INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS 2005 EDITION

SYSTEM DRIVERS

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SYSTEM DRIVERS

SCOPE

Future semiconductor manufacturing and design technology capability is developed in response to economic drivers within the worldwide semiconductor industry. The ITRS must comprehend how technology requirements arise for product classes whose business and retooling cycles drive the semiconductor sector. Until 2001, the ITRS focused on microprocessor (MPU), dynamic random-access memory (DRAM), and application-specific integrated circuit (ASIC) product classes, with some mention of system-on-chip (SOC) and analog/mixed-signal circuits. The unstated assumption was that technological advances needed only be straight-ahead and “linear,” and would be deployed in all semiconductor products. For this reason, specifics of the product classes (e.g., MPU or ASIC) were not required. Today, introduction of new technology solutions is increasingly application-driven, with products for different markets making use of different combinations of technologies at different times. General-purpose digital microprocessors for personal computers have been joined as drivers by mixed-signal systems for wireless communication and embedded applications. Wall-plugged servers are being replaced by battery-powered mobile devices. In-house, single-source chip designs are being supplanted by SOC and system-in-package (SIP) designs that incorporate building blocks from multiple sources.

The purpose of the 2005 ITRS System Drivers Chapter is to update and more clearly define the system drivers as used in previous ITRS editions. Together with the Overall Roadmap Technology Characteristics, the System Drivers Chapter provides consistent framework and motivation for technology requirements across the respective ITRS technology areas and the 15-year span of the ITRS. The main contribution of the chapter consists of quantified, self-consistent models of the system drivers that support extrapolation into future technologies and adapt more smoothly to future technology developments. We focus on four system drivers: system-on-chip, microprocessor, analog/mixed-signal (AMS), and embedded memory. Before describing these drivers, we briefly survey key market drivers for semiconductor products. The reader is also referred to the International Electronics Manufacturing Initiative (iNEMI) roadmap, *http://www.inemi.org*.

MARKET DRIVERS

Table 8 contrasts semiconductor product markets according to such factors as manufacturing volume, die size, integration heterogeneity, system complexity, and time-to-market. Influence on the SOC, AMS, and MPU drivers is noted.[1]

[1] The market drivers are most clearly segmented according to cost, time-to-market, and production volume. System cost is equal to Manufacturing cost + Design cost. Manufacturing cost breaks down further into non-recurring engineering (NRE) cost (masks, tools, etc.) and silicon cost (raw wafers + processing + test). The total system depends on function, number of I/Os, package cost, power and speed. Different regions of the (Manufacturing Volume, Time To Market, System Complexity) space are best served by FPGA, Structured-ASIC, or SOC implementation fabrics, and by single-die or system-in-package integration. This partitioning is continually evolving.

Table 8 Major Product Market Segments and Impact on System Drivers

Market Drivers	*SOC*	*Analog/MS*	*MPU*
I. Portable/consumer			
1. Size/weight ratio: peak in 2004 2. Battery life: peak in 2004 3. Function: 2×/2 years 4. Time-to-market: ASAP	Low power paramount Need SOC integration (DSP, MPU, I/O cores, etc.)	Migrating on-chip for voice processing, A/D sampling, and even for some RF transceiver function	Specialized cores to optimize processing per microwatt
II. Medical			
1. Cost: slight downward pressure (~1/2 every 5 years) 2. Time-to-market: >12 mos 3. Function: new on-chip functions 4. Form factor often not important 5. Durability/safety 6. Conservation/ ecology	High-end products only. Reprogrammability possible. Mainly ASSP, especially for patient data storage and telemedicine; more SOC for high-end digital with cores for imaging, real-time diagnostics, etc.	Absolutely necessary for physical measurement and response but may not be integrated on chip	Often used for programmability especially when real-time performance is not important. Recent advances in multi-core processors have made programmability and real-time performance possible
III. Networking and communications			
1. Bandwidth: 4×/3–4 yrs. 2. Reliability 3. Time-to-market: ASAP 4. Power: W/m^3 of system	Large gate counts High reliability More reprogrammability to accommodate custom functions	Migrating on-chip for MUX/DEMUX circuitry MEMS for optical switching.	MPU cores, FPGA cores and some specialized functions
IV. Defense			
1. Cost: not prime concern 2. Time-to-market: >12 mos 3. Function: mostly on SW to ride technology curve 4. Form factor may be important 5. High durability/safety	Most case leverage existing processors but some requirements may drive towards single-chip designs with programmability	Absolutely necessary for physical measurement and response but may not be integrated on chip	Often used for programmability especially when real-time performance is not important Recent advances in multi-core processors have made programmability and real-time performance possible
V. Office			
1. Speed: 2×/2 years 2. Memory density: 2×/2 years 3. Power: flat to decreasing, driven by cost and W/m^3 4. Form factor: shrinking size 5. Reliability	Large gate counts High speed Drives demand for digital functionality Primarily SOC integration of custom off-the-shelf MPU and I/O cores	Minimal on-chip analog Simple A/D and D/A Video i/f for automated camera monitoring, video conferencing Integrated high-speed A/D, D/A for monitoring, instrumentation, and range-speed-pos resolution	MPU cores and some specialized functions Increased industry partnerships on common designs to reduce development costs (requires data sharing and reuse across multiple design systems)
VI. Automotive			
1. Functionality 2. Ruggedness (external environment, noise) 3. Reliability and safety 4. Cost	Mainly entertainment systems. Mainly ASSP, but increasing SOC for high end using standard HW platforms with RTOS kernel, embedded software.	Cost-driven on-chip A/D and D/A for sensor and actuators Signal processing shifting to DSP for voice, visual Physical measurement ("communicating sensors" for proximity, motion, positioning). MEMS for sensors	

A/D—analog to digital ASSP—application-specific standard product D /A—digital to analog DEMUX—demultiplexer DSP—digital signal processing FPGA—field programmable gate array i/f—intermediate frequency I/O—input/output HW—hardware MEMS—microelectromechanical systems MUX—multiplex RTOS—real-time operating system

SYSTEM ON CHIP DRIVER

SOC is a yet-evolving *product class and design style.* The most important observation is that SOC integrates technology and design elements from other system driver classes (MPU, embedded memory, AMS—as well as reprogrammable logic) into a wide range of high-complexity, high-value semiconductor products. Manufacturing and design technologies for SOC are typically developed originally for high-volume custom drivers. The SOC driver class most closely resembles, and is evolved most directly from, the ASIC category since reduced design costs and higher levels of system integration are its principal goals.[2] The primary difference from ASIC is that in SOC design, the goal is to maximize reuse of existing blocks or "cores"—i.e., minimize the amount of the chip that is newly or directly created. Reused blocks in SOC include analog and high-volume custom cores, as well as blocks of software technology. A key challenge is to invent, create and maintain reusable blocks or cores so that they are available to SOC designers.[3] Economic viability of SOC also requires that validation of reuse-based SOC designs becomes easier than for equivalent "from-scratch" designs.

SOC represents a confluence of previous product classes in several ways. As noted above, SOCs integrate building blocks from the other system driver classes, and are subsuming the ASIC category. The quality gap between full-custom and ASIC/SOC is diminishing: 1) starting in the 2001 ITRS, overall ASIC and MPU logic densities are modeled as being equal; and 2) "custom quality on an ASIC schedule" is increasingly achieved by on-the-fly ("liquid") or tuning-based standard-cell methodologies. Finally, MPUs have evolved into SOCs: 1) MPUs are increasingly designed as cores to be included in SOCs, and 2) MPUs are themselves designed as SOCs to improve reuse and design productivity (as discussed below, the ITRS MPU model has multiple processing cores and resembles an SOC in organization[4]). The most basic SOC challenge is presented by implementation productivity and manufacturing cost, which require greater reuse as well as platform-based design, silicon implementation regularity, or other novel circuit and system architecture paradigms. Another challenge is the heterogeneous integration of components from multiple implementation *fabrics* (such as, reprogrammable, memory, analog and radio frequency (RF), MEMS, and software).

The SOC driver class is characterized by heavy reuse of intellectual property (IP) to improve design productivity, and by system integration that potentially encompasses heterogeneous technologies. SOCs exist to provide low cost and high integration. Cost considerations drive the deployment of low-power process and low-cost packaging solutions, along with fast-turnaround time design methodologies. The latter, in turn, require new standards and methodologies for IP description, IP test (including built-in self-test and self-repair), block interface synthesis, etc. Integration considerations drive the demand for heterogeneous technologies (Flash, DRAM, analog and RF, MEMS, ferroelectric RAM (FeRAM), magnetic RAM (MRAM), chemical sensors, etc.) in which particular system components (memory, sensors, etc.) are implemented, as well as the need for chip-package co-optimization. Thus, SOC is the driver for convergence of multiple technologies not only in the same system package, but also potentially in the same manufacturing process. This chapter discusses the nature and evolution of SOCs with respect to three variants driven respectively by multi-technology integration (MT), high performance (HP), and low power and low cost (LP). This partition is by no means disjointed, but rather reflects separate driving concerns (e.g., low-power design *is* high-performance design, but must also reduce package and system cost).

SOC/SIP MULTI-TECHNOLOGY

The need to build heterogeneous systems on a single chip is driven by such considerations as cost, form-factor, connection speed/overhead, and reliability. Thus, process technologists seek to meld CMOS with MEMS, and other sensors. Process complexity is a major factor in the cost of SOC-MT applications, since more technologies assembled on a single chip requires more complex processing. The total cost of processing is difficult to predict for future new materials and combinations of processing steps. However, cost considerations limit the number of technologies on a given SOC: processes are increasingly modular (e.g., enabling a Flash add-on to a standard low-power logic process), but the modules are not generally "stackable." First integrations of each technology within standard CMOS processes—not necessarily together with other technologies, and not necessarily in volume production—will evolve over time. CMOS integration of the latter technologies (electro-optical, electro-biological) is less certain, since this depends not only on basic technical

[2] *Most digital designs today are considered to be ASICs. ASIC connotes both a business model (with particular "handoff" from design team to ASIC foundry) and a design methodology (where the chip designer works predominantly at the functional level, coding the design at Verilog/very high description language (VHDL) or higher level description languages and invoking automatic logic synthesis and place-and-route with a standard-cell methodology). For economic reasons, custom functions are rarely created; reducing design cost and design risk is paramount. ASIC design is characterized by relatively conservative design methods and design goals (cf. differences in clock frequency and layout density between MPU and ASIC in previous ITRS editions) but aggressive use of technology, since moving to a scaled technology is a cheap way of achieving a better (smaller, lower power, and faster) part with little design risk (cf. convergence of MPU and ASIC process geometries in previous ITRS editions). Since the latter half of the 1990s, ASICs have been converging with SOCs in terms of content, process technology, and design methodology.*

[3] *For example, reusable cores might require characterization of specific noise or power attributes ("field of use," or "assumed design context") that are not normally specified. Creation of an IC design artifact for reuse by others is substantially more difficult (by factors estimated at between 2× and 5×) than creation for one-time use.*

[4] *The corresponding ASIC and structured-custom MPU design methodologies are also converging to a common "hierarchical ASIC/SOC" methodology. This is accelerated by customer-owned tooling business models on the ASIC side, and by tool limitations faced by both methodologies.*

advances but also on SOC-MT being more cost-effective than multi-die SIP alternatives. Today, a number of technologies (MEMS, GaAs) are more cost-effectively flipped onto or integrated side-by-side with silicon in the same module depending also on the area and pin-count restrictions of the respective product (such as Flash, DRAM). Physical scale in system applications (ear-mouth = speaker-microphone separation, or distances within a car) also affect the need for single-die integration, particularly of sensors.

