

OVERALL ROADMAP TECHNOLOGY CHARACTERISTICS

BACKGROUND

The Overall Roadmap Technology Characteristics (ORTC) tables are created early in the Roadmap process and are used as the basis for initiating the activities of the International Technology Working Groups (ITWGs) in producing their detailed chapters. These tables are also used throughout the renewal effort of the Roadmap as a means of providing synchronization among the TWGs by highlighting inconsistencies between the specific tables. The process to revise the tables include increasing levels of cross-TWG and international coordination and consensus building to develop underlying models of trends and to reach agreement on target metrics. As a result, the ORTC tables went through several iterations and reviews.

The metric values of the ORTC tables can be found throughout the Roadmap in greater detail in each Technology Working Group chapter. The information in this section is intended to highlight the current rapid pace of advancement in semiconductor technology. It represents a completion of the revision update and renewal work that began in 2000. Additionally, an ORTC Glossary is provided as an appendix.

OVERVIEW OF 2001 REVISIONS

DEFINITIONS

As noted above, the Overall Roadmap Technology Characteristics tables provide a consolidated summary of the key technology metrics. Please note that the year header on the tables may refer to different points in the development/life cycle of integrated circuits (ICs), depending on the individual line item metric. However, unless otherwise specified for a particular line item, the default year header still refers (as in previous Roadmaps) to the year when product shipment first exceeds 10,000 units per month of ICs from a manufacturing site using “production tooling.” Additional clarification was provided this year by the *ITRS* executive International Roadmap Committee (IRC), requiring a second company to start production within three months. To satisfy this definition, ASIC production may represent the cumulative volume of many individual product line items processed through the facility. Please see the Glossary section for additional details on “Technology Node” and “Production” timing definitions.

Furthermore, new IRC guidelines clarified the definition of a technology node as the achievement of significant advancement in the process technology. To be explicit, a technology node was defined as the achievement of an approximate 0.7× reduction per node (0.5× per two nodes) (Figure 5). The period of time in which a new technology node is reached is called a “technology-node cycle” (Figure 6). It is acknowledged that continuous improvement occurs between the technology nodes, and this is reflected by including data between nodes in the annual columns of the “Near-term years” tables. The “Long-term years” table columns are three-year increments of the 2001 *ITRS* timeframe.

ROADMAP TIMELINE

The 2001 edition of the Roadmap maintains a 15-year projection, from 2001 as a reference year and through 2016. However, the timing of future technology nodes has changed from the 1999 edition.

By international consensus, the 130 nm node was pulled in an additional year, continuing the historical precedent for two-year technology-node cycles since 350 nm/1995 (250 nm node in 1997, 180 nm node in 1999). Although there is the possibility of a continuation of the two-year-node cycle trend, the present consensus projects a three-year cycle for DRAM interconnect half-pitch nodes throughout the 2001–2016 Roadmap period (Figure 7).

DRAM interconnect (metal or polysilicon—refer to the Glossary for additional detail) half-pitch will continue to be used as the most representative feature of leading-edge semiconductor manufacturing technology for defining the achievement of a technology node. However, additional data analysis indicates an aggressive trend for the lagging MPU and ASIC interconnect half-pitches to catch up to DRAM half-pitch by 2004 (Figure 7).

The 2001 *ITRS* includes a correction of the past “rounding” convention for the technology node labels. The actual mathematical trend reduces the nodes by 50% every other node, resulting in an actual versus rounded node number targets, starting from 350 nm in 1995 as follows:

Table C Rounded versus Actual Trend Numbers

<i>YEAR OF PRODUCTION</i>	<i>1995</i>	<i>1997</i>	<i>1999</i>	<i>2001</i>	<i>2004</i>	<i>2007</i>	<i>2010</i>	<i>2013</i>	<i>2016</i>
<i>Actual Trend Numbers (nm)</i>	350	247.5	175	123.7	87.5	61.9	43.8	30.9	21.9
<i>ITRS Rounded Node Numbers (nm)</i>	350	250	180	130	90	65	45	32	22

Note the new “rounding” corrections become more critical as the industry moves into the double-digit technology nodes. Some regions, for past publication consistency, will continue to track the previous technology nodes beginning with 100 nm/2003, resulting in milestones that are placed one year earlier than the present 2001 roadmap convention (70 nm/2006; 50 nm/2009; 35nm/2012; 25 nm/2015).

The *printed* MPU gate length received a major correction to more an aggressive starting point in 2001. In addition, a new *physical* gate length is being tracked that further reduces the bottom gate length dimension of a fully-processed transistor. Both the printed and physical gate length trends are forecast to continue scaling by about 70% per two-year cycle through the 32nm physical MPU gate length in 2005, but are expected to return to a three-year cycle trend thereafter, consistent with the present DRAM half-pitch trend forecast (Figure 8).

The ORTC metrics, which guide the Roadmap, are often used by semiconductor companies as a set of targets that need to be achieved ahead of schedule to achieve industry leadership. Thus, the highly competitive environment of the semiconductor industry quickly tends to make obsolete many portions of the ORTC metrics and, consequently, the Roadmap. Hopefully, our annual update process will provide sufficiently close tracking of the evolving international consensus on technology directions to maintain the usefulness of the *ITRS* to the industry.

For example, there is some anticipation that DRAM half-pitch nodes could undergo an additional one-year pull-in. This possibility will be re-evaluated during the year 2002 *ITRS* Update process, along with the possibility of using a two-year node cycle as a longer-term trend. To reflect the variety of cycles and to allow for closer monitoring of future roadmap shifts, it was agreed to continue the practice of publishing annual technology requirements from 2001 through 2007, called the “Near-Term Years,” and at three-year (node) intervals thereafter, called the “Long-Term Years” (2010, 2013, 2016).

MOS Transistor *Scaling* (1974 to present)

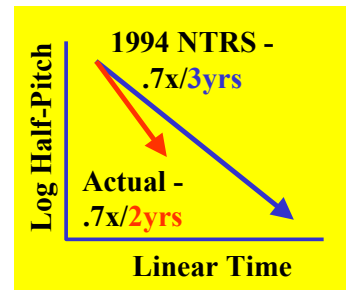
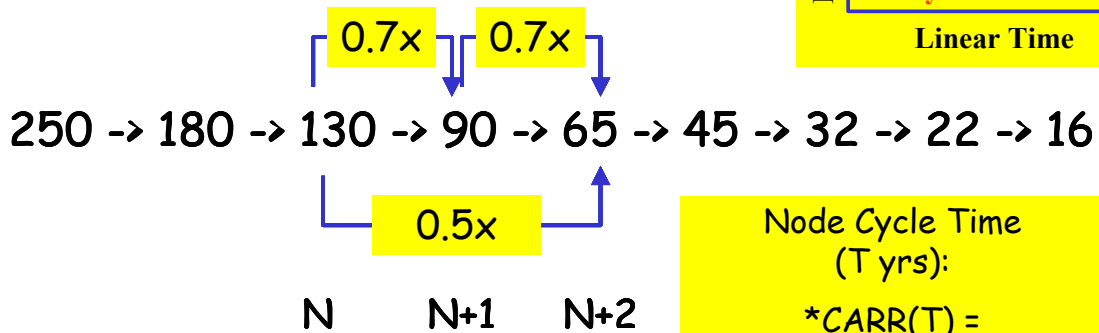
$$S=0.7$$

[0.5x per 2 nodes]



Figure 5 MOS Transistor Scaling—1974 to present

Scaling Calculator + Node Cycle Time:



* CARR(T) = Compound Annual Reduction Rate (@ cycle time period, T)

Node Cycle Time (T yrs):
*CARR(T) = $[(0.5)^{(1/2T \text{ yrs})}] - 1$
CARR(3 yrs) = -10.9%
CARR(2 yrs) = -15.9%

