danel, Suman Datta, Carlos H. Diaz, Paul Feeney, Nicholas Fuller, Mike Fury, Yehiel Gotkis, Mike Goldstein, Christina Gottschalk, Aomar Halimaoui, Qingyuan Han, Chris Hobbs, Andrew Hoff, Yoshimasa Horii, Dick James; Raj Jammy, Andreas Kadavanich, Ravi Kanjolia, Hyungsup Kim, Simon Kirk, Brian Kirkpatrick, Hiroshi Kitajima, Martin Knotter, Daniel Koos, Hwasung Kyunggi-Do, Larry Larson, Jeff Lauerhaas, Yannick Le Tiec, Kuntack Lee, Jooh Yun Lee, Tom Lii, Hong Lin, Wei-Yip Lo, Prashant Majhi, Tom McKenna, Kevin McLaughlin, Paul Mertens, Katsuhiko Miki, Ichiro Mizushima, Dan Mocuta, Mansour Monipour, Paul Morgan, Anthony Muscat, Sadao Nakajima, Yasuo Nara, Bich-Yen Nguyen, Masaaki Niwa, Toshihide Ohgata, Hiroshi Oji, Jin Goo Park, Friedrich Passek, Eric Persson, Darryl Peters, Gerd Pfeiffer, Er-Xuan Ping, Jagdish Prasad, Srini Raghavan, Maud Rao, Rick Reidy, Karen Reinhardt, Hwa sung Rhee, Rob Rhoades, Marcello Riva, Jae Sung Roh, Akira Sakai, Archita Sengupta, YugYun Shin, James Shen, Wolfgang Sievert, Chris Sparks, Robert Standley, Sing Pin Tay, Bob Turkot, Allan Upham, Steven Verhaverbeke, Rita Voss, Hitoshi Wakabayashi, Mike Walden, Masaharu Watanabe, Han Xu.

ITRS Roadmap Lithography Working Group

Tsukasa Azuma, Tatsuo Chijimatsu, Masayuki Endo, Ted Fedynyshyn, Heiko Feldmann, Emily Gallagher, Reiner Garreis, Cesar Garza, TS Gau, Frank Goodwin, Roel Gronheid, Naoya Hayashi, Franklin Kalk, Kunihiko Kasama, Insung Kim, Patrick Kearney, Sachiko Kobayashi, Toshio Konishi, Y.C. Ku, Keishiro Kurihara, David Kyser, Shy-Jay Lin, Dan Millward, Matt Malloy, Pawitter Mangat, Venkat Nagaswami, Hideo Nakashima, Patrick Naulleau, Mark Neisser, Yasushi Okubo, Katsumi Ohmori, Masahiko Okumura, Moshe Preil, Doug Resnick, Kazuhiro Takahashi, Hirofumi Takeda, Walt Trybula, Fumikatsu Uesawa, Mauro Vasconi, Geert Vandenberghe, Phil Ware, John Wiesner, Jim Wiley, Grant Willson, Obert Wood, Stefan Wurm, Tetsuo Yamaguchi, Jiro Yamamoto, John Zimmerman.

ITRS Roadmap Metrology Working Group

Carlos Beitia, Mark Berry, Benjamin Bunday, Umberto Celano, SB Chin, Hyun Mo Cho, Alain Diebold, Christina Hacker, Karey Holland, Masahiko Ikeno, Yun Jung Jee, Eiichi Kawamura, Adrain Kiermasz, Joseph Kline, Stephen Knight, Delphine Le-Cunff, Scott List, Seong-Min Ma, Philippe Maillot, Jack Martinez, Niwa Masaaki, Yaw Obeng, George Orji, Brennan Peterson, Narenda Rana, Victor Vartanian, Andras Vladar, Yuichiro Yamazaki.

ITRS Yield Enhancement Working Group

Scott Anderson, Dwight Beal, David Blackford, Yannick Bordel, Marc Camezind, Jan Cavelaars, Jeff Chapman, Hubert Chu, Robert Clark, John DeGenova, Charley Dobson, Arnaud Favre, Guiseppe Fazio, Francois Finck, Dan Fuchs, Takashi Futatsuki, Guillaume Gallet, Delphine Gerin, Astrid Gettel, Rick Godec, Milton Goldwin, Barry Gotlinsky, Mathias Haeuser, Asad Haider, Jeff Hanson, Teruyuki Hayashi, Christoph Hocke, Masahiko Ikeno, Francesca Illuzzi, Hans Jansen, Jost Kames, Barry Kennedy, Keith Kerwin, Suhas Ketkar, Y.J. Kim, Katsunobu Kitami, Kaoru Kondoh, Naoki Kotani, John Kurowski, Sumio Kuwabara, Bob Latimer, Ravi Laxman, Slava Libman, Rushikesh Matkar, Yasuhiko Matsumoto, Fumio Mizuno, William Moore, Chris Muller, Hiroshi Nagaishi, Andreas Neuber, Kazuo Nishihagi, Andreas Nutsch, Kevin Pate, Dilip Patel, Ruben Pessina, Lothar Pfitzner, Larry Rabellino, Dieter Rathei, Rich Riley, David Roberts, Biswanath Roy, Koichiro Saga, Hideyuki Sakaizawa, Tony Schleisman, Yoshimi Shiramizu, Drew Sinha, Terry Stange, Isamu Sugiyama, Paul Tan, Yoshitaka Tatsumoto, Ines Thurner, Hiroshi Tomita, Takahiro Tsuchiya, Ken Tsugane, Dan Wilcox, Hubert Winzig.

ITRS Roadmap Modeling and Simulation Working Group

Nobutoshi Aoki, Jean-Charles Barbe, Francis Benistant, Augusto Benvenuti, Jaehoon Choi, Mauro Ciappa, Derren Dunn, Andreas Erdmann, Roman Gafiteanu, Wladek Grabinski, Tibor Grasser, An De Keersgieter, Yongwoo Kwon, Wolfgang Demmerle, Ronald Gull, Bert Huizing, Herve Jaouen, Dick Klaassen, Gerhard Klimeck, BongHoon Lee, Wilfried Lerch, Jürgen Lorenz, Rainer Minixhofer, Wolfgang Molzer, Victor Moroz, Chandra Mouli, Aneesh Nainani, Yon-Sup Pang, Hwasik Park, Jin Kyu Park, Paul Pfäffli, Peter Pichler, Shigeo Satoh, Yi Ming Sheu, Vivek Singh, I.C. Yang.