SOC High-performance

Examples of SOC-HP include network processors and high-end gaming applications. Since it reflects MPU-SOC convergence, SOC-HP follows a similar trend as MPU and is not separately modeled here. However, one aspect of SOC-HP merits discussion, namely, that instances in the high-speed networking domain drive requirements for off-chip I/O signaling (which in turn create significant technology challenges to test, assembly and packaging, and design). Historically, chip I/O speed (per-pin bandwidth) has been scaling much more slowly than internal clock frequency. This is partly due to compatibility with existing slow I/O standards, but the primary limitation has been that unterminated CMOS signals on printed circuit boards are difficult to run at significantly greater than 100 MHz due to slow settling times. During the past decade, high-speed links in technology initially developed for long-haul communication networks have found increasing use in other applications. The high-speed I/O eliminates the slow board settling problems by using point-to-point connections and treating the wire as a transmission line. Today the fastest of these serial links can run at 10 Gbit/s per pin.

A high-speed link has four main parts: 1) a transmitter to convert bits to an electrical signal that is injected into the board-level wire, 2) the wire itself, 3) a receiver that converts the signal at the end of the wire back to bits, and 4) a timing recovery circuit that compensates for the delay of the wire and samples the signal on the wire at the right place to get the correct data. Such links are intrinsically mixed-signal designs since receivers, transmitters, and timing recovery all require analog blocks (for example, the voltage control oscillator (VCO) discussed as part of the mixed-signal driver is a key component of a timing recovery circuit). Broadly speaking, high-speed links are used in optical systems, chip-to-chip connections, and backplane connections.

Optical links generally push link performance the hardest; since there are generally a small number of optical signals, these links can tolerate relatively complex and power hungry interface circuits. Today, optical links run at 10 Gbit/s per pin, and are expected to continue to scale up in frequency as projected in the Test chapter (high-speed serial links discussion). Initially, electronics for these links were created in non-CMOS technologies, since CMOS was thought incapable of meeting the high-speed requirements. However, over the past five years, many researchers have developed circuits that can run at 10 Gbit/s. While some papers have demonstrated links that run as fast as 1 fanout-of-4 (FO4) delay per bit, most links run at 2–4 FO4 delays per bit, which yields 10 Gbit/s in the 180 nm generation. Continuing to scale link speed with technology should be possible from the circuits' standpoint, but will become difficult due to parasitics and packaging. Signals at this speed are highly sensitive to any discontinuities in their signal path. Even if controlled impedance packaging is used, vias in the package or board can cause impedance changes that will degrade the signal. The 1–2 pF parasitic capacitance from the electrostatic discharge (ESD) device will also significantly degrade the signal. Thus, continued performance scaling will require significant work in ESD, package, and board design.

Chip-to-chip interconnections communicate information between two chips located on the same board, usually close to each other. The main metric driving the design of these links is not Gbit/s since it is generally possible to use a number of links in parallel to connect these chips. For example, if going twice as fast requires 10× the area and 10× the power, it is better to use two links in parallel. Thus, these links are optimized for performance and cost, not just performance. In general, the highest chip-to-chip link speeds are 2–4 times slower than the highest optical link speeds. Bit times for these links vary dramatically, e.g., point-to-point links are available today with bit times ranging from about 2.5 ns (400 Mbit/s) to .4 ns (2.5 Gbit/s). This wide range of performance reflects dependencies on the number of IO required (higher IO counts have slower speeds), the degree of risk the designer is willing to take, and sometimes an existing I/O standard. Design of robust high-speed I/O is still a mixed-signal problem that cannot be automated or checked with current tools. Thus, many design teams are still conservative when choosing I/O rates. As technology scales and design tools become more robust, bit times should approach 4–8 FO4 delays, but this will require additional circuitry to compensate for package and other parasitic effects.

The last major application for high-speed links is in networking, where two chips on different boards must communicate. The signal path is still point-to-point, but travels from one chip through its package to the local board, through a connector to another board, through another connector to the destination board, and then through that board and receiver package to the receiver chip. For high bandwidth each chip generally has a large number of links, so that performance per unit cost is critical. The principal difference from chip-to-chip links is that the "wire" between the two chips has worse electrical properties. Wire issues are a serious concern as speeds increase through 10 Gbit/s, which are achieved in the

90 nm generation. These I/O considerations also show the trade-off between SOC and SIP solutions in the high-speed area.

SOC Power-Efficient

The SOC-PE is a driver that increasingly represents SOC designs—it includes portable and wireless applications such as smart media-enabled phones or digital camera chips, but also chips for other processing purposes such as high-performance computing and enterprise applications. Figure 11 depicts requirements for various attributes of a power-efficient, consumer-driven, possibly wireless device with multimedia processing capabilities, based in part on the model created by the Japan Semiconductor Technology Roadmap Design Working Group. Key aspects of the model are as follows:

- Its typical application area is electronic equipment categorized as "Mobile Consumer Platforms'" because this application area will make rapid progress in the foreseeable future across semiconductor technology generations.
- From the typical requirements of this type of SOC ("Mobile Consumer Platforms") an explosive increase of processing power is required under some upper bound for battery life.
- As a result, the requirement for processing power will be 1000× in the next ten years, while the requirement for dynamic power consumption will not change noticeably.
- The life cycle of "Mobile Consumer Platform" products is short, and will stay short in the future. Therefore, the design effort cannot be increased—it needs to stay at the current level for the foreseeable future.

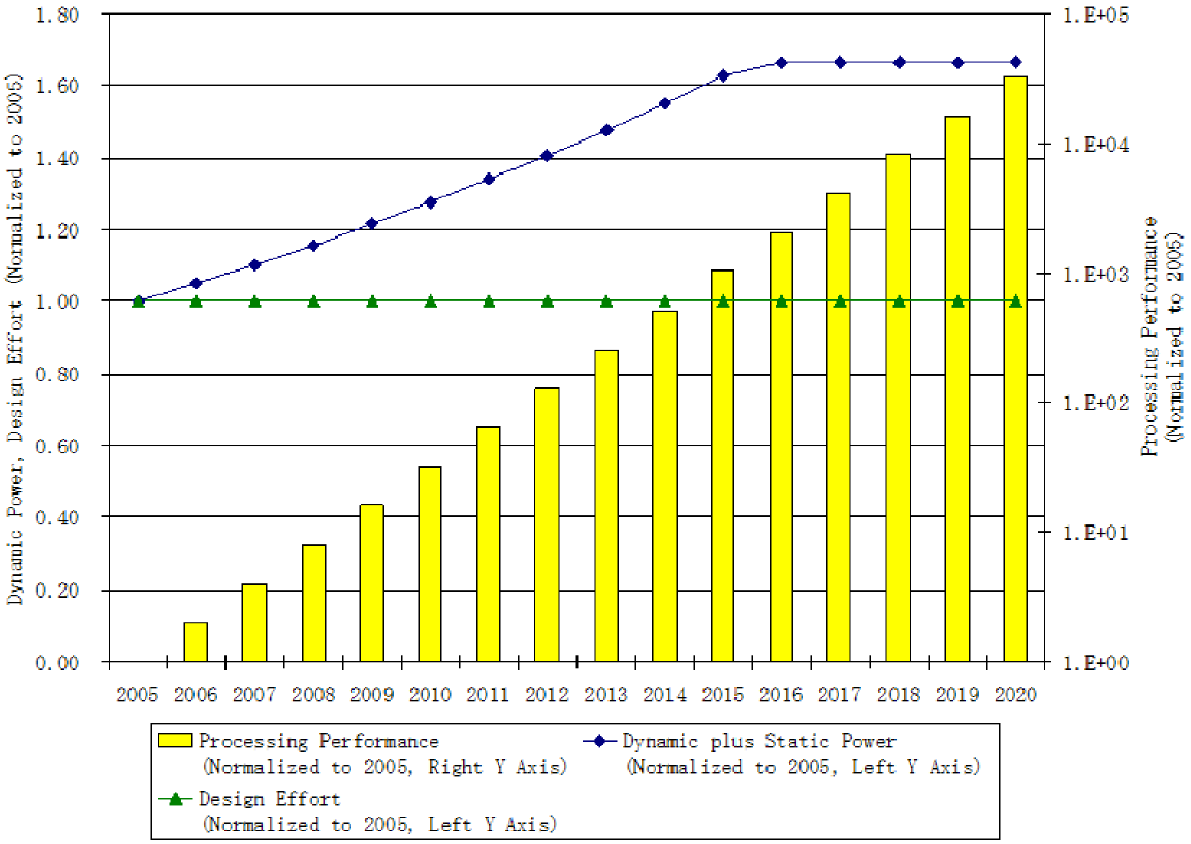

Figure 11 Several Trends for SOC-PE Driver

As shown in Figure 12, a typical power-efficient SOC can be seen under an architecture template for it Mobile Consumer Platform application. The SOC will feature a highly parallel architecture, and consist of a main processor, a number of PEs (Processing Engines), peripherals, and memory.

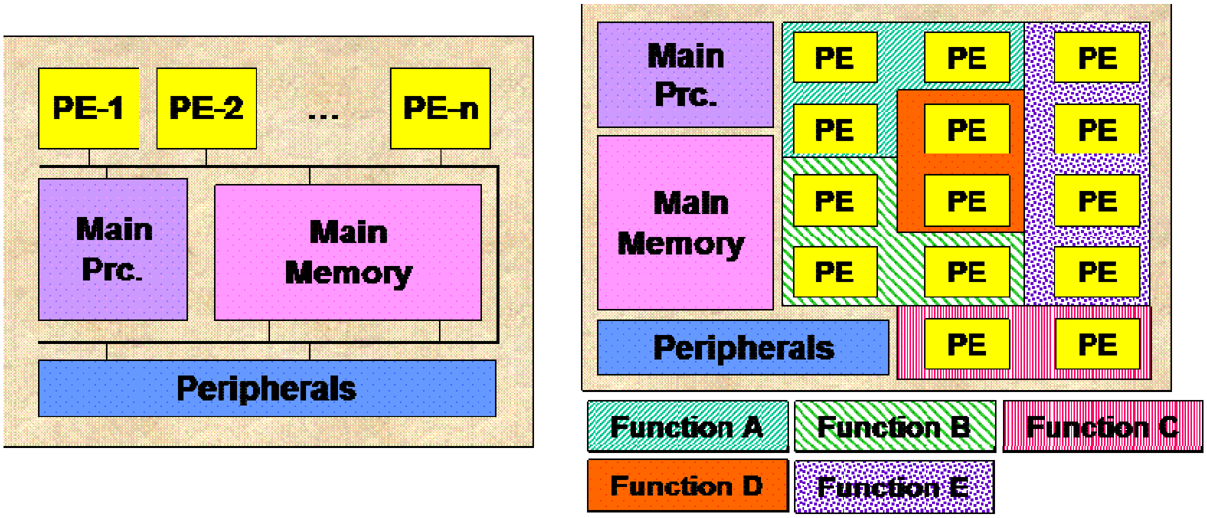

Figure 12 SOC-PE Architecture Template

A PE is a processor customized for a specific function. Functions with large-scale, highly complicated structures will be implemented as a set of PEs. This scheme enables both high processing performance and low-power consumption for the SOC, by virtue of parallel processing and hardware realization of the specific function. A structural characteristic of most future SOCs will be a high number of PEs.

Note that the proposed architecture does not require specific processor array architectures or symmetric processors. The essential feature of the architecture is a high number of PEs embedded on the SOC implementing a set of required functions.

SOC-PE DESIGN COMPLEXITY TRENDS

Based on this template, design-quantified design complexity trends for the SOC-PE driver are shown below in Figure 13. The most interesting factor is the number of PEs and the size of main memory. Specifically, the number of PEs will rapidly grow in subsequent years.