Figure 6 Scaling Calculator

ITRS Roadmap Acceleration Continues...Half Pitch

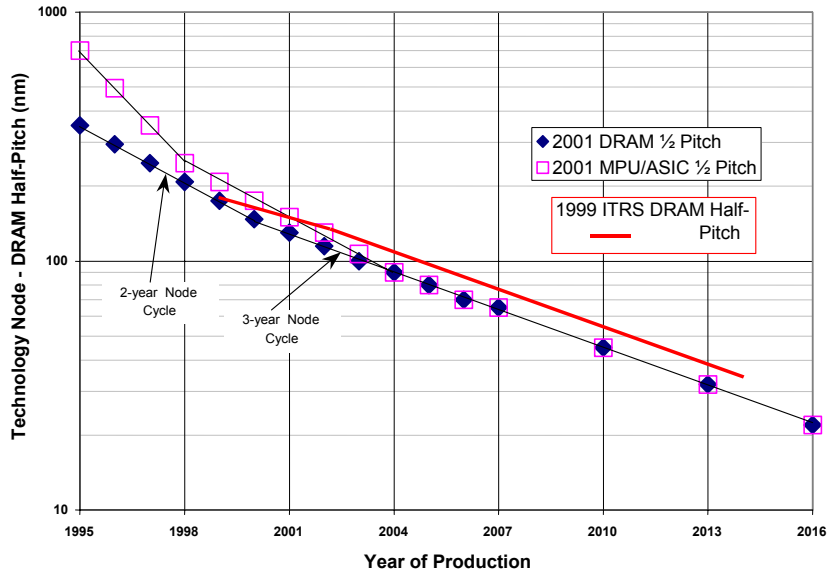


Figure 7 ITRS Roadmap Acceleration Continues—Half Pitch Trends

ITRS Roadmap Acceleration Continues...Gate Length

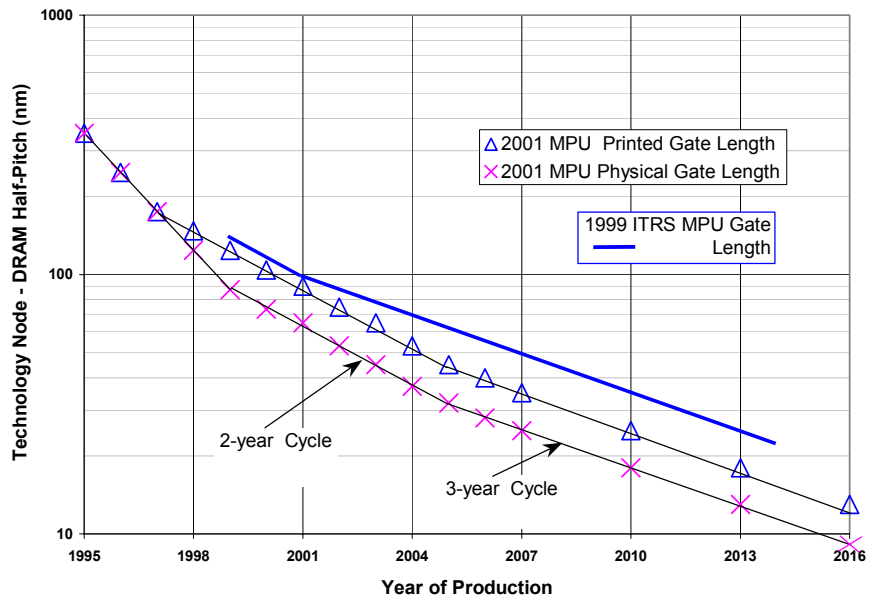


Figure 8 ITRS Roadmap Acceleration Continues—Gate Length Trends

PRODUCT GENERATIONS AND CHIP-SIZE MODEL

In this section, we will discuss “product generations” and their relationship to the technology nodes, since, in the past, these terms have often been used interchangeably. However, the historically simple picture of a new DRAM product generation every three years (at 4× the previous density and based on an essentially new set of technology features) has become obsolete as a way to define technology nodes. For this 2001 edition, the “technology node” is still linked to an anticipated DRAM feature size (minimum metal or polysilicon half-pitch). However, implications of this connection are diminishing as the product evolution/shrink path becomes more complex.

Historically, DRAM products have been recognized as the technology drivers for the entire semiconductor industry. Prior to the early 1990s, logic (as exemplified by MPU) technology was developed at a slower pace than DRAM technology. During the last few years, the development rate of new technologies used to manufacture microprocessors has accelerated. Microprocessor products are closing the half-pitch technology gap with DRAM, and are now driving the most leading-edge lithography tools and processes—especially for the capability to process the isolated-line feature of the printed and physical gate length. With this 2001 Roadmap it is recognized that DRAM and microprocessor products share the technology leadership role.

However, several fundamental differences exist between the two families of products. Due to strong commodity market economic pressure to reduce cost and increase fab output productivity, DRAM product emphasizes the minimization of the chip size. Therefore development of DRAM technology focuses mainly on minimization of the area occupied by the memory cell. However, this pressure to minimize cell size is in conflict with the requirement to maximize the capacitance of the cell for charge storage performance, which puts pressure on memory cell designers to find creative ways through design and materials to meet minimum capacitance requirements while reducing cell size. In addition, to closely pack the highest number of DRAM cells in the smallest area requires minimization of cell pitch.

Microprocessors have also come under strong market pressure to reduce costs while maximizing performance. Performance is enabled primarily by the length of the transistor gate and by the number of interconnect layers. The 2001 *ITRS* teams have reached consensus on models for the required functionality, chip size, cell area, and density for the ORTC tables. Additional line items were added to communicate the model consensus, and the underlying model assumptions are included in notations. Table 1a and 1b summarize the near and long technology node metrics. As agreed, the key *ITRS* technology node identifier would continue to be the DRAM half-pitch, but also included are the aggressive MPU gate-length performance-driven feature sizes. For completeness, the MPU/ASIC product metal half-pitch are also tracked and that will trail slightly behind or equal to the DRAM half-pitch. The ASIC/low power gate lengths are also included, and lag behind the leading-edge MPU in order to maximize standby and operating current drain. See the Glossary section for additional detail on the definition of the half-pitch and gate-length features. For each product generation, both the leading-edge (“at introduction”) and the high-volume (“at production”) DRAM products are indicated

It should be noted that the long-term average annualized reduction rate in feature size is projected to continue at approximately 11%/year (~30% reduction/three years), even though this rate accelerated to approximately 16%/year (~30% reduction/two years) in the time interval 1995–2001 (refer to Figure 5). As mentioned above, the overall schedule for introduction of a new product generation has been accelerated by one additional year.

Table 1a Product Generations and Chip Size Model Technology Nodes—Near-term Years

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm) ††	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
ASIC/Low Power Printed Gate Length (nm) ††	130	107	90	75	65	53	45
ASIC/Low Power Physical Gate Length (nm)	90	75	65	53	45	37	32

Table 1b Product Generations and Chip Size Model Technology Nodes—Long-term years

Year of Production	2010	2013	2016
DRAM ½ Pitch (nm)	45	32	22
MPU/ASIC ½ Pitch (nm)	45	32	22
MPU Printed Gate Length (nm) ††	25	18	13
MPU Physical Gate Length (nm)	18	13	9
ASIC/Low Power Printed Gate Length (nm) ††	32	22	16
ASIC/Low Power Physical Gate Length (nm)	22	16	11

Notes for Tables 1a and 1b:

†† MPU and ASIC gate-length (in resist) node targets refer to the most aggressive requirements, as printed in photoresist (which was by definition also “as etched in polysilicon,” in the 1999 ITRS).

However, during the 2000/2001 ITRS development, trends were identified, in which the MPU and ASIC “Physical” gate lengths may be reduced from the “as-printed” dimension. These “Physical” gate-length targets are driven by the need for maximum speed performance in logic Microprocessor (MPU) products, and are included in the Front End Processes (FEP), Process Integration, Devices, and Structures (PIDs), and Design ITWG Tables as needs that drive device design and process technology requirements.

Table 1c DRAM Production Product Generations and Chip Size Model—Near-term Years

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
Cell area factor [a]	8	8	6	6	6	6	6
Cell area [Ca = af ²] (μm ²)	0.130	0.103	0.061	0.049	0.039	0.031	0.024
Cell array area at production (% of chip size) §	54.8%	55.3%	55.7%	56.1%	56.4%	56.7%	57.0%
Generation at production §	512M	512M	1G	1G	2G	2G	4G
Functions per chip (Gbits)	0.54	0.54	1.07	1.07	2.15	2.15	4.29
Chip size at production (mm ²)§	127	100	118	93	147	116	183
Gbits/cm ² at production §	0.42	0.54	0.91	1.15	1.46	1.85	2.35

Table 1d DRAM Production Product Generations and Chip Size Model—Long-term Years

Year of Production	2010	2013	2016
DRAM ½ Pitch (nm)	45	32	22
MPU/ASIC ½ Pitch (nm)	45	32	22
MPU Printed Gate Length (nm)	25	18	13
MPU Physical Gate Length (nm)	18	13	9
Cell area factor [a]	6	4	4
Cell area [Ca = af ²] (μm ²)	0.012	0.004	0.002
Cell array area at production (% of chip size) §	57.7%	58.1%	58.4%
Generation at production §	8G	32G	64G
Functions per chip (Gbits)	8.59	34.36	68.72
Chip size at production (mm ²)§	181	239	238
Gbits/cm ² at production §	4.75	14.35	28.85

Notes for Tables 1c and 1d:

§ DRAM Model—Cell Factor (design/process improvement) targets are as follows:

1999–2002/8×; 2003–2010/6×; 2011–2016/4×

DRAM product generations are usually increased by 4×bits/chip every four years with interim 2×bits/chip generations, except:

- at the Introduction phase, after the 8Gbit interim generation, the introduction rate is 4×/five years (2×/two–three years); and
- at the Production phase, after the interim 32Gbit generation, the introduction rate is 4×/five years (2×/two–three years).

InTER-generation chip size growth rate varies to maintain one chip per 572mm² field at Introduction and two chips per 572mm² field at Production. The more aggressive “best case opportunity” technology node trends allow the Production-phase products to remain at 2×bits/chip every two years and still fit within the target of two DRAM chips per 572mm² field size, through the 32Gbit interim generation. The InTRA-generation chip size shrink model is 0.5× every technology node in-between cell factor reductions.

Refer to the Glossary for definitions of Introduction, Production, InTERgeneration, and InTRAgeneration terms.

Table 1e DRAM Introduction Product Generations and Chip Size Model—Near-term Years

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
Cell area factor [a]	8	8	6	6	6	6	6
Cell area [Ca = af ²] (μm ²)	0.130	0.103	0.061	0.049	0.039	0.031	0.024
Cell array area at introduction (% of chip size) §	71.3%	71.8%	72.2%	72.6%	72.9%	73.2%	73.5%
Generation at introduction §	2G	2G	4G	4G	8G	8G	16G
Functions per chip (Gbits)	2.15	2.15	4.29	4.29	8.59	8.59	17.18
Chip size at introduction (mm ²) §	390	308	364	287	454	359	568
Gbits/cm ² at introduction §	0.55	0.70	1.18	1.49	1.89	2.39	3.03

Table 1f DRAM Introduction Product Generations and Chip Size Model—Long-term Years

Year of Production	2010	2013	2016
DRAM ½ Pitch (nm)	45	32	22
MPU/ASIC ½ Pitch (nm)	45	32	22
MPU Printed Gate Length (nm)	25	18	13
MPU Physical Gate Length (nm)	18	13	9
Cell area factor [a]	6	4	4
Cell area [Ca = af ²] (μm ²)	0.012	0.004	0.002
Cell array area at introduction (% of chip size) §	74.2%	74.6%	74.9%
Generation at introduction §	32G	64G	64G
Functions per chip (Gbits)	34.36	68.72	68.72
Chip size at introduction (mm ²) §	563	373	186
Gbits/cm ² at introduction §	6.10	18.42	37.00

Notes for Tables 1e and 1f:

§ DRAM Model—Cell Factor (design/process improvement) targets are as follows:

1999–2002/8×; 2003–2010/6×; 2011–2016/4×

DRAM product generations are usually increased by 4×bits/chip every four years with interim 2×bits/chip generations, except:

5. at the Introduction phase, after the 8Gbit interim generation, the introduction rate is 4×/five years (2×/two–three years); and
6. at the Production phase, after the interim 32Gbit generation, the introduction rate is 4×/five years (2×/two–three years).

InTER-generation chip size growth rate varies to maintain one chip per 572mm² field at Introduction and two chips per 572mm² field at Production. The more aggressive “best case opportunity” technology node trends allow the Production-phase products to remain at 2×bits/chip every two years and still fit within the target of two DRAM chips per 572mm² field size, through the 32Gbit interim generation. The InTRA-generation chip size shrink model is 0.5× every technology node in-between cell factor reductions.

Refer to the Glossary for definitions of Introduction, Production, InTERgeneration, and InTRAgeneration terms.

Table 1g MPU (High-volume Microprocessor) Cost-Performance Product Generations and Chip Size Model—Near-term Years

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
SRAM Cell (6-transistor) Area factor ++	126.1	123.0	120.3	117.8	115.6	113.7	111.9
Logic Gate (4-transistor) Area factor ++	320.0	320.0	320.0	320.0	320.0	320.0	320.0
SRAM Cell (6-transistor) Area efficiency ++	0.63	0.63	0.63	0.63	0.63	0.63	0.63
Logic Gate (4-transistor) Area efficiency ++	0.50	0.50	0.50	0.50	0.50	0.50	0.50
SRAM Cell (6-transistor) Area w/overhead ++	3.3	2.5	2.0	1.5	1.2	0.93	0.73
Logic Gate (4-transistor) Area w/overhead ++	10.4	8.2	6.5	5.2	4.1	3.3	2.6
Transistor density SRAM (Mtransistors/cm ²)	184	237	305	393	504	646	827
Transistor density logic (Mtransistors/cm ²)	38.6	48.6	61.2	77.2	97.2	122.5	154.3
Generation at introduction *	p04c	—	—	p07c	—	—	p10c
Functions per chip at introduction (million transistors [Mtransistors])	193	243	307	386	487	614	773
Chip size at introduction (mm ²) ‡	280	280	280	280	280	280	280
Cost performance MPU (Mtransistors/cm ² at introduction) (including on-chip SRAM) ‡	69	87	110	138	174	219	276
Generation at production *	p01c	—	—	p04c	—	—	p07c
Functions per chip at production (million transistors [Mtransistors])	97	122	153	193	243	307	386
Chip size at production (mm ²) §§	140	140	140	140	140	140	140
Cost performance MPU (Mtransistors/cm ² at production, including on-chip SRAM) ‡	69.0	87.0	109.6	138.0	173.9	219.1	276.1

Notes for Tables 1g and 1h:

++ The MPU area factors are analogous to the "cell area factor" for DRAMs. The reduction of area factors has been achieved historically through a combination of many factors, for example—use of additional interconnect levels, self-alignment techniques, and more efficient circuit layout. However, recent data has indicated that the improvement (reduction) of the area factors is slowing, and is virtually flat for the logic gate area factor.

* p is processor, numerals reflect year of production; c indicates cost-performance product. Examples—the cost-performance processor, p01c, was introduced in 1999, but not ramped into volume production until 2001; similarly, the p04c, is introduced in 2001, but is targeted for volume production in 2004.

‡ MPU Cost-performance Model—Cost-performance MPU includes Level 2 (L2) on-chip SRAM (512Kbyte/1999), and the combination of both SRAM and logic transistor functionality doubles every technology node cycle.

§§ MPU Chip Size Model—Both the cost-performance and high-performance MPUs InTER-generation chip size growth rates are targeted to be flat through 2016, made possible by doubling the on-chip functionality every technology node cycle. The InTRA-generation chip size shrink model is 0.5× every two-year technology node through 2001, then 0.5× every three-year technology node after 2001.