The following are basic assumptions made for these design complexity trends. There will be one main processor with approximately constant complexity. Peripherals will keep the same complexity. As for PEs, the average circuit complexity will stay constant, and the number of PEs will keep growing as long as the die size stays around 64 mm^2. Finally, the amount of main memory will increase proportionally with the numbers of PE.

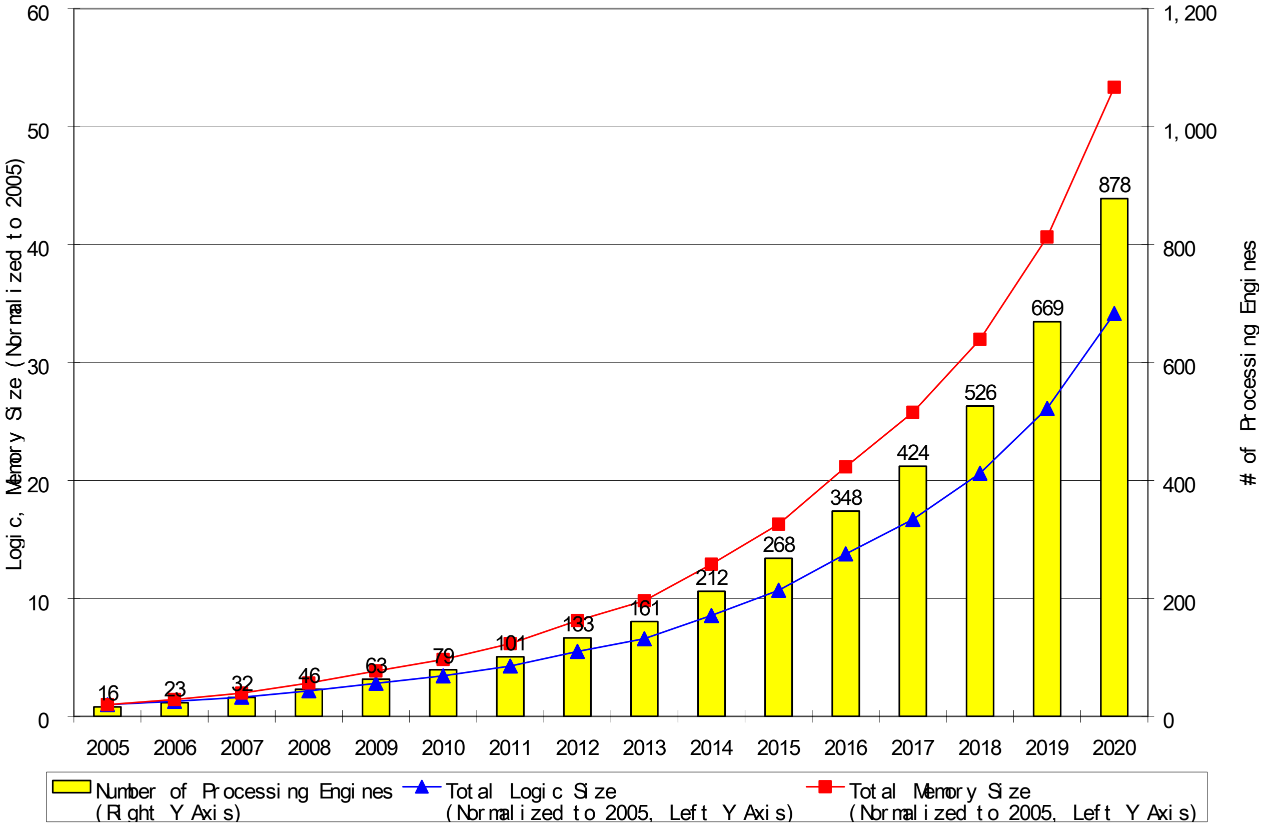

Figure 13 SOC-PE Design Complexity Trends

The following are assumptions made for specific items such as reuse rate and hardware design effort. Required design effort is assumed constant. Design effort is simply assumed to be proportional to the size of the logic circuit portion. Design effort for the reused logic portion is assumed to require half of the effort needed for newly designed logic for equal size. It has been observed that design effort for reuse logic is never free. Its overhead is assumed to be 50%. The reused logic portion requires effort in terms of modifying its functionality and efforts for design steps up to implementation and final physical verification. Design reuse effort is free for non-logic circuits, such as memory and pure analog. Reuse rate is set to 30% in 2005 and 90% in 2020, respectively. Reuse rate for each year between 2005 and 2020 is determined via linear interpolation.

The abovementioned assumptions result in 10× design productivity improvement required for the newly designed portion in the next ten years to 2016, in order to keep design effort constant.

To solve this challenging productivity improvement, several approaches must be combined. First, design abstraction levels must be raised. Second, the level of automation, particularly in design verification and design implementation, must be increased. Finally, the increase in reuse rate is really the key aspect of the solution. However, increasing reuse rate is insufficient. Design overhead for reuse must also be a target for reduction.

Table 9 SOC-PE Design Productivity Trends

	2005	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
Trend: SOC total logic size (normalized to 2005)	**1.00**	**1.27**	**1.62**	**2.15**	**2.81**	**3.42**	**4.27**	**5.50**	**6.58**	**8.54**	**10.69**	**13.77**	**16.69**	**20.62**	**26.12**	**34.15**
Requirement % of reused design	**30%**	**34%**	**38%**	**42%**	**46%**	**50%**	**54%**	**58%**	**62%**	**66%**	**70%**	**74%**	**78%**	**82%**	**86%**	**90%**
Requirement productivity for new designs (normalized to 2005)	**1.00**	**1.24**	**1.54**	**2.00**	**2.54**	**3.02**	**3.67**	**4.59**	**5.34**	**6.73**	**8.18**	**10.2**	**12.0**	**14.3**	**17.5**	**22.1**
Requirement productivity for reused designs (normalized to productivity for new designs at 2005)	**2.00**	**2.48**	**3.08**	**4.00**	**5.09**	**6.04**	**7.33**	**9.19**	**10.7**	**13.5**	**16.4**	**20.4**	**24.0**	**28.6**	**35.0**	**44.2**

SOC-PE Power Consumption Trends

Design complexity is a key trend, but power consumption is a critical factor as well for the design of SOC-LP chips. Figure 14 shows the trend for total chip power, using transistor performance parameters from the PIDS requirements table, interconnect performance parameters from the “Interconnect Technology Requirements” in the *Interconnect chapter*, and circuit complexity parameters from the “Design Complexity Trends” table above. To better understand the trends shown in the figure, we note the following:

- The model applied here simply extrapolates the current level of state-of-the-art technology, hence the resulting power consumption substantially exceeds the requirements.
- Potential solutions are discussed and addressed in the *Design chapter*. Specific solutions for SOC-PE include architecture optimization in high-level design stages based upon power consumption analysis, and customized PE realization.
- Due to voltage supply non-continuous transition in the future, logic dynamic power shows up-and-down transition behavior from 2009 to 2010, from 2012 to 2013, and from 2015 to 2016, respectively.

SOC-PE Processing Performance Trends

As a basic assumption, processing performance will be improved in proportion to the product of device performance itself times the increase in the number of PEs on the SOC.

As shown in Figure 15, the gap between the requirement and the estimate for processing performance spreads out beyond a linear trend. This mentioned gap can be solved by increasing the number of PEs.

Potential solutions will be discussed and addressed in detail in the Design chapter. Important solutions for this driver include appropriate hardware/software (HW/SW) partitioning in high-level design stages, and automated interface technology from high-level design stages to implementation design stages, such as high-level synthesis.

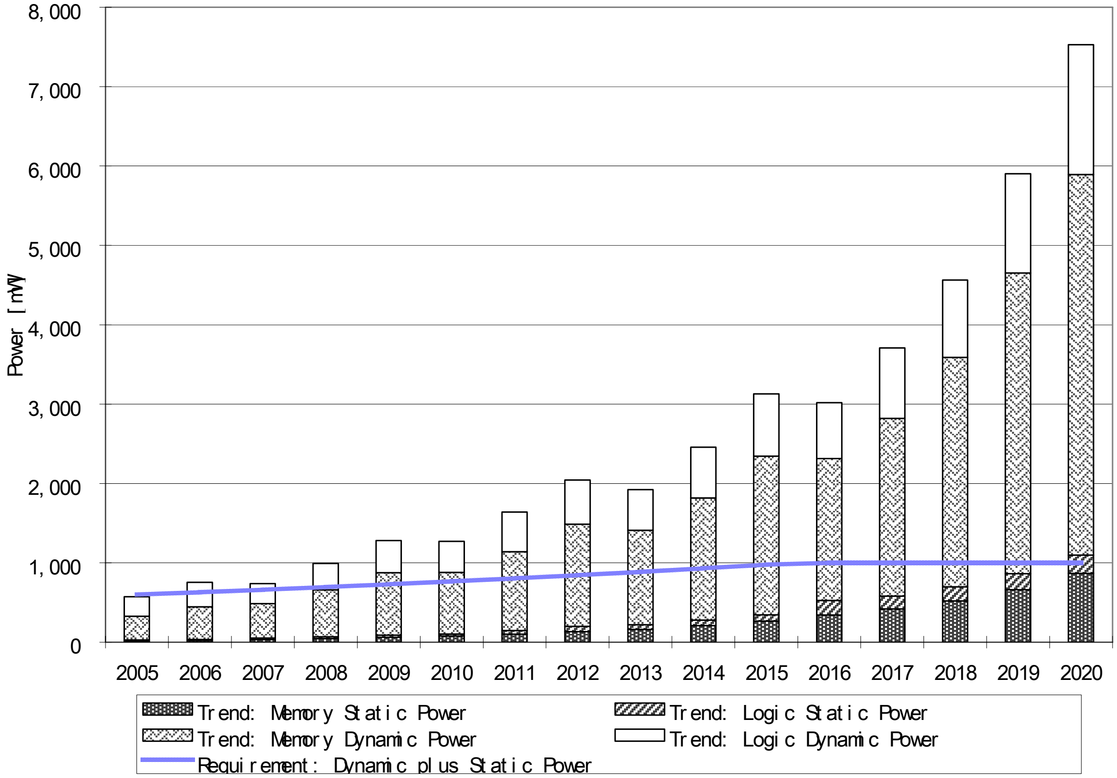

Figure 14 SOC-PE Power Consumption Trends

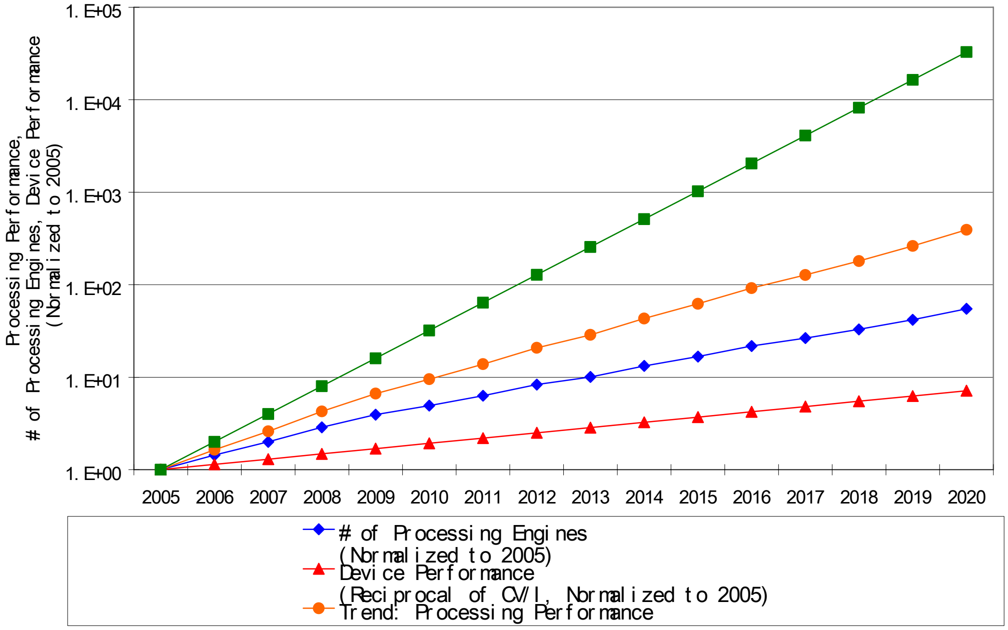

Figure 15 SOC-PE Processing Performance Trends

MICROPROCESSOR (MPU) DRIVER

In high-volume custom designs, performance and manufacturing cost issues outweigh design or other non-recurring engineering cost issues, primarily because of the large profits that these chips can potentially produce. These large profits result from very large sales volumes. Large volumes alone are neither necessary nor sufficient to warrant the custom design style, special process engineering and equipment, etc. often associated with such parts; the key is that the expected return on the combined NRE and manufacturing investment must be positive. Within the high-volume custom arena, three dominant classes today are MPUs, memory[5] and reprogrammable (e.g., FPGA). In this section, MPU is a focus as one of the key system drivers for semiconductor products. MPUs use the most aggressive design styles and manufacturing technologies to achieve their goals. It is for these high-volume parts that changes to the manufacturing flow are made, new design styles and supporting tools are created (the large revenue streams can pay for new tool creation), and subtle circuits issues are uncovered (not all risks taken by designers work out). Indeed, MPUs drive the semiconductor industry with respect to integration density and design complexity, power-speed performance envelope, large-team design process efficiency, test and verification, power management, and packaged system cost. While MPUs (and high-volume custom designs in general) are extremely labor-intensive, they create new technology and automation methods (in both design and fabrication) that are leveraged by the entire industry.