Table 1h MPU (High-volume Microprocessor) Cost-Performance Product Generations and Chip Size Model—Long-term Years

<i>Year of Production</i>	<i>2010</i>	<i>2013</i>	<i>2016</i>
<i>DRAM ½ Pitch (nm)</i>	45	32	22
<i>MPU/ASIC ½ Pitch (nm)</i>	45	32	22
<i>MPU Printed Gate Length (nm)</i>	25	18	13
<i>MPU Physical Gate Length (nm)</i>	18	13	9
<i>SRAM Cell (6-transistor) Area factor ++</i>	107.8	106.7	105.7
<i>Logic Gate (4-transistor) Area factor ++</i>	320.0	320.0	320.0
<i>SRAM Cell (6-transistor) Area efficiency ++</i>	0.63	0.63	0.63
<i>Logic Gate (4-transistor) Area efficiency ++</i>	0.50	0.50	0.50
<i>SRAM Cell (6-transistor) Area w/overhead ++</i>	0.22	0.17	0.13
<i>Logic Gate (4-transistor) Area w/overhead ++</i>	0.82	0.65	0.51
<i>Transistor density SRAM (Mtransistors/cm²)</i>	1718	3532	7208
<i>Transistor density logic (Mtransistors/cm²)</i>	309	617	1235
<i>Generation at introduction *</i>	p13c	p16c	p19c
<i>Functions per chip at introduction (million transistors [Mtransistors])</i>	1546	3092	6184
<i>Chip size at introduction (mm²) ‡</i>	280	280	280
<i>Cost performance MPU (Mtransistors/cm² at introduction) (including on-chip SRAM) ‡</i>	552	1104	2209
<i>Generation at production *</i>	p10c	p13c	p16c
<i>Functions per chip at production (million transistors [Mtransistors])</i>	773	1546	3092
<i>Chip size at production (mm²) §§</i>	140	140	140
<i>Cost performance MPU (Mtransistors/cm² at production, including on-chip SRAM) ‡</i>	552	1104	2209

Table 1i High-Performance MPU and ASIC Product Generations and Chip Size Model—Near-term Years

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
Logic (Low-volume Microprocessor) High-performance ‡							
Generation at production **	p01h	—	p03h	—	p05h	—	p07h
Functions per chip (million transistors)	276	348	439	553	697	878	1106
Chip size at production (mm ²) §§	310	310	310	310	310	310	310
High-performance MPU Mtransistors/cm ² at production (including on-chip SRAM) ‡	89	112	142	178	225	283	357
ASIC							
ASIC usable Mtransistors/cm ² (auto layout)	89	112	142	178	225	283	357
ASIC max chip size at production (mm ²) (maximum lithographic field size)	800	800	572	572	572	572	572
ASIC maximum functions per chip at production (Mtransistors/chip) (fit in maximum lithographic field size)	714	899	810	1020	1286	1620	2041

Table 1j High-Performance MPU and ASIC Product Generations and Chip Size Model—Long-term Years

Year of Production	2010	2013	2016
DRAM ½ Pitch (nm)	45	32	22
MPU/ASIC ½ Pitch (nm)	45	32	22
MPU Printed Gate Length (nm)	25	18	13
MPU Physical Gate Length (nm)	18	13	9
Logic (Low-volume Microprocessor) High-performance ‡			
Generation at production **	—	p13h	—
Functions per chip (million transistors)	2212	4424	8848
Chip size at production (mm ²) §§	310	310	310
High-performance MPU Mtransistors/cm ² at production (including on-chip SRAM) ‡	714	1427	2854
ASIC			
ASIC usable Mtransistors/cm ² (auto layout)	714	1427	2854
ASIC maximum chip size at production (mm ²) (maximum lithographic field size)	572	572	572
ASIC maximum functions per chip at ramp (Mtransistors/chip) (fit in maximum lithographic field size)	4081	8163	16326

Notes for Tables 1i and 1j:

** *p* is processor, numerals reflect year of production; *h* indicates high-performance product. Examples—the high-performance processor, *p99h*, was ramped into volume production in 1999; similarly, the *p01h*, is introduced in 2001.

‡ MPU High-performance Model—High-performance MPU includes large L2 and L3 on-chip SRAM (2MByte/1999) plus a larger logic core (P99h core = 25M transistor (Mtransistors) both SRAM and Logic functionality doubles every technology node cycle.

§§ MPU Chip Size Model—Both the cost-performance and high-performance MPUs InTER-generation chip size growth rates are targeted to be flat through 2016, made possible by doubling the on-chip functionality every technology node cycle. The InTRA-generation chip size shrink model is 0.5x every two-year technology node through 2001, then 0.5x every three-year technology node after 2001.

CHIP-SIZE, LITHOGRAPHIC-FIELD, AND WAFER-SIZE TRENDS

Despite the continuous reduction in feature size of about 30% every three years, the size of first DRAM product demonstration in technical forums such as the IEEE International Solid State Circuits Conference (ISSCC) has continued to double every six years (an increase of about 12%/year). This increase in chip area has been necessary to accommodate 59% more bits/capacitors/transistors per year in accordance with Moore's Law (historically doubling functions per chip every 1.5–2 years). However, to maintain the historical trend of reducing cost/function by ~25–30%/year, it is necessary to continuously enhance equipment productivity, increase manufacturing yields, use the largest wafer size available, and, most of all, increase the number of chips available on a wafer.

The increase in the gross number of chips available on a wafer is primarily obtained by reducing the area of the chip by means of a combination of smaller feature size (shrink/scaling) and product/process redesign (compaction). For instance, using the latest models, it is forecast that the introduction chip area of a cost-effective product generation [which doubles the inter-generation (generation-to-generation) functionality every two years] must either remain flat or grow no faster than 20% every four years. Furthermore, the area must be shrunk at an intra-generation (within a generation) annual reduction rate of 50% (the square of the $.7\times$ lithography reduction rate) during every technology node period.

Affordable DRAM products must also achieve virtually flat intra-generation chip-sizes, and they must also maintain a cell area array efficiency ratio of less than 70% of total chip area. Therefore, DRAM products require aggressive cell area factors (cell area in units of minimum-feature-size-squared). The Front End Processes International Technology Working Group has provided the cell area factors and detailed the challenges and needs for solutions to meet the aggressive cell area goals in the Front End Processes chapter. Due to the importance of tracking/coordinating these new challenges, the DRAM cell area factor, the target cell sizes, and the cell array area percentage of total chip-size line items will continue to be tracked in ORTC Tables 1c, d, e, and f. (Refer to the Glossary for additional details.)

In 2001 the Design ITWG improved the MPU chip size model to update with the latest transistor densities, large on-chip SRAM, and smaller target chip sizes. The Design ITWG has also added additional detail to the model, including transistor design improvement factors. The Design ITWG notes that design improvements occur at a slow rate in SRAM transistors and very little in logic gate transistors. Almost all the “shrink” and density improvement comes from lithography-enabled interconnect half-pitch scaling alone.

The present MPU chip size model reflects additional competitive requirements for affordability and power management by targeting flat chip size trends for both high-performance MPUs (310mm^2) and cost-performance MPUs (140mm^2). Due to the MPU two-year-cycle half-pitch “catch-up phase” through the year 2004, the MPU products may be able to maintain flat chip sizes due to lithography improvements alone. However, after 2004, the intra-generation chip size of MPUs can remain flat only by slowing the rate of on-chip transistors to double every technology node.

Due to the forecasted return to a three-year technology node cycle, the present MPU chip-size model slows the Moore's Law rate of on-chip transistors to $2\times$ every three years. In order to maintain a flat chip size target and also return to the historical doubling every two years of on-chip functionality (transistors), MPU chip and process designers must add additional design/process improvements to the fundamental lithography-based scaling trends. The new target metrics of the MPU model are summarized in Tables 1g, h, I, and j.

To improve productivity, it is necessary to increase the output of good chips at each step in the fabrication process. The ability of printing multiple chips in a single exposure is a key productivity driver and is determined by the field size of the lithographic tool and the size and aspect ratio of the chips being printed on the wafer. In the past, lithography exposure field sizes doubled every other technology node to meet the demand for increasing chip sizes. The result was the achievement of very large step-and-scan fields ($25\times 32 = 800\text{mm}^2$) by 1999. However, the Lithography ITWG indicates that maintaining the large field size under continued reduction of exposure features is increasing costs dramatically. Therefore, the ITWG forecasts a requirement for the economically affordable lithography field to be reduced to a 572mm^2 level (22×26) by the 90 nm node. That trend is shown in Tables 2a and b.

DRAM chip sizes were deemed to be the most appropriate driver of affordable lithography field sizes. In the present *ITRS* chip-size model for DRAMs, the introduction-level chip size is targeted to be smaller than the new affordable 572mm^2 lithography field size, fitting at least one introduction-level chip size within the field. The production-level DRAM model fits at least two die within the affordable field. The combination of technology-node scaling and cell design improvements (A-factor reduction) accomplishes that goal, while also maintaining a goal of doubling on-chip bits every two years. However, the slowing of DRAM design improvements causes a requirement to add fewer on-chip bits to stay under the

162 Overall Roadmap Technology Characteristics

affordable lithography field limit. This accomplished in the present DRAM model by slowing the short-term Moore's Law bits/chip rate from $2\times/1.5$ years to $2\times/$ two years. In the later years of the Roadmap the rate is stretched even more to $2\times/2.5$ to three years, as required. The data targets for the DRAM model are included in Tables 1c, d, e, and f.

Both the DRAM and MPU models depend upon achieving the aggressive DRAM and MPU design and process improvement targets. If those targets slip, then pressure will increase to print chip sizes larger than the present roadmap, or further slow the rate of "Moore's-Law" on-chip functionality. Either of these consequences will result in a negative impact upon cost-per-function reduction rates—the classical measure of our industries productivity-improvement and competitiveness.

With increasing cost reduction pressures, the need for the 300 mm productivity boost will also increase in urgency, especially for leading-edge manufacturers, but the poor economy will create challenges and limit capital investment. The 2001 Wafer-Diameter Generation roadmap (see Tables 2a and b) is consistent with the ramp of 300 mm capacity beginning 2001. Also, the first manufacturing capability for the next $1.5\times$ wafer size conversion to 450 mm diameter is not anticipated to be required until 2013 in the present roadmap. However, should the other productivity-improvement drivers (lithography and design/process improvements) fail to stay on schedule, there would be a need to accelerate the use of increased wafer diameter as a productivity improvement.

The affects of future technology acceleration/deceleration and the timing of the next wafer generation conversion requires the development and application of comprehensive long-range factory productivity and industry economic models. Such industry economic modeling (IEM) work is being sponsored and carried out jointly by SEMI and International SEMATECH.

Table 2a Lithographic-Field and Wafer-Size Trends—Near-term Years

(Note: 2001 Lithographic field sizes represent current capability)

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length(nm)	65	53	45	37	32	28	25
<i>Lithography Field Size</i>							
Lithography Field Size—area (mm ²)	800	800	800	800	800	800	572
Lithographic field size — length (mm)	32	32	32	32	32	32	26
Lithographic field size — width (mm)	25	25	25	25	25	25	22
<i>Maximum Substrate Diameter (mm) — High-volume Production (>20K wafer starts per month)</i>							
Bulk or epitaxial or SOI wafer	300	300	300	300	300	300	300

Table 2b Lithographic-Field and Wafer Size Trends—Long-term Years

Year of Production	2010	2013	2016
DRAM ½ Pitch (nm)	45	32	22
MPU/ASIC ½ Pitch (nm)	45	32	22
MPU Printed Gate Length (nm)	25	18	13
MPU Physical Gate Length (nm)	18	13	9
<i>Lithography Field Size</i>			
Maximum lithographic field size—area (mm ²)			
Lithography Field Size—area (mm ²)	572	572	572
Maximum lithographic field size—length (mm)	26	26	26
Maximum lithographic field size—width (mm)	22	22	22
<i>Maximum Substrate Diameter (mm)—High-volume Production (>20K wafer starts per month)</i>			
Bulk or epitaxial or SOI wafer	300	450	450

PERFORMANCE OF PACKAGED CHIPS

NUMBER OF PADS AND PINS / PAD PITCH, COST PER PIN, FREQUENCY

The demand for a higher number of functions on a single chip requires the integration of an increased number of transistors or bits (memory cells) for each product generation. Typically, the number of pads and pins necessary to allow Input/Output (I/O) signals to flow to and from an integrated circuit increases as the number of transistors on a chip increases. (Refer to Tables 3a and b)

Additional power and ground connections to the chip are also necessary to optimize power management and to increase noise immunity. Based upon chip pad-count numbers supplied by the Test ITWG, logic products (MPUs and high-performance ASICs) both approach 4-6K pads over the *ITRS* period. The MPU products are forecast to increase the total number of pads through this period by nearly 50%, and the ASICs double the maximum number of pads per chip. The two product types also differ significantly in the ratio of power/ground pads. The MPU product pad counts typically have 1:3 signal I/O pads and 2:3 power and ground pads, or two power/ground pads for every signal I/O pad. Unlike MPUs, high-performance ASIC product pad counts typically include one power/ground pad for each signal I/O pad.

Table 3a Performance of Packaged Chips: Number of Pads and Pins—Near-term Years

<i>Year of Production</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>
<i>DRAM ½ Pitch (nm)</i>	130	115	100	90	80	70	65
<i>MPU/ASIC ½ Pitch (nm)</i>	150	130	107	90	80	70	65
<i>MPU Printed Gate Length (nm)</i>	90	75	65	53	45	40	35
<i>MPU Physical Gate Length (nm)</i>	65	53	45	37	32	28	25
<i>Number of Chip I/Os (Number of Total Chip Pads) — Maximum</i>							
<i>Total pads—MPU</i>	3072	3072	3072	3072	3072	3072	3072
<i>Signal I/O—MPU (1/3 of total pads)</i>	1024	1024	1024	1024	1024	1024	1024
<i>Power and ground pads—MPU (2/3 of total pads)</i>	2048	2048	2048	2048	2048	2048	2048
<i>Total pads—ASIC high-performance</i>	3000	3200	3400	3600	4000	4200	4400
<i>Signal I/O pads—ASIC high-performance</i>	1500	1600	1700	1800	2000	2100	2200
<i>Power and ground pads—ASIC high-performance (½ of total pads)</i>	1500	1600	1700	1800	2000	2100	2200
<i>Number of Total Package Pins—Maximum [1]</i>							
<i>Microprocessor/controller, cost-performance</i>	480–1,200	480–1320	500–1452	500–1600	550–1760	550–1936	600–2140
<i>Microprocessor/controller, high-performance</i>	1200	1320	1452	1,600	1,760	1,936	2,140
<i>ASIC (high-performance)</i>	1700	1870	2057	2263	2489	2738	3012

Notes for Tables 3a and 3b:

[1] Pin counts will be limited for some applications where fine pitch array interconnect is used by PWB technology and system cost.

The highest pin count applications will as a result use larger pitches and larger package sizes.