The ITRS MPU driver reflects general-purpose instruction-set architectures (ISAs) that are found standalone in desktop and server systems, and embedded as cores in SOC applications. The MPU system driver is subject to market forces that have historically led to 1) emergence of standard architecture platforms and multiple generations of derivatives, 2) strong price sensitivities in the marketplace, and 3) extremely high production volumes and manufacturing cost awareness. Key elements of the MPU driver model are as follows (studies in this chapter can be run in the *GTX tool*; MPU content is provided in the linked study in the electronic chapter version.

1. *Three types of MPU*—Historically, there have been three types of MPU: 1) cost-performance (CP), reflecting “desktop,” and 2) high-performance (HP), reflecting “server,” and 3) power-connectivity-cost (PCC). As predicted in the 2001 ITRS, the increasing market acceptance of battery-limited mobile designs (often with wireless connectivity) lead to the creation of a new PCC category for MPUs. At the same time, the CP segment that traditionally referred to “desktops” is now expanding to span a much larger portion of the price-performance tradeoff curve, ranging from low-end, low-cost traditional “servers” to “mobile desktops” (i.e., laptops used primarily in AC mode) and “blade” servers. As a consequence, the performance gap between the CP and HP categories is shrinking. However, there will remain a market for truly high-end servers, driving design effort disproportionate to product volume because of large margins involved. As predicted previously, the new PCC category will start taking on characteristics of high-performance, low-power SOC design, with an emphasis on convenience through battery life extension and wireless connectivity. However, the larger margins and volumes of a PCC design will justify much greater design effort as compared to a traditional SOC.
2. *Constant die area*—Die areas are constant (140 mm^2 for CP, 310 mm^2 for HP, 70–100 mm^2 for PCC) over the course of the roadmap, and are broken down into logic, memory, and integration overhead. Integration overhead reflects the presence of white space for interblock channels, floor plan packing losses, and potentially growing tradeoff of layout density for design turnaround time. The core message, in contrast to previous ITRS models, is that power, cost and interconnect cycle latency are strong limiters of die size. Additional logic content would not be efficiently usable due to package power limits, and additional memory content (e.g., larger caches, more levels of memory hierarchy integrated on-chip) would not be cost-effective beyond a certain point.[6] Furthermore, the difficulty of accurate architectural performance simulations with increasingly deeper interconnect pipelining (caused due to process scaling) will also limit die growth size.
3. *Multi-core organization*—MPU logic content reflects multiple processing units on-chip starting at the 130 nm generation, primarily in the HP and high-end CP categories. This integrates several factors: 1) organization of recent and planned commercial MPU products (both server and desktop); 2) increasing need to reuse verification and logic design, as well as standard ISAs; 3) ISA “augmentations” in successive generations (for example, x86, multi-media instructions (MMX), and explicitly parallel instruction computing (EPIC) with likely continuations for encryption, graphics, and multimedia, etc.; 4) the need to enable flexible management of power at the architecture, operating

[5] Memory is a special class of high-volume custom design because of the very high replication rate of the basic memory cells and supporting circuits. Since these cells are repeated millions of times on a chip, and millions of chips are sold, the amount of custom design for these parts is extraordinary. This aspect has led to separate fabrication lines for DRAM devices, with some of the most careful circuit engineering needed to ensure correct operation.

[6] Multi-core organization and associated power efficiencies may permit slight growth in die size, but the message is still that die areas are flattening out.

system (OS) and application levels via SOC-like integration of less efficient, general-purpose processor cores with more efficient, special-purpose “helper engines”[7]; 5) the limited size of processor cores (the estimate of a *constant* 20–25 million transistors per core is a conservative upper bound with respect to recent trends); [8] and 6) the convergence of SOC and MPU design methodologies due to design productivity needs. While increasingly complex single core designs will continue for a few more years, they will compete with equivalent multi-core designs especially in the HP and high-end CP categories. During this period, the number of cores in multi-core designs is projected to double with each successive technology generation.

4. *Memory content*—The MPU memory content is initially 512 KBytes (512 × 1024 × 9 bits) of SRAM for CP and 2 MBytes for HP at 180 nm. Memory content, like logic content, is projected to double with each successive technology generation, not with respect to absolute time intervals (e.g., every 18 months).[9, 10]

5. *Layout density*—Due to their high levels of system complexity and production volume, MPUs are the driver for improved layout density.[11] Thus, MPU driver sets the layout densities, and hence the transistor counts and chip sizes, stated in the Overall Roadmap Technology Characteristics. The logic and SRAM layout densities are analogous to the DRAM “A-factor,” and have been calibrated to recent MPU products. Logic layout densities reflect average standard-cell gate layouts of approximately $320F^2$, where F is the minimum feature size of a technology generation.[12] At 65 nm and below layout density scaling will slow due to the increased complexity of spacing constraints imposed by sub-resolution lithography techniques such as optical proximity correction (OPC) and phase shift mask (PSM). Projections show this impact to be as much as 20% at the 65 nm generation. As a result, the scale factor of 0.7 will yield a density improvement of only 0.55–0.6. As noted above, the logic layout density may improve significantly with the advent of novel devices. SRAM layout densities reflect use of a 6-transistor bit cell (the fitted expression for area per bit cell in units of $F^2 = 223.19F\ (\mu m) + 97.74$) in MPUs, with 60% area overhead for peripheral circuitry.

6. *Maximum on-chip (global) clock frequency*—MPUs also drive maximum on-chip clock frequencies in the Overall Roadmap Technology Characteristics; these in turn drive various aspects of the *Interconnect*, *Process Integration, Devices, and Structures (PIDS), Front End Processes (FEP) and Test* roadmaps. The MPU maximum on-chip clock frequency has historically increased by a factor of 2 per generation. Of this, approximately 1.4× has been from device scaling (which runs into t_{ox} and other limits); the other 1.4× has been from reduction in number of logic stages in a pipeline stage (e.g., equivalent of 32 fanout-of-4 inverter (FO4 INV) delays[13] at 180 nm, and 24–26 FO4 INV delays at 130 nm). There are several reasons why this historical trend will not continue: 1) well-formed clock pulses cannot be generated with period below 6–8 FO4 INV delays; 2) there is increased overhead (diminishing returns) in pipelining (2–3 FO4 INV delays per flip-flop, 1–1.5 FO4 INV delays per pulse-mode latch); 3) thermal envelopes imposed by affordable packaging discourage very deep pipelining, and 4) architectural and circuit innovations will increasingly counter the impact of worsening interconnect RCs (relative to devices) rather than contribute directly to frequency improvements. The 2005 ITRS MPU model continues the historic rate of advance for maximum on-chip

[7] *A “helper engine” is a form of “processing core” for graphics, encryption, signal processing, etc. The trend is toward architectures that contain more special-purpose, and less general-purpose, logic.*

[8] *The CP core has 20 million transistors, and the HP core has 25 million transistors. The difference allows for more aggressive microarchitectural enhancements (trace caching, various prediction mechanisms, etc.) and other performance support.*

[9] *The doubling of logic and memory content with each technology generation, rather than with each 18- or 24-month time interval, is due to essentially constant layout densities for logic and SRAM, as well as conformance with other parts of the ITRS. Specifically, the ITRS remains planar CMOS-centric, there is evidence that non-planar “emerging research devices” are moving into development, possibly as early as 45 nm (VLSI Symp’03). Adoption of such novel device architectures would allow improvements of layout densities beyond what is afforded by scaling alone.*

[10] *Deviation from the given model will likely occur around 90 nm with adoption of denser embedded memories (eDRAM). Adoption of eDRAM, and integrated on-chip L3 cache, will respectively increase the on-chip memory density and memory transistor count by factors of approximately 3 from the given values. While this will significantly boost transistor counts, it is not projected to significantly affect the chip size or total chip power roadmap. Adoption of eDRAM will also depend strongly on compatibility with logic processes (notably the limited process window that arises from scaling of oxide thickness), the size and partitioning of memory within the individual product architecture, and density-performance-cost sensitivities.*

[11] *ASIC/SOC and MPU system driver products have access to similar processes, as forecast since the 1999 ITRS. This reflects emergence of pure-play foundry models, and means that fabric layout densities (SRAM, logic) are the same for SOC and MPU. However, MPUs drive high density and high performance, while SOCs drive high integration, low cost, and low power.*

[12] *A 2-input NAND gate is assumed to lay out in an 8 × 4 standard cell, where the dimensions are in units of contacted local metal pitch (MP = 3.16 × F). In other words, the average gate occupies* $32\times(3.16)^2 = 320F^2$. *For both semi-custom (ASIC/SOC) and full-custom (MPU) design methodologies, an overhead of 100% is assumed.*

[13] *A FO4 INV delay is defined to be the delay of an inverter driving a load equal to 4× its own input capacitance (with no local interconnect). This is equivalent to roughly 14× the CV/I device delay metric that is used in the PIDS Chapter to track device performance. An explanation of the FO4 INV delay model used in the 2005 ITRS is provided as a link.*

global clock frequencies, but flattens the clock period at 12 FO4 INV delays at 90 nm (a plot of historical *MPU clock period data* is provided). This is a change from the projection of 16 FO4 INV delays made in the 2001 ITRS; projections based on circuit and architecture advances indicate that the minimum achievable logic depth is closer to 10–12 F04. The message remains that after the 90 nm generation, clock frequencies will advance only with device performance in the absence of novel circuit and architectural approaches.[14]

MPU Evolution

An emerging "centralized processing" context integrates 1) centralized computing servers that provide high-performance computing via traditional MPUs (this driver), and 2) *interface remedial processors* that provide power-efficient basic computing via, such as SOC integration of RF, analog/mixed-signal, and digital functions within a wireless handheld multimedia platform (refer to the low-power SOC PE model in Figure 14). Key contexts for the future evolution of the traditional MPU are with respect to design productivity, power management, multi-core organization, I/O bandwidth, and circuit and process technology.