The reference to signal pin ratio will also vary greatly dependent on applications with an expected range from 2:1 to 1:4

Table 3b Performance of Packaged Chips: Number of Pads and Pins—Long-term Years

Year of Production	2010	2013	2016
DRAM ½ Pitch (nm)	45	32	22
MPU/ASIC ½ Pitch (nm)	45	32	22
MPU Printed Gate Length (nm)	25	18	13
MPU Physical Gate Length (nm)	18	13	9
<i>Number of Chip I/Os (Number of Total Chip Pads) — Maximum</i>			
Total pads—MPU	3840	4224	4416
Signal I/O—MPU (1/3 of total pads)	1280	1408	1472
Power and ground pads—MPU (2/3 of total pads)	2560	2816	2944
Total pads—ASIC high-performance	4800	5400	6000
Signal I/O pads—ASIC high-performance	2400	2700	3000
Power and ground pads—ASIC high-performance (½ of total pads)	2400	2700	3000
<i>Number of Total Package Pins—Maximum [1]</i>			
Microprocessor/controller, cost-performance	780–2782	1014–3616	1318–4702
Microprocessor/controller, high-performance	2782	3616	4702
ASIC (high-performance)	4009	5335	7100

Package pin count (Tables 3 a and b) and cost-per-pin (Tables 4 a and b), provided by the Assembly and Packaging ITWG, point out challenges to future manufacturing economics. Based upon the projected growth in the number of transistors/chip, it is forecast that the number of package pin/balls will continue to grow at an annual rate of approximately 10%, while the cost/pin decreases at 5%/year. These trends make it more challenging for suppliers of packaging technologies to deliver cost-effective solutions, because the overall average cost of packaging will increase annually at 5%/year ($.95 \text{ cost/pin} \times 1.10 \text{ pins/year} = 1.05 \text{ cost/year}$).

In the very competitive consumer electronics product environment, prices for high-volume, high-tech products such as PCs and cell phones tend to remain flat or even decrease. These same high-tech products typically also deliver twice the performance every two years. This is the end-use market environment of the leading-edge semiconductor manufacturer, and it is the fundamental economic driver behind the ITRS economic requirement to reduce cost per function (bits, transistors) at an annual 30% or faster rate ($2 \times \text{functionality/chip at flat price every two years} = 29\%/\text{year}$).

If future semiconductor component products must be targeted to maintain constant or decreasing prices and the average number of pins per unit increases at 10% while the average cost per pin decreases at only 5%, then the following will occur:

1. the average packaging share of total product cost will double over the 15-year roadmap period, and
2. the ultimate result will be greatly reduced gross profit margins and limited ability to invest in R&D and factory capacity.

This conclusion is one of the drivers behind the industry trends to reduce the overall system pin requirements by combining functionality into Systems-on-Chip (SoC) and through the use of multi-chip modules, bumped chip-on-board (COB), and other creative solutions.

In addition to the need to increase functionality while exponentially decreasing cost per function, there is also a market demand for higher-performance, cost-effective products. Just as Moore's Law predicts that functions-per-chip will double every 1.5–2 years to keep up with consumer demand, there is a corresponding demand for processing electrical signals at progressively higher rates. In the case of MPUs, processor instructions/second have also historically doubled every 1.5–2 years. For MPU products, increased processing power, measured in millions of instructions per second (MIPs), is accomplished through a combination of "raw technology performance" (clock frequency) multiplied by "architectural performance" (instructions per clock cycle). The need for a progressively higher operational frequency associated with an

166 Overall Roadmap Technology Characteristics

increasing average chip size will continue to demand the development of novel process, design, and packaging techniques.

These considerations are reflected in Tables 4c and d, which includes line items contributed by the Design and Assembly and Packaging ITWGs to forecast the maximum on-chip and chip-to-board frequency trends. The highest frequency obtainable in each product generation is directly related to the intrinsic transistor performance (on-chip, local clock). The difference between this “local” frequency and the frequency of signals traveling across the chip increases due to degradation of signal propagation delay caused by line-to-line and line-to-substrate capacitive coupling. Additional signal degradation is associated with the inductance of wire bonds and package leads. Direct chip attachment may eventually be the only viable way to eliminate any parasitic effect introduced by the package. To optimize signal and power distribution across the chip, it is expected that the number of layers of interconnect will continue to increase. As size downscaling of interconnect also continues, wider use of copper (low resistivity) and various inter-metal insulating materials of progressively lower dielectric constant ($\kappa \sim 2-3$) will be adopted in the chip fabrication process. Multiplexing techniques will also be used to increase the chip-to-board operating frequency (off-chip).

Table 4a Performance and Package Chips: Pads, Cost—Near-term Years [Update]

Year of Production		2001	2002	2003	2004	2005	2006	2007
	DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
	MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
	MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
	MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
Chip Pad Pitch (micron)								
Was	Pad pitch—ball bond	45	35	30	25	20	20	20
Is	Pad pitch—ball bond	45	<u>40</u>	<u>35</u>	<u>30</u>	<u>25</u>	20	20
Was	Pad pitch—wedge bond	40	35	30	25	20	20	20
Is	Pad pitch—wedge bond	<u>50</u>	<u>50</u>	<u>40</u>	<u>40</u>	<u>35</u>	<u>35</u>	<u>30</u>
Was	Pad Pitch—area array flip-chip (cost-performance, high-performance)	160	160	150	150	130	130	120
Is	Pad Pitch—area array flip-chip (cost-performance, high-performance)	<u>200</u>	<u>180</u>	150	150	<u>100</u>	<u>100</u>	<u>80</u>
Was	Pad Pitch—peripheral flip-chip (handheld, low-cost, harsh)	150	130	120	110	100	90	80
Is	Pad Pitch—peripheral flip-chip (handheld, low-cost, harsh)	<u>80</u>	<u>80</u>	<u>60</u>	<u>60</u>	<u>40</u>	<u>40</u>	<u>30</u>
Cost-Per-Pin								
Was	Package cost (cents/pin) (cost-performance)—minimum—maximum	0.80–1.60	0.75–1.44	0.70–1.30	0.66–1.17	0.61–1.06	0.56–1.03	0.64–1.00
Is	Package cost (cents/pin) (cost-performance)—minimum—maximum	0.80–1.60	0.75– <u>1.30</u>	0.70– <u>1.24</u>	0.66–1.17	0.61– <u>1.11</u>	0.56– <u>1.06</u>	0.64– <u>1.01</u>
	Package cost (cents/pin) (Memory)—minimum—maximum	0.36–1.54	0.34–1.39	0.32–1.26	0.30–1.14	0.28–1.03	0.27–0.93	0.27–0.84

Table 4b Performance and Package Chips: Pads, Cost—Long-term Years [*Update*]

Year of Production		2010	2013	2016
	DRAM ½ Pitch (nm)	45	32	22
	MPU/ASIC ½ Pitch (nm)	45	32	22
	MPU Printed Gate Length (nm)	25	18	13
	MPU Physical Gate Length (nm)	18	13	9
Chip Pad Pitch (micron)				
	Pad pitch—ball bond	20	20	20
	Pad Pitch—wedge bond	20	20	20
Was	Pad Pitch—area array (cost-performance, high-performance)	90	80	70
Is	Pad Pitch—area array (cost-performance, high-performance)	<u>70</u>	<u>70</u>	<u>50</u>
Was	Pad Pitch—peripheral flip-chip (handheld, low-cost, harsh)	60	45	30
Is	Pad Pitch—peripheral flip-chip (handheld, low-cost, harsh)	<u>20</u>	<u>20</u>	<u>15</u>
Cost-Per-Pin				
	Package cost (cents/pin) (cost-performance)— minimum—maximum	0.49–0.98	0.42–0.93	0.36–0.79
	Package cost (cents/pin) (Memory)— minimum—maximum	0.22–0.54	0.19–0.39	0.19–0.33

Table 4c Performance and Package Chips: Frequency On-Chip Wiring Levels—Near-term Years [*Update*]

YEAR OF PRODUCTION		2001	2002	2003	2004	2005	2006	2007
	DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
	MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
	MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
	MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
Chip Frequency (MHz)								
	On-chip local clock	1,684	2,317	3,088	3,990	5,173	5,631	6,739
	Chip-to-board (off-chip) speed (high-performance, for peripheral buses)[1]	1,684	2,317	3,088	3,990	5,173	5,631	6,739
Was	Maximum number wiring levels—maximum	7	8	8	8	9	9	9
Is	Maximum number wiring levels—maximum	<u>10</u>	<u>10</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>14</u>	<u>14</u>
Was	Maximum number wiring levels—minimum	7	7	8	8	8	9	9
Is	Maximum number wiring levels—maximum	<u>10</u>	<u>10</u>	<u>12</u>	<u>9</u>	<u>10</u>	<u>10</u>	<u>10</u>

Table 4d Performance and Package Chips: Frequency, On-Chip Wiring Levels—Long-term Years [Update]

Year of Production		2010	2013	2016
	DRAM ½ Pitch (nm)	45	32	22
	MPU/ASIC ½ Pitch (nm)	45	32	22
	MPU Printed Gate Length (nm)	25	18	13
	MPU Physical Gate Length (nm)	18	13	9
Chip Frequency (MHz)				
	On-chip local clock	11,511	19,348	28,751
	Chip-to-board (off-chip) speed (high-performance, for peripheral buses)[1]	11,511	19,348	28,751
Was	Maximum number wiring levels—maximum	10	10	10
Is	Maximum number wiring levels—maximum	14	15	15
Was	Maximum number wiring levels—minimum	9	9	10
Is	Maximum number wiring levels—minimum	10	11	11

Note for Tables 4c and 4d:

[1] The off chip frequency is expected to increase for a small number of high speed pins which will be used in combination with a large number of lower speed pins

[2] In 2001, high-speed serial communications transceiver devices are achieving chip-board frequencies of 3.125 GHz using CMOS, and 10 GHz using SiGe. In 2002 it is expected that 10 GHz transceivers will be fabricated using CMOS. 40 GHz SiGe devices are expected in 2003. The roadmap for higher levels of integration with wider bus widths, is shown in the High Frequency Serial Communications section in the Test chapter.

Add

[3] The Minimum number of wiring levels represents the Interconnect metal levels, and the Maximum number of Interconnect wiring levels includes the Minimum number of wiring levels plus additional optional levels required for power, ground, signal conditioning, and integrated passives (i.e., capacitors).

ELECTRICAL DEFECT DENSITY

The latest targets for electrical defect density of DRAM, MPU, and ASIC (necessary to achieve 83–89.5 % chip yield in the year of volume production) are shown in Tables 5a and b. The allowable number of defects is calculated by taking into account the different chip sizes based on the latest chip size model forecasts, as reported in Table 1 for DRAM and microprocessors. In addition, the data in the table are now reported only at the production-level of the product life-cycle. Other defect densities may be calculated at different chip sizes at the same technology node by using the formula found in the *Yield Enhancement* chapter. The approximate number of masks for logic devices is included as an indicator of the ever-increasing process complexity.

Table 5a Electrical Defects—Near-term Years

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
DRAM Random Defect D_0 at production chip size and 89.5% yield (faults/m ²) §	1,963	2,493	2,148	2,748	1,752	2236	1426
MPU Random Defect D_0 at production chip size and 83% yield (faults/m ²) §§	1,356	1,356	1,356	1,356	1,356	1,356	1,356
# Mask Levels – MPU	25	25	25	25	25	27	27
# Mask Levels – DRAM	21	22	24	24	24	24	24

Table 5b Electrical Defects—Long-term Years

Year of Production	2010	2013	2016
DRAM ½ Pitch (nm)	45	32	22
MPU/ASIC ½ Pitch (nm)	45	32	22
MPU Printed Gate Length (nm)	25	18	13
MPU Physical Gate Length (nm)	18	13	9
DRAM Random Defect D_0 at production chip size and 89.5% yield (faults/m ²) §	1356	1356	1356
MPU Random Defect D_0 at production chip size and 83% yield (faults/m ²) §§	1464	1116	1134
# Mask Levels – MPU	27	29	29
# Mask Levels – DRAM	26	26	26

Notes for Tables 5a and 5b:

D_0 —defect density

§ DRAM Model—Cell Factor (design/process improvement) targets are as follows:

1999–2002/8×; 2003–2010/6×; 2011–2016/4×

DRAM product generations are usually increased by 4×bits/chip every four years with interim 2×bits/chip generations, except:

1. at the Introduction phase, after the 8Gbit interim generation, the introduction rate is 4×/five years (2×/two–three years); and
2. at the Production phase, after the interim 32Gbit generation, the introduction rate is 4×/five years (2×/two–three years).

InTER-generation chip size growth rate varies to maintain one chip per 572mm² field at Introduction and two chips per 572mm² field at Production. The more aggressive “best case opportunity” technology node trends allow the Production-phase products to remain at 2×bits/chip every two years and still fit within the target of two DRAM chips per 572mm² field size, through the 32Gbit interim generation. The InTRA-generation chip size shrink model is 0.5×every technology node in-between cell factor reductions.

Refer to the Glossary for definitions of Introduction, Production, InTERgeneration, and InTRAgeneration terms.

§§ MPU Chip Size Model—Both the cost-performance and high-performance MPUs InTER-generation chip size growth rates are targeted to be flat through 2016, made possible by doubling the on-chip functionality every technology node cycle. The InTRA-generation chip size shrink model is 0.5× every two-year technology node through 2001, then 0.5× every three-year technology node after 2001.

POWER SUPPLY AND POWER DISSIPATION

Reduction of power supply voltage (see Tables 6a and b) is driven by several factors—reduction of power dissipation, reduced transistor channel length, and reliability of gate dielectrics. As seen in Tables 6a and b, the value of the power supply voltage is now given as a range.

Selection of a specific V_{dd} value continues to be a part of the analysis undertaken to simultaneously optimize speed and power for an IC, leading to a range of usable power supply voltages in each product generation. Values of V_{dd} as low as 0.5 volts are expected to be achieved by 2013, but the lowest target is now 0.4V by 2016 (versus 0.3V by 2014 in the 1999 ITRS).

Maximum power trends (e.g., for MPUs) are presented in three categories—1) high-performance desktop applications, for which a heat sink on the package is permitted; 2) cost-performance, where economical power management solutions of the highest performance are most important; and 3) portable battery operations. In all cases, total power consumption continues to increase, despite the use of a lower supply voltage. The increased power consumption is driven by higher chip operating frequencies, the higher interconnect overall capacitance and resistance; and the increasing gate leakage of exponentially-growing and scaled on-chip transistors.

Table 6a Power Supply and Power Dissipation—Near-term Years

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
Power Supply Voltage (V)							
V_{dd} (high performance)	1.1	1.0	1.0	1	0.9	0.9	0.7
V_{dd} (Low Operating Power, high V_{dd} transistors)	1.2	1.2	1.1	1.1	1.0	1.0	0.9
V_{dd} (Low Standby Power, high V_{dd} transistors)	1.2	1.2	1.2	1.2	1.2	1.2	1.1
Allowable Maximum Power [1]							
High-performance with heatsink (W)	130	140	150	160	170	180	190
Cost-performance (W)	61	75	81	85	92	98	104
Battery (W)—(hand-held)	2.4	2.6	2.8	3.2	3.2	3.5	3.5

Table 6b Power Supply and Power Dissipation—Long-term Years

Year of Production	2010	2013	2016
DRAM ½ Pitch (nm)	45	32	22
MPU/ASIC ½ Pitch (nm)	45	32	22
MPU Printed Gate Length (nm)	25	18	13
MPU Physical Gate Length (nm)	18	13	9
Power Supply Voltage (V)			
V_{dd} (high performance)	0.6	0.5	0.4
V_{dd} (Low Operating Power, high V_{dd} transistors)	0.8	0.7	0.6
V_{dd} (Low Standby Power, high V_{dd} transistors)	1.0	0.9	0.9
Allowable Maximum Power [1]			
High-performance with heatsink (W)	218	251	288
Cost-performance (W)	120	138	158
Battery (W)—(hand-held)	3.0	3.0	3.0

Note for Table 6a and 6b:

[1] Power will be limited more by system level cooling and test constraints than packaging

COST

Tables 7a and b are dedicated to cost trends. The ability to reduce the cost per function by an average 25–30% each year represents one of the unique features of the semiconductor industry and is a direct consequence of the market pressure to continue to deliver twice the functionality on-chip every 1.5–2 years in an environment of constant or reducing prices. In support of this cost reduction, a continuously increasing amount of investment is needed for R&D and manufacturing capital. Even on a per-factory basis, the capital cost of manufacturing continues to escalate. Yet, the semiconductor industry has historically delivered two times as many functions per chip every 1.5–2 years with no or only a moderate increase in chip size and cost (approximately constant cost per cm² of silicon). This technological and economic performance is the fundamental engine behind the growth of the semiconductor industry.