Design productivity—The complexity and cost of design and verification of MPU products has rapidly increased to the point where thousands of engineer-years (and a design team of hundreds) are devoted to a single design, yet processors reach market with hundreds of bugs. This aspect is leading to a decreasing emphasis on the use of heavy customization and fancy circuit families resulting in an increasing use of design automation such as logic synthesis and automatic circuit tuning. The resulting productivity increases have allowed processor development schedules and team sizes to flatten out. Improvements in design tools for analysis for timing, noise, power, and electrical rules checking have also contributed to a steady increase in design quality.

Power management—Power dissipation limits of packaging (despite being estimated to reach 200 W/cm^2 by the end of the 2005 ITRS timeframe) cannot continue to support high supply voltages (historically scaling at 0.85× per generation instead of 0.7× ideal scaling) and frequencies (historically scaling by 2× per generation instead of 1.4× ideal scaling).[15] Past clock frequency trends in the MPU system driver have been interpreted as future CMOS device performance (switching speed) requirements that lead to large off-currents and extremely thin gate oxides, as specified in the *PIDS chapter*. Given such devices, MPUs that simply continue existing circuit and architecture techniques would exceed package power limits by factors of nearly 4× by the end of 2020; alternatively, MPU logic content and/or logic activity would need to decrease to match package constraints. Portable and low-power embedded contexts have more stringent power limits, and will encounter such obstacles earlier. Last, power efficiencies (for example, GOPS/mW) are up to four orders of magnitude greater for direct-mapped hardware than for general-purpose MPUs; this gap is increasing. As a result, traditional processing cores will face competition from application-specific or reconfigurable processing engines for space on future SOC-like MPUs.

Multi-core organization—In an MPU with multiple cores per die, the cores can be 1) smaller and faster to counter global interconnect scaling, and 2) optimized for reuse across multiple applications and configurations. Multi-core architectures allow-power savings as well as the use of redundancy to improve manufacturing yield.[16] Organization of the MPU model also permits increasing amounts of the memory hierarchy on chip (consistent with processor-in-memory, or large on-chip eDRAM L3 starting in the 90 nm generation). Higher memory content can, if only in a relatively trivial way, afford better "control" of leakage and total chip power.

[14] *Unlike the ITRS clock frequency models used through 2000 (refer to Fisher/Nesbitt 1999), the 2005 model does not have any local or global interconnect component in its prototypical "critical path". This is because local interconnect delays are negligible, and scale with device performance. Furthermore, buffered global interconnect does not contribute to the minimum clock period since long global interconnects are pipelined—i.e., the clock frequency is determined primarily by the time needed to complete local computation loops, not by the time needed for global communication. Pipelining of global interconnects will become standard as the number of clock cycles required to signal cross-chip continues to increase beyond 1. "Marketing" emphases for MPUs necessarily shift from "frequency" to "throughput" or "utility."*

[15] *To maintain reasonable packaging cost, package pin counts and bump pitches for flip-chip are required to advance at a slower rate than integration densities (refer to the Assembly and Packaging chapter). This increases pressure on design technology to manage larger wakeup and operational currents and larger supply voltage IR drops; power management problems are also passed to the architecture, OS, and application levels of the system design.*

[16] *Replication enables power savings through lowering of frequency and V_{dd} while maintaining throughput (e.g., two cores running at half the frequency and half the supply voltage will save a factor of 4 in CV^2f dynamic capacitive power, versus the "equivalent" single core). (Possibly, this replication could allow future increases in chip size.) More generally, overheads of time-multiplexing of resources can be avoided, and the architecture and design focus can shift to better use of area than memory. Redundancy-based yield improvement occurs if, for example, a die with k-1 instead of k functional cores is still useful.*

Evolutionary microarchitecture changes (super-pipelining, super-scalar, predictive methods) appear to be running out of steam. ("*Pollack's Rule*" observes that in a given process technology, a new microarchitecture occupies 2–3× the area of the old (previous-generation) microarchitecture, while providing only 1.4–1.6× the performance.) Thus, more multithreading support will emerge for parallel processing, as well as more complex "hardwired" functions and/or specialized engines for networking, graphics, security, etc. Flexibility-efficiency tradeoff points shift away from general-purpose processing.

Input/output bandwidth—In MPU systems, I/O pins are mainly used to connect to memory, both high-level cache memory and main system memory. Increased processor performance has been pushing I/O bandwidth requirements. The highest-bandwidth port has traditionally been used for L2 or L3 cache, but recent designs are starting to integrate the memory controller on the processor die to reduce memory latency. These direct memory interfaces require more I/O bandwidth than the cache interface. In addition to the memory interface, many designs are replacing the system bus with high-speed point-to-point interfaces. These interfaces require much faster I/O design, exceeding Gbit/s rates. While serial links have achieved these rates for a while, integrating a large number of these I/O on a single chip is still challenging for design (each circuit must be very low power), test (need to have a tester that can run this fast) and packaging (packages must act as balanced transmission lines, including the connection to the chip and the board).

Circuit and process technology—Parametric yield ($/wafer after bin-sorting) is severely threatened by the growing process variability implicit in feature size and device architecture roadmaps, *Lithography* and *PIDS*, including thinner and less reliable gate oxides, subwavelength optical lithography requiring aggressive reticle enhancement, and increased vulnerability to atomic-scale process variability (e.g., implant). This will require more intervention at the circuit and architecture design levels. Circuit design use of dynamic circuits, while attractive for performance in lower-frequency or clock-gated regimes, may be limited by noise margin and power dissipation concerns; less pass gate logic will be used due to body effect. Error-correction for single-event upset (SEU) in logic will increase, as will the use of redundancy and reconfigurability to compensate for yield loss. Design technology will also evolve to enable consideration of process variation during design and analysis and its impact on parametric yield (bin-splits). The need for power management will require a combination of techniques from several component technologies:

- Application-, OS- and architecture-level optimizations including parallelism and adaptive voltage and frequency scaling
- Process innovations including increased use of silicon-on-insulator (SOI)
- Circuit design techniques including the *simultaneous* use of multi-V_{th}, multi-V_{dd}, minimum-energy sizing under throughput constraints, and multi-domain clock gating and scheduling
- Novel devices that decrease leakage

MPU CHALLENGES

The MPU driver strongly affects design and test technologies (distributed/collaborative design process, verification, at-speed test, tool capacity, power management), as well as device (off-current), lithography/FEP/interconnect (variability) and packaging (power dissipation and current delivery). The most daunting challenges are:

- *Design and verification productivity* (e.g., total design cost, number of bug escapes) (Design)
- *Power management and delivery* (e.g., gigaoperations per second (GOPS) per mW) (Design, PIDS, Assembly and Packaging)
- *Parametric yield at volume production* (Lithography, PIDS, FEP, Design)

MIXED-SIGNAL DRIVER

Analog/mixed-signal chips are those that at least partially deal with input signals whose precise values matter. This broad class includes RF, analog, analog-to-digital and digital-to-analog conversion, and, more recently, a large number of mixed-signal chips where at least part of the chip design needs to measure signals with high precision. These chips have very different design and process technology demands than digital circuits. While technology scaling is always desirable for digital circuits due to reduced power, area and delay, it is not necessarily helpful for analog circuits since dealing with precision requirements or signals from a fixed voltage range is more difficult with scaled voltage supplies. Thus, scaling of analog circuits into new technologies is a difficult challenge. In general, AMS circuits (such as RF and analog design styles) and process technologies (e.g., silicon-germanium, embedded passives) present severe challenges to cost-effective

CMOS integration. However, clever system combinations of analog and digital circuitry also offer potential for functionality and cost scaling at almost the same rate as digital circuits.

The need for precision also affects tool requirements for analog design. Digital circuit design creates a set of rules that allow logic gates to function correctly: as long as these rules are followed, precise calculation of exact signal values is not needed. Analog designers, on the other hand, must be concerned with a large number of “second-order effects” to obtain the required precision. Relevant issues include coupling (capacitance, inductance, resistance and substrate affecting the integrity of signals and supply voltages) and asymmetries (local variation of implantation, alignment, etching, and other fabrication steps all affect the predictability of the electrical performance). Analysis tools for these issues are mostly in place, but require expert users and their accuracy are still insufficient for many problems both for low-power analog and high speed mixed-signal and RF design. Synthesis tools are preliminary and should concentrate on analog specific layout synthesis. Manufacturing test for AMS circuits still needs to be improved but the trend towards SOC also gives opportunities for analog built-in self test (BIST).

Most analog and RF circuitry in today’s high-volume applications is part of SOCs. The economic regime of a mainstream product is usually highly competitive—it has a high production volume, and hence a high level of R&D investment by which its technology requirements can drive mixed-signal technology as a whole. Mobile communication platforms are the highest volume circuits driving the needs of mixed signal circuits. When formulating an analog and mixed-signal (AMS) roadmap, simplification is necessary because there are many different circuits and architectures. This section discusses four basic analog circuits. Those are not only most critical components, but their performance requirements are also representative and most important for RF and analog parts of the SOC:

1. Low-noise amplifier (LNA)

2. Voltage-controlled oscillator (VCO)

3. Power amplifier (PA)

4. Analog-to-digital converter (ADC)

The design and process technology used to build these basic analog circuits also determines the performance of many other mixed-signal circuits. Thus, the performance of these four circuits, as described by figures of merit (FoMs), is a good basis for a mixed-signal roadmap.

The following discussion develops these FoMs in detail. Unless otherwise noted, all parameters (e.g., gain *G*) are given as absolute values and not on a decibel scale. Preferences for specific solutions to given design problems are avoided; rather, different types of solutions are encouraged since unexpected solutions have often helped to overcome barriers. (Competition, such as between alternative solutions, is a good driving force for all types of advances related to technology roadmapping.) Any given type of circuit will have different requirements depending on its purposes. Therefore, certain performance indicators can be contradictory in different applications.[17] To avoid such situations, the figures of merit correlate to the analog and RF needs of a mobile communication platform. Last, this section evaluates the dependence of the FoMs on device parameters, so that circuit design requirements can lead to specific device and process technology specifications. Extrapolations are proposed that lead to a significant advance of analog circuit performance as well as to realistic and feasible technology advances. These parameters are given in the *RF and Analog/Mixed-signal Technologies for Wireless Communications chapter*.

LOW-NOISE AMPLIFIER (LNA)

Digital processing systems require interfaces to the analog world. Prominent examples of these interfaces are transmission media in wired or wireless communication. The LNA amplifies the input signal to a level that makes further signal processing insensitive to noise. The key performance issue for an LNA is to deliver the undistorted but amplified signal to downstream signal processing units without adding further noise.

LNA applications (global standard for mobile (GSM), code division multiple access (CDMA), wireless local area network (WLAN), global positioning system (GPS), Bluetooth, etc.) operate in many frequency bands. The operating frequency and, in some cases, the operating bandwidth of the LNA will impact the maximum achievable performance; nonlinearity must also be considered to meet the specifications of many applications. These parameters must be included in the FoM. On the other hand, different systems may not be directly comparable, and have diverging requirements. For example, very

[17] *Certain cases of application are omitted for the sake of simplicity, and arguments are given for the cases selected. Considerations focus on CMOS since it is the prime technological driving force and in most cases the most important technology. Alternative solutions (especially other device families) and their relevance will be discussed for some cases, as well as at the end of this section.*

wide bandwidth is needed for high-performance wired applications, but this increases power consumption. Low power consumption is an important design attribute for low-bandwidth wireless applications. For wide-bandwidth systems, bandwidth may be more important than linearity to describe the performance of an LNA. To avoid contradictory design constraints, the *wireless* communication context is presented.