However, the customers in today's competitive market environment are resistant to even "moderate" increases in cost and the rate of doubling functions per chip (Moore's Law) is slowing. Therefore, the semiconductor manufacturers must seek a new model to deliver the same cost-per-function reduction requirements that have fueled industry growth. Consequently, the 1999 ITRS proposed a new model for achieving the required reduction: provide the customer twice the functionality every two years at a constant cost and average selling price (ASP) per chip. The 2001 ITRS uses the model and results in the same 29% cost reduction of a function (bit, transistor, etc.) that has been achieved historically by delivering four times the functionality per chip every three years at 1.4× increase in cost per unit. The DRAM and MPU cost models continue to use the need for that 29% cost-per-function productivity reduction rate as an economic driver of the industry. Therefore, that core cost-per-function trend has been used to set the INTRA-generation trends for the affordable cost/bit and cost/transistor for DRAM and microprocessors, respectively. Extrapolation of historical trends would indicate an "at introduction" affordable cost/bit of approximately 21 microcents for 2-Gbit DRAMs in 2001. In addition, the historical trends indicate that, within a DRAM generation, a 45%/year reduction in cost/bit should be expected.² A corresponding analysis conducted from published data for microprocessors yields similar results.³ As a result the 29%/year target for reduction in affordable cost/transistor from generation to generation is also being used in the MPU model, along with the 45%/year reduction rate within the same generation.

However, the Design ITWG has updated the MPU model, based upon recent data. The new data indicates that logic transistor size is improving only at the rate of the lithography (0.7× linear, 0.5×area reduction every technology node). Therefore in order to keep the MPU chip sizes flat, the number of transistors can be doubled only every technology node. The technology node rate is projected to return to a three-year cycle after 2001, therefore the transistors per MPU chip can double only every three years after 2001. DRAM memory bit cell design improvements are also slowing down, and the rate of bits per chip will also be slowing in the future to keep chip sizes under control. To compensate for slowing DRAM and MPU functions-per-chip, there will be increasing pressure to find alternative productivity enhancements from the equivalent productivity scaling benefits of chip and system-level architecture and designs.

Even though the rate of increase of on-chip functionality could slow in the future, the amount of functions/chip is still growing exponentially. As the number of functions/chip continues to increase, it becomes increasingly difficult and, therefore, costly to test the final products. This is reflected in the escalating cost of testers. Even though the cost/pin of testers is forecast to decline between 0% and 10% per year (Tables 7 a and b), the number of pins grows at 10%/year (Tables 4 a and b). Therefore, the need for accelerated implementation of Built-In-Self-Test (BIST) and Design-For-Testability (DFT) techniques will continue within the time frame of the 2001 International Technology Roadmap for Semiconductors. Further discussion is detailed in the *Test* chapter.

² McClean, William J., ed. *Mid-Term 1994: Status and Forecast of the IC Industry*. Scottsdale: Integrated Circuit Engineering Corporation, 1994.

McClean, William J., ed. *Mid-Term 1995: Status and Forecast of the IC Industry*. Scottsdale: Integrated Circuit Engineering Corporation, 1995.

³ a) Dataquest Incorporated. *x86 Market: Detailed Forecast, Assumptions, and Trends*. MCRO-WW-MT-9501. San Jose: Dataquest Incorporated, January 16, 1995.

b) Port, Otis; Reinhardt, Andy; McWilliams, Gary; and Brull, Steven V. "The Silicon Age? It's Just Dawning," Table 1. *Business Week*, December 9, 1996, 148–152.

Table 7a Cost—Near-term Years

Year of Production	2001	2002	2003	2004	2005	2006	2007
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65
MPU/ASIC ½ Pitch (nm)	150	130	107	90	80	70	65
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25
<i>Affordable Cost per Function</i> ++							
DRAM cost/bit at (packaged microcents) at samples/introduction	21	14.8	10.5	7.4	5.3	3.7	2.6
DRAM cost/bit at (packaged microcents) at production §	7.7	5.4	3.8	2.7	1.9	1.4	0.96
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at introduction §§	176	124	88	62	44	31	22
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at production §§	107	75	53	38	27	19	13.3
High-performance MPU (microcents/transistor) (including on-chip SRAM) at production §§	97	69	49	34	24	17	12
<i>Cost-Per-Pin</i>							
<i>Test Cost</i>							
Volume tester cost per high-frequency signal pin (\$K/pin) (high-performance ASIC)—maximum	4.0	3.0	3.0	3.0	3.0	3.0	3.0
Volume tester cost per high-frequency signal pin (\$K/pin) (high-performance ASIC)—minimum	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Notes for Tables 7a and 7b:

++ Affordable packaged unit cost per function based upon Average Selling Prices (ASPs) available from various analyst reports less Gross Profit Margins (GPMs); 35% GPM used for commodity DRAMs and 60% GPM used for MPUs; 0.5×/two years inTER-generation reduction rate model used; .55×/year inTRA-generation reduction rate model used; DRAM unit volume life-cycle peak occurs when inTRA-generation cost per function is crossed by next generation, typically seven–eight years after introduction; MPU unit volume life-cycle peak occurs typically after four–six years, when the next generation processor enters its ramp phase (typically two–four years after introduction).

§ DRAM Model—Cell Factor (design/process improvement) targets are as follows:

1999–2002/8×; 2003–2010/6×; 2011–2016/4×

DRAM product generations are usually increased by 4× bits/chip every four years with interim 2× bits/chip generations, except:

- 1) at the Introduction phase, after the 8Gbit interim generation, the introduction rate is 4×/five years (2×/two–three years); and
- 2) at the Production phase, after the interim 32Gbit generation, the introduction rate is 4×/five years (2×/two–three years).

InTER-generation chip size growth rate varies to maintain one chip per 572mm² field at Introduction and two chips per 572mm² field at Production. The more aggressive “best case opportunity” technology node trends allow the Production-phase products to remain at 2× bits/chip every two years and still fit within the target of two DRAM chips per 572mm² field size, through the 32Gbit interim generation. The InTRA-generation chip size shrink model is 0.5× every technology node in-between cell factor reductions.

Refer to the Glossary for definitions of Introduction, Production, InTERgeneration, and InTRAgeneration terms.

§§ MPU Chip Size Model—Both the cost-performance and high-performance MPUs InTER-generation chip size growth rates are targeted to be flat through 2016, made possible by doubling the on-chip functionality every technology node cycle. The InTRA-generation chip size shrink model is 0.5× every two-year technology node through 2001, then 0.5× every three-year technology node after 2001.

Table 7b Cost—Long-term Years

<i>Year of Production</i>	<i>2010</i>	<i>2013</i>	<i>2016</i>
<i>DRAM ½ Pitch (nm)</i>	45	32	22
<i>MPU/ASIC ½ Pitch (nm)</i>	45	32	22
<i>MPU Printed Gate Length (nm)</i>	25	18	13
<i>MPU Physical Gate Length (nm)</i>	18	13	9
<i>Affordable Cost per Function ++</i>			
<i>DRAM cost/bit (packaged microcents) at samples/introduction</i>	0.93	0.33	0.12
<i>DRAM cost/bit (packaged microcents) at production §</i>	0.34	0.12	0.042
<i>Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at introduction §§</i>	7.78	2.75	0.97
<i>Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at production §§</i>	4.71	1.66	0.59
<i>High-performance MPU (microcents/transistor) (including on-chip SRAM) at production §§</i>	4.31	1.52	0.54
<i>Cost-Per-Pin</i>			
<i>Test Cost</i>			
<i>Volume tester cost per high-frequency signal pin (\$K/pin) (high-performance ASIC)—maximum</i>	4	4	4
<i>Volume tester cost per high-frequency signal pin (\$K/pin) (high-performance ASIC)—minimum</i>	2	3	4