The linearity of a low noise amplifier can be described by the output referenced third order intercept point ($OIP3 = G \times IIP3$ where G is the gain and $IIP3$ is the input referenced third order intercept point). A parameter determining the minimum signal that is correctly amplified by a LNA is directly given by the noise figure of the amplifier, NF. However, (NF-1) is a better measure of the contribution of the amplifier to the total noise, since it allows the ratio between the noise of the amplifier $N_{\mathrm{amplifier}}$ and the noise already present at the input N_{input} to be directly evaluated. These two performance figures can be combined with the total power consumption P. The resulting figure of merit captures the dynamic range of an amplifier versus the necessary DC power. For roadmapping purposes it is preferable to have a performance measure that is independent of frequency and thus independent of the specific application. This can be achieved by assuming that the LNA is formed by a single amplification stage, so that the FoM scales linearly with operating frequency *f*. With these approximations and assumptions, a figure of merit ($\mathrm{FoM}_{\mathrm{LNA}}$) for LNAs is defined:

$$FoM_{LNA} = \frac{G \cdot IIP3 \cdot f}{(NF-1) \cdot P} \qquad [1]$$

Making further simplifying assumptions, and neglecting "design intelligence", the evolution of the FoM with technology scaling can be extrapolated.[18] Future trends of relevant device parameters for LNA design, including maximum oscillation frequency f_{max}, quality of inductors, inner gain of the MOSFETs ($g_m/g_{ds}|_{L\_min}$), and RF supply voltages are shown in the *RF and Analog/Mixed-signal Technologies for Wireless Communications chapter*. Extrapolating these data into the future, an estimate of future progress in LNA design is obtained as shown in Table 10. Especially linearity issues in long-term future may increasingly be solved by digital calibration techniques.

VOLTAGE CONTROL OSCILLATOR

The VCO is the key part of a phase-locked loop (PLL), which synchronizes communication between an integrated circuit and the outside world in high-bandwidth and/or high-frequency applications. The key design objectives for VCOs are to minimize the timing jitter of the generated waveform (or, equivalently, the phase noise) and to minimize the power consumption. From these parameters a figure of merit ($\mathrm{FoM}_{\mathrm{VCO}}$) is defined:

$$FoM_{VCO} = \left(\frac{f_0}{\Delta f}\right)^2 \frac{1}{L\{\Delta f\} \cdot P} \qquad [2]$$

Here, f_0 is the oscillation frequency, $L\{\Delta f\}$ is the phase noise power spectral density measured at a frequency offset Δf from f_0 and taken relative to the carrier power, and P is the total power consumption.

This definition does not contain the absolute value of the operating frequency since there is no clear correlation between the operating frequency and the figure of merit. The definition also neglects the tuning range of the VCO since the necessary tuning range strongly depends on the application. In this tuning range, $\mathrm{FoM}_{\mathrm{VCO}}$ should be evaluated at the frequency where phase noise is maximal.

Phase noise is mainly determined by thermal noise of the active and passive components in the VCO, the quality factor of the LC tank, the amplitude of the oscillation, and, close to the carrier frequency, by the 1/f noise of the active components of the VCO. $\mathrm{FoM}_{\mathrm{VCO}}$ is roughly proportional to the overdrive voltage of the active elements in the VCO, inversely proportional to VDD, and proportional to the square of the quality factor of the LC tank. The value of the chosen overdrive voltage is a compromise between minimization of the contribution of 1/f noise and keeping the amplitude of the oscillation sufficiently high. In this way, $\mathrm{FoM}_{\mathrm{VCO}}$ is linked to technology development. Based on a prediction of the relevant device parameters for future technology generations (see the data in the *RF and Analog/Mixed-signal Technologies for Wireless Communications chapter.*), an extrapolation of the VCO FoM for future technology

[18] *R. Brederlow, S. Donnay, J. Sauerer, M. Vertregt, P. Wambacq, and W. Weber, "A mixed signal design roadmap for the International Technology Roadmap for Semiconductors (ITRS)," IEEE Design and Test, December 2001.*

generations is given in Table 10. In addition to those technology scaling related trends, a further design-related trend towards digital controlled oscillators is observed where frequency is tuned by switching of capacitors.

POWER AMPLIFIER

Power amplifiers (PAs) are key components in the transmission path of wired or wireless communication systems. They deliver the transmission power required for transmitting information off-chip with high linearity to minimize adjacent channel power. For battery-operated applications in particular, minimum DC power at a given output power is required.

CMOS PAs due to technological issues are restricted to applications where relatively small transmit power is needed. For discrete PAs with higher transmit power (maybe integrated within a SIP), other technologies like bipolar or compound semiconductor technologies have advantages (*RF and Analog/Mixed-signal Technologies for Wireless Communications chapter*).

To establish a performance figure of merit, several key parameters must be taken into account. These include output power P_{out}, power gain G, carrier frequency *f*, linearity (in terms of IIP3), and power-added-efficiency (PAE). Unfortunately, linearity strongly depends on the operating class of the amplifiers, making it difficult to compare amplifiers of different classes. In addition linearity issues in future may increasingly be solved by digital calibration techniques. To remain independent of the design approach and the specifications of different applications, this parameter is omitted in the figure of merit. To compensate for the 20 dB/decade roll-off[19] of the PA's RF-gain, a factor of f^2 is included into the figure of merit. This results in:

$$FoM_{PA} = P_{out} \cdot G \cdot PAE \cdot f^2 \qquad [3]$$

Finally, restricting to the simplest PA architecture (class A operation)[20] and making further simplifications enables correlation between the FoM and device parameters.[21] The key device parameters are seen to be the quality factor of the available inductors and f_{max}. FoMs of best-in-class CMOS PAs have increased by approximately a factor of two per technology generation in recent years strongly correlated with progress in active and passive device parameters. From required device parameters for future technology generations (see the *Power Amplifier Tables in RF and Analog/Mixed-signal Technologies for Wireless Communications chapter*), we can deduce requirements for future PA FoM values, as shown in Table 10.

ANALOG-TO-DIGITAL CONVERTER

Digital processing systems have interfaces to the analog world—audio and video interfaces, interfaces to magnetic and optical storage media, and interfaces to wired or wireless transmission media. The analog world meets digital processing at the ADC, where continuous-time and continuous-amplitude analog signals are converted to discrete-time (sampled) and discrete-amplitude (quantized). The ADC is therefore a useful vehicle for identifying advantages and limitations of future technologies with respect to system integration. It is also the most prominent and widely used mixed-signal circuit in today's integrated mixed-signal circuit design.

The main specification parameters of an ADC relate to sampling and quantization. The resolution of the converter, i.e., the number of quantization levels, is 2^n where n is the "number of bits" of the converter. This parameter also defines the maximum signal to noise level $SNR = n \cdot 6.02 + 1.76 \quad [dB]$. The sampling rate of the converter, i.e., the number of *n*-wide samples quantized per unit time, is related to the bandwidth that needs to be converted and to the power consumption required for reaching these performance points. The Shannon/Nyquist criterion states that a signal can be reconstructed whenever the sample rate exceeds twice the converted bandwidth: $f_{sample} > 2 \times BW$.

To yield insight into the potential of future technology generations, the ADC FoM should combine dynamic range, sample rate f_{sample} and power consumption *P*. However, these nominal parameters do not give accurate insight into the effective performance of the converter; a better basis is the effective performance extracted from measured data. Dynamic

[19] *Most CMOS PAs are currently operated in this regime. Using DC-gain for applications far below* f_t *would result in a slightly increased slope.*

[20] *R. Brederlow, S. Donnay, J. Sauerer, M. Vertregt, P. Wambacq, and W. Weber, "A mixed signal design roadmap for the International Technology Roadmap for Semiconductors (ITRS)," IEEE Design and Test, December 2001.*

[21] *R. Brederlow, S. Donnay, J. Sauerer, M. Vertregt, P. Wambacq, and W. Weber, "A mixed signal design roadmap for the International Technology Roadmap for Semiconductors (ITRS)," IEEE Design and Test, December 2001.*

range is extracted from low frequency signal-to-noise-and-distortion (SINAD$_0$) measurement minus quantization error (both values in dB). From SINAD$_0$ an “effective number of bits” can be derived as $ENOB_0 = (SINAD_0 - 1.76)/6.02$. Then, the sample rate may be replaced by twice the effective resolution bandwidth (2×ERBW) if it has a lower value, to establish a link with the Nyquist criterion:

$$FoM_{ADC} = \frac{\left(2^{ENOB_0}\right) \times \min(\{f_{sample}\},\{2 \times ERBW\})}{P} \qquad [4]$$

For ADCs, the relationship between FoM and technology parameters is strongly dependent on the particular converter architecture and circuits used. The complexity and diversity of ADC designs makes it nearly impossible to come up with a direct relationship, as was possible for the basic RF circuits. Nevertheless, some general considerations regarding the parameters in the FoM are proposed,[22] in some cases, it is possible to determine performance requirements of the design from the performance requirements of a critical subcircuit. The device parameters are relevant for the different ADC designs (refer to the data in the *RF and Analog/Mixed-signal Technologies for Wireless Communications chapter*). The trend in recent years shows that the ADC FoM improves by approximately a factor of 2 every three years. Taking increasing design intelligence into account, these past improvements are in good agreement with improvements in analog device parameters. Current best-in-class is approximately 1600 [giga-conversion-steps per second and watt] for stand-alone CMOS/bipolar CMOS (BiCMOS), and approximately 800 [giga-conversion-steps per second and watt] for embedded CMOS. Expected future values for the ADC FoM are shown in Table 10. Major advances in design are needed to maintain performance increases for ADCs in the face of decreased voltage signal swings and supplies. In the long run, fundamental physical limitations (thermal noise) may block further improvement of the ADC FoM.

Table 10 Projected Mixed-Signal Figures of Merit for Four Circuit Types.

Year of Production	*2005*	*2008*	*2011*	*2014*	*2017*	*2020*	*Driver*
RF-CMOS ½ Pitch	*90*	*65*	*45*	*32*	*22*	*18*	
FoM_{LNA} [GHz]	*80*	*160*	*200–300*	*300–400*	*400–600*	*500–700*	*Refer to the RF and AMS Technologies for Wireless chapter*
FoM_{VCO} [1/J] 10^{22}	*0.9*	*1.1*	*1.5*	*2*	*2.4*	*3*	
FoM_{PA} [W× GHz^2] 10^4	*10*	*20*	*40*	*60–80*	*90–100*	*110–130*	
FoM_{ADC} [GHz/W]10^3	*1.2*	*2*	*3–4*	*4–10*	*6–20*	*8–40*	

Mixed-Signal Evolution

Evolution of the mixed-signal driver, including its scope of application, is completely determined by the interplay between cost and performance. The figures of merit in Table 10 measure mixed-signal *performance*. However, *cost* of production is also a critical issue for practical deployment of AMS circuits. Together, cost and performance determine the sufficiency of given technology trends relative to existing applications, as well as the potential of given technologies to enable and address entirely new applications.

Cost estimation—Unlike high-volume digital products where cost is mostly determined by chip area, in mixed-signal designs area is only one of several cost factors. The area of analog circuits in an SOC is typically in the range of 5–30%; economic forces to reduce mixed-signal area are therefore not as strong as for logic or memory. Related considerations include:

- Analog area can sometimes be reduced by shifting the partitioning of a system between analog and digital parts (for example, auto-calibration of ADCs, linearity tuning of PAs)
- Process complexity is increased by introducing high-performance analog devices, so that solutions can have less area but greater total cost
- Technology choices can impact design cost by introducing greater risk of multiple hardware passes (tapeout iterations)
- Manufacturing cost can also be impacted via parametric yield sensitivities

[22] *R. Brederlow, S. Donnay, J. Sauerer, M. Vertregt, P. Wambacq, and W. Weber, “A Mixed-signal Design Roadmap for the International Technology Roadmap for Semiconductors (ITRS),” IEEE Design and Test, December 2001.*

- A SIP solution with multiple die (e.g., large, low-cost digital and small, high-performance analog) can be cheaper than a single SOC solution

Such considerations make cost estimation very difficult for mixed-signal designs. It is possible to quantify mixed-signal cost by first restricting our attention to high-performance applications, since these also drive technology demands. Next, note that analog features are embodied as high-performance passives or analog transistors, and that area can be taken as a proxy for cost.[23] Since scaling of transistors is driven by the need to improve density of the digital parts of a system, analog transistors can simply follow, thus rendering it unnecessary to specifically address their layout density. At the same time, total area in most current AMS designs is determined by embedded passives; their area consumption dominates the cost of the mixed-signal part of a system. Therefore, the tables in the *Wireless Chapter* set a roadmap of layout density for on-chip passive devices that is necessary to improve the cost/performance ratio of high-performance mixed-signal designs.

Estimation of technology sufficiency—Figure 16 shows ADC requirements for recent applications in terms of a power/performance relationship. Under conditions of constant performance (resolution × bandwidth), a constant power consumption is represented by a straight line with slope –1. Increasing performance—which is achievable with better technology or circuit design—is equivalent to a shift of the power consumption lines towards the upper right. The data show a technological "barrier-line" moving with an order of magnitude per ten years (Table 10) for ADCs for a power consumption of 1W. Most of today's ADC technologies (silicon, SiGe, and III-V compound semiconductor technologies and their hybrids) lie below the 1W barrier-line, and though near-term solutions for moving the barrier-line more rapidly are unknown, the 2005 position (1 GHz/milliWatt) of the barrier enables emerging high data-rate communication fields with acceptable dissipation in the conversion function.

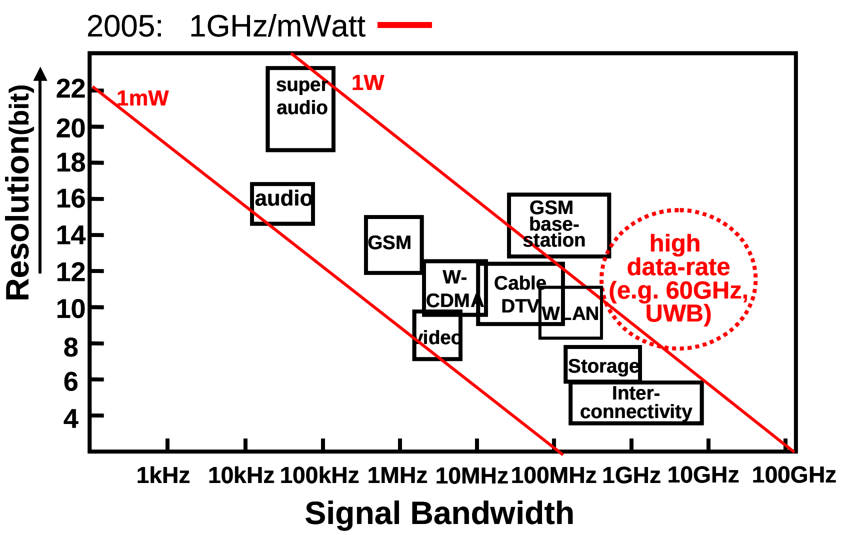

Figure 16 Recent ADC Performance Needs for Important Product Classes

While the rate of improvement in ADC performance has been adequate for handset applications, this is clearly not the case for applications such as digital linearization of GSM base-stations, or handheld/mobile high-data rate digital video applications. For example, a multi-carrier GSM base-station with a typical setup of 32 carriers requires over 80 dB of dynamic range. Implementing digital linearization in such a base-station with a 25 MHz transmitter band requires ADCs that have sampling rates of 300 MHz and 14 bits of resolution at a power consumption of less than 1W. According to Table 10 and assuming progress at recent rates, it will be perhaps until after 2010 before ADCs with such performance are manufactured in volume. While system designers would like to have such ADCs now, silicon and SiGe technologies have the necessary bit resolution (large numbers of devices per unit area) but not the speed; on the other hand, III-V compound semiconductor technologies have the speed but not the bit resolution. This motivates consideration of solutions that potentially increase the rate of ADC improvement at reasonable costs—e.g., use of hybrids of both CMOS and compound semiconductor technologies. For applications that need high performance PAs often a SIP solution with Si-Ge

[23] *In analog designs, power consumption is often proportional to area—and since power is included in all four figures of merit, area and cost criteria are considered. Nonetheless, area requirements should be stated explicitly in a roadmap.*

heterojunction bipolar transistors (HBTs) and III-V devices for the PA and CMOS for the other parts of the analog front-end are the best choice.

Enabling new applications—For a given product, the usual strategy to increase unit shipments is to reduce cost while increasing product performance. However, this is not the only driver for the semiconductor business, especially for products that include mixed-signal parts. Rather, improving technology and design performance enables *new* applications (comparable to the realization of the mobile handset in recent years), thus pushing the semiconductor industry into new markets. Analysis of mixed-signal designs as in Figure 12 can also be used to estimate design needs and design feasibility for future applications and new markets. We see that increasing performance is equivalent to the ability to develop new products that need higher performance or lower power consumption than is available in today's technologies. Alternatively, when specifications of a new product are known, one can estimate the technology needed to fulfill these specifications, and/ or the timeframe in which the semiconductor industry will be able to build that product with acceptable cost and performance. In this way, the FoM concept can be used to evaluate the feasibility and the market of potential new mixed-signal products. The ability to build high performance mixed-signal circuitry at low cost will continuously drive the semiconductor industry into such new products and markets.

MIXED-SIGNAL CHALLENGES

For most of today's mixed-signal designs—and particularly in classical analog design—the processed signal is represented by a voltage difference, so that the supply voltage determines the maximum signal. Decreasing supplies, a consequence of constant-field scaling, means decreasing the maximum achievable signal level. This has a strong impact on mixed-signal product development for SOC solutions. Typical development time for new mixed-signal parts is much longer than for digital and memory parts; sheer lack of design resources thus becomes another key challenge. An ideal design process would reuse existing mixed-signal designs and adjust parameters to meet interface specifications between a given SOC and the outside world, but such reuse depends on a second type of MOSFET that does not scale its maximum operating voltage. This challenge has led to the specification in the *PIDS chapter* of a mixed-signal CMOS transistor that uses a higher analog supply voltage and stays unchanged across multiple digital technology generations. Even with such a device, voltage reduction and development time of analog circuit blocks are major obstacles to low-cost and efficient scaling of mixed-signal functions. In summary, the most daunting mixed-signal challenges are:

- *Decreasing supply voltage*, with needs including current-mode circuits, charge pumps for voltage enhancement, and thorough optimization of voltage levels in standard-cell circuits (PIDS, Design)
- *Increasing relative parametric variations*, with needs including active mismatch compensation, and tradeoffs of speed versus resolution in product definition (PIDS, FEP, Lithography, Design)
- *Increasing numbers of analog transistors per chip*, with needs including faster processing speed, more accurate compact models, and improved convergence of mixed-signal simulation tools (Modeling and Simulation, Design)
- *Increasing processing speed (carrier or clock frequencies)*, with needs including more accurate modeling of devices and interconnects, as well as test capability and package- and system-level integration (Test, Assembly and Packaging, Modeling and Simulation)
- *Increasing crosstalk* arising from SOC integration, with needs including more accurate modeling of parasitics, fully differential design for RF circuits, as well as technology measures outlined in the PIDS Chapter (PIDS, Modeling and Simulation, Design)
- *Shortage of design skills and productivity* arising from lack of training and poor automation, with needs including education and basic design tools research (Design)

EMBEDDED MEMORY DRIVER

SOC designs contain an increasing number and variety of embedded RAM, read only memory (ROM), and register file memories. Interconnect and I/O bandwidths, design productivity, and system power limits all point to a continuing trend of high levels of memory integration in microelectronic systems. Driving applications for embedded memory technology include code storage in reconfigurable applications (such as automotive), data storage in smart or memory cards, and the high memory content and high performance logic found in gaming or mass storage systems.

The balance between logic and memory content reflects overall system cost, power and I/O constraints, hardware-software organization, and overall system and memory hierarchy. With respect to cost, the device performance and added mask levels of monolithic logic-memory integration must be balanced against chip-laminate-chip or other system-in-package integration alternatives. Levels of logic-memory integration will also reflect tradeoffs in hardware-software partitioning (for example, software is more flexible, but must be booted and consumes more area) as well as code-data balance (software must be available to fill code memory, and both non-volatility and applications must be present for data memory). I/O pin count and signaling speeds determine how system organization trades off bandwidth versus storage, such as 1) memory access can be made faster at the cost of peripheral overhead by organizing memory in higher or lower bank groups; and 2) access speed also depends on how pin count and circuit complexity are balanced between high-speed low pin count connections or higher pin count lower speed connections.

Memory hierarchy is crucial in matching processor speed requirements to memory access capabilities. This fact is well known in the traditional processor architecture domain and has led to the introduction of several layers of hardware-controlled caches between "main" memory and foreground memory (e.g. register files) in the processor core. At each layer, typically one physical cache memory is present. However, the choice of hierarchy also has strong implications for power. Conventional architectures increase performance largely at the cost of energy-inefficient control overheads, for example, prediction/history mechanisms and extra buffers that are included around highly associative caches. From the system point of view, the embedded multimedia and communication applications that are dominant on portable devices can profit more from software-controlled and distributed memory hierarchies. Different layers of the memory hierarchy also require highly different access modes and internal partitionings. The use of page/burst/interleaving modes and the physical partitioning in banks, subarrays, divided word/bitlines must in general be optimized per layer. Increasingly dominant leakage power constraints also lead to more heterogeneous memory hierarchies.

Scaling presents a number of challenges to embedded memory fabrics. At the circuit level, amplifier sense margins for SRAM, and decreased I_{on} drive currents for DRAM, are two clear challenges. Smaller feature sizes imply greater impact of variability, e.g., with fewer dopants per device. With larger numbers of devices integrated into a single product, variability leads to greater parametric yield loss with respect to both noise margins and leakage power (there is an exponential dependence of leakage current on V_{th}). Future circuit topologies and design methodologies will need to address these issues. Error-tolerance is another challenge that becomes severe with process scaling and aggressive layout densities. Embedded memory soft-error rate (SER) increases with diminishing feature sizes, and affects both embedded static random-access memory (SRAM) and embedded DRAM, as discussed in the *Design chapter*. Moving bits in non-volatile memory may also suffer upsets. Particularly for highly reliable applications such as in the automotive sector, error correction is a requirement going forward, and will entail tradeoffs of yield and reliability against access time, power, and process integration. Finally, cost-effective manufacturing test and built-in self-test, for both large and heterogeneous memory arrays, is a critical requirement in the SOC context.

Since memory cell size and performance due to its high multiplication rate has very direct impact on cost and performance the amount of engineering work spend for optimization is much higher compared to all other basic circuits discussed here. Tables 11a and 11b give technology requirements for the three currently dominant types of embedded memory: CMOS embedded SRAM, embedded non-volatile memory (NVM), and embedded DRAM. Those parameters arise from the balance of circuit design consideration and technology boundary conditions given by the logic requirements tables in the *PIDS chapter*. Aggressive scaling of CMOS SRAM continues due to high-performance and low-power drivers, which require scaling of read cycle time by 0.7× per generation. Voltage scaling involves multiple considerations, such as the relationship between retention time and read operating voltage, or the impact of supply and threshold voltage scaling on pMOS device requirements starting at 45 nm. More nascent ferroelectric RAM, magnetoresistive RAM, and phase-change memory technologies are discussed in the *Emerging Research Devices chapter*.

Table 11a Embedded Memory Requirements—Near-term

Year of Production	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*
DRAM ½ Pitch (nm)	*80*	*70*	*65*	*55*	*50*	*45*
CMOS SRAM High-performance, low standby power (HP/LSTP) *DRAM ½ pitch (nm), Feature Size – F*	*90*	*90*	*65*	*65*	*65*	*45*
6T bit cell size (F^2) [1]	**$140F^2$**	**$140F^2$**	**$140F^2$**	**$140F^2$**	**$140F^2$**	**$140F^2$**
Array efficiency [2]	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**
Process overhead versus standard CMOS – number of added mask layers [3]	**1**	**2**	**2**	**2**	**2**	**2**
Operating voltage – V_{dd} (V) [4]	**1.1/1.2**	**1.1/1.2**	**1.1**	**1/1.1**	**1/1.1**	**1**
Static power dissipation (mW/Cell) [5]	**1.5E-4/6E-7**	**1.5E-4/6E-7**	**3E-4/1E-6**	**3E-4/1E-6**	**3E-4/1E-6**	**5E-4/1.2E-6**
Dynamic power consumption per cell – (mW/MHz) [6]	**7E-7/8.5E-7**	**6E-7/8E-7**	**4.5E-7/7E-7**	**4E-7/6.5E-7**	**4E-7/6E-7**	**3E-7/5E-7**
Read cycle time (ns) [7]	**0.4/2**	**0.4/2**	**0.3/1.5**	**0.3/1.5**	**0.3/1.5**	**0.2/1.2**
Write cycle time (ns) [7]	**0.4/2**	**0.4/2**	**0.3/1.5**	**0.3/1.5**	**0.3/1.5**	**0.2/1.2**
Percentage of MBU on total SER	**8%**	**8%**	**16%**	**16%**	**16%**	**32%**
Soft error rate (FIT/Mb) [8]	**1100**	**1100**	**1150**	**1150**	**1150**	**1200**
Embedded Non-Volatile Memory (code/data), DRAM ½ pitch (nm)	*130*	*130*	*90*	*90*	*90*	*65*
Cell size (F^2) – NOR FLOTOX /NAND FLOTOX [9]	**$10F^2/5F^2$**	**$10F^2/5F^2$**	**$10F^2/5F^2$**	**$10F^2/5F^2$**	**$10F^2/5F^2$**	**$10F^2/5F^2$**
Array efficiency – NOR FLOTOX/NAND FLOTOX [10]	**0.6/0.8**	**0.6/0.8**	**0.6/0.8**	**0.6/0.8**	**0.6/0.8**	**0.6/0.8**
Process overhead versus standard CMOS – number of added mask layers [11]	**6–8**	**6–8**	**6–8**	**6–8**	**6–8**	**6–8**
Read operating voltage (V)	**2.5V**	**2.5V**	**2V**	**2V**	**2V**	**1.8V**
Write (program/erase) on chip maximum voltage (V) – NOR/NAND [12]	**12V/15V**	**12V/15V**	**12V/15V**	**12V/15V**	**12V/15V**	**12V/15V**
Static power dissipation (mW/Cell) [5]	**1.E-06**	**1.E-06**	**1.E-06**	**1.E-06**	**1.E-06**	**1.E-06**
Dynamic power consumption per cell – (mW/MHz) [6]	**0.8E-08**	**0.8E-08**	**0.6E-08**	**0.6E-08**	**0.6E-08**	**0.6E-08**
Read cycle time (ns) NOR FLOTOX /NAND FLOTOX [7]	**14/70**	**14/70**	**10/50**	**10/50**	**10/50**	**7/35**
Program time per cell (μs) NOR FLOTOX /NAND FLOTOX [13]	**1.0/1000.0**	**1.0/1000.0**	**1.0/1000.0**	**1.0/1000.0**	**1.0/1000.0**	**1.0/1000.0**
Erase time per cell (ms) NOR FLOTOX /NAND FLOTOX [13]	**10.0/0.1**	**10.0/0.1**	**10.0/0.1**	**10.0/0.1**	**10.0/0.1**	**10.0/0.1**
Data retention requirement (years) [13]	**10**	**10**	**10**	**10**	**10**	**10**
Endurance requirement [13]	**100000**	**100000**	**100000**	**100000**	**100000**	**100000**
Embedded DRAM, ½ pitch (nm)	*130*	*90*	*90*	*90*	*65*	*65*
1T1C bit cell size (F^2) [14]	**12–30**	**12–30**	**12–30**	**12–30**	**12–30**	**12–30**
Array efficiency [2]	**0.6**	**0.6**	**0.6**	**0.6**	**0.6**	**0.6**
Process overhead versus standard CMOS – number of added mask layers [3]	**3–5**	**3–5**	**3–5**	**3–5**	**3–5**	**3–5**
Read operating voltage (V)	**2.5**	**2**	**2**	**2**	**1.8**	**1.7**
Static power dissipation (mW/Cell) [5]	**1E-11**	**1E-11**	**1E-11**	**1E-11**	**1E-11**	**1E-11**
Dynamic power consumption per cell – (mW/MHz) [6]	**1.E-07**	**1.E-07**	**1.E-07**	**1.E-07**	**1.E-07**	**1.5E-07**
DRAM retention time (ms) [13]	**64**	**64**	**64**	**64**	**64**	**64**
Read/Write cycle time (ns) [7]	**1**	**0.7**	**0.7**	**0.7**	**0.5**	**0.4**
Soft error rate (FIT/Mb) [8]	**60**	**60**	**60**	**60**	**60**	**60**

FIT—failures in time FLOTOX—floating gate tunnel oxide MBU—multiple bit upsets NAND—not an "AND" logic operation NOR—not an "OR" logic operation

*Table 11b Embedded Memory Requirements—Long-term**

Year of Production	*2012*	*2015*	*2018*
DRAM ½ Pitch (nm)	*36*	*25*	*18*
CMOS SRAM High-performance, low standby power (HP/LSTP) *DRAM 1/2 pitch (nm), Feature Size – F*	*35*	*25*	*18*
6T bit cell size (F^2) [1]	**$140F^2$**	**$140F^2$**	**$140F^2$**
Array efficiency [2]	**0.7**	**0.7**	**0.7**
Process overhead versus standard CMOS – number of mask adders [3]	**2**	**2**	**2**
Operating voltage – V_{dd} (V)	**0.9/1**	**0.8/0.9**	**0.7/0.8**
Static power dissipation (mW/Cell) [5]	**1E-3/1.5E-6**	**2E-3/2E-6**	**3E-3/2.5E-6**
Dynamic power consumption per cell – (mW/MHz) [6]	**2.5E-7/4.5E-7**	**2E-7/4E-7**	**1.5E-7/3E-7**
Read cycle time (ns) [7]	**0.15/0.8**	**0.1/0.5**	**0.07/0.3**
Write cycle time (ns) [7]	**0.15/0.8**	**0.1/0.5**	**0.07/0.3**
Percentage of MBU on total SERs	**64%**	**100%**	**100%**
Soft error rate (FIT/Mb) [8]	**1250**	**1300**	**1350**
Embedded Non-Volatile Memory (code/data), DRAM ½ pitch (nm)	*45*	*35*	*25*
Cell size (F^2) – NOR FLOTOX/NAND FLOTOX [9]	**$10F^2/5F^2$**	**$10F^2/5F^2$**	**$10F^2/5F^2$**
Array efficiency – NOR FLOTOX/NAND FLOTOX [10]	**0.6/0.8**	**0.6/0.8**	**0.6/0.8**
Process overhead versus standard CMOS – number of mask adders [3]	**6–8**	**6–8**	**6–8**
Read operating voltage (V) [4]	**1.5V**	**1.3V**	**1.2V**
WRITE (program/erase) on chip maximum voltage (V) – NOR/NAND [4]	**12V/15V**	**12V/15V**	**12V/15V**
Static power dissipation (mW/Cell) [5]	**1.E-06**	**1.E-06**	**1.E-06**
Dynamic power consumption per cell – (mW/MHz) [6]	**0.4E-8**	**0.35E-8**	**0.3E-8**
Read cycle time (ns)	**5/25**	**3.5/18**	**2.5/12**
Program time per cell (µs) [13]	**1.0/1000.0**	**1.0/1000.0**	**1.0/1000.0**
Erase time per cell (ms) [13]	**10.0/0.1**	**10.0/0.1**	**10.0/0.1**
Data retention requirement (years) [13]	**10**	**10**	**10**
Endurance requirement [13]	**100000**	**100000**	**100000**
Embedded DRAM, ½ pitch (nm)	*45*	*35*	*25*
1T1C bit cell size (F^2) [14]	**12–30**	**12–30**	**12–30**
Array efficiency [2]	**0.6**	**0.6**	**0.6**
Process overhead versus standard CMOS – number of mask adders [3]	**3–6**	**3–6**	**3–6**
Read operating voltage (V)	**1.6**	**1.5**	**1.5**
Static power dissipation (mW/Cell) [5]	**1E-11**	**1E-11**	**1E-11**
Dynamic power consumption per cell – (mW/MHz) [6]	**1.6E-07**	**1.7E-07**	**1.7E-07**
DRAM retention time (ms) [13]	**64**	**64**	**64**
Read/Write cycle time (ns) [7]	**0.3**	**0.25**	**0.2**
Soft error rate (FIT/Mb) [8]	**60**	**60**	**60**

**Table 11b data will be annualized in 2006. For the 2005 ITRS, long-term years are 2014–2020.*

Definitions of Terms for Tables 11a and 11b:

[1] Size of the standard 6T CMOS SRAM cell as a function of minimum feature size.

[2] Typical array efficiency defined as (core area/memory instance area).

[3] Typical number of extra masks is needed over standard CMOS logic process of equivalent technology. This is typically zero, however for some high-performance or highly reliable (noise immune) SRAMs special process options are sometimes applied like additional high—V_{th} pMOS cell transistors and using higher V_{dd} for better noise margin or zero-V_{th} access transistors for fast read-out.

[4] Nominal operating voltage refers to the HP and LSTP devices in the logic device requirements table in the PIDS chapter.

[5] Static power dissipation per cell in standby mode. This is measured at I_standby x V_{dd}. (off-current and V_{dd} are taken from the HP and LSTP devices in the logic device requirements table in the PIDS Chapter.

[6] This parameter is a strong function of array architecture. However, a parameter for technology can be determined per cell level. Assume full V_{dd} swing on the Wordline (WL) and 0.8 V_{dd} swing on the Bitline (BL). Determine the WL capacitance per cell (CWL) and BL capacitance per cell (CBL). Then: dyn. power cons. per MHz per cell = V_{dd} × CWL (per cell) × (V_{dd}) + V_{dd} x CBL (per cell) x (V_{dd}) ×10^6.

[7] Read cycle time is the typical time it takes to complete a READ operation from an address. Depends on memory size and architecture. Write cycle time is the typical time it takes to complete a WRITE operation to an ADDR. Depends on memory size and architecture.

[8] A FIT is a failure in 1 billion hours. This data is presented as FIT per megabit.

[9] Size of the standard 1T FLOTOX cell/size of the standard 2T select gate (SG) cell/size of the standard NAND cell. Cell size is somewhat enhanced compared to stand-alone NVM due to integration issues.

[10] Array efficiency of the standard stacked gate NOR architecture/standard split gate NOR architecture/standard NAND architecture. Data refer to PIDS table the NVM device requirements table in the PIDS chapter.

[11] Extra process steps needed to realize the technology as compared to standard CMOS process.

[12] Maximum voltage required for operation, typically used in WRITE operation. Data refer to the NVM device requirements table in the PIDS chapter.

[13] Program time per cell is typically the time needed to program data to a cell. Erase time per cell is typically the time needed to erase a cell. Data retention requirement is the duration for which the data must remain non-volatile even under worst-case conditions. Endurance requirement specifies the number of times the cell can be programmed and erased.

[14] Size of the standard cell for embedded trench DRAM cell. Data refers to the DRAM requirements table in the PIDS chapter.