International Technology Roadmap for Semiconductors

1999 Edition

Overall Roadmap Technology Characteristics

&

Glossary

Table of Contents

List of Tables

OVERALL ROADMAP TECHNOLOGY CHARACTERISTICS

OVERVIEW

The Overall Roadmap Technology Characteristics (ORTC) tables provide a consolidated summary of the key technology metrics. As described in the Introduction, the year header on the tables may refer to different points in the development/life cycle of integrated circuits (ICs) characterized by this technology, depending on the individual line item metric. However, unless otherwise specified for a line, the default year header still refers (as in previous Roadmaps) to the year when product shipment first exceeds 10,000 units of ICs from a manufacturing site using "production tooling."

The ORTC tables are created early in the Roadmap process and are used as the basis for initiating the activities of the Technology Working Groups in producing their detailed chapters. The ORTC tables were 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. As a result, the ORTC tables went through several iterations until the Roadmap document was completed. The metric values of the ORTC tables can be found throughout the Roadmap in greater detail in each Technology Working Group chapter. Additionally, an ORTC glossary is provided as an appendix.

The new tables include an unprecedented level of cross-TWG and international coordination and consensus building to develop underlying models of trends and to reach agreement on target metrics. The ORTC tables have also expanded in line items compared to the 1997 NTRS. The expansion is due to the addition of several line items, which clarify the underlying models and communicate additional details of the interrelationships between technology areas.

ROADMAP TIMELINE

The timing at which future technology nodes are introduced has not changed since the ITRS 1998 Update. However, it has changed since the 1997 NTRS, which was the last edition with "explanatory text." Thus, we will next describe the node timing changes between the 1997 NTRS and the 1998/1999 ITRS. By international consensus, the "150-nm node" was eliminated, and the subsequent technology nodes, beginning with the 130-nm node, were pulled in by one year. Note that products built with the 180-nm technology node are shipping in 1999, fulfilling the forecast of a 2-year technology-node cycle from the 250-nm node in 1997. In addition, the introduction into the market place of the 130-nm technology node, which, in 1997, was predicted to occur in the year 2003, is now forecast to occur in the year 2002, representing a continuation of the 3-year cycle for DRAM metal half-pitch nodes. By agreement and driven by performance demand, the MPU gate length is forecast to continue scaling by about 70% per 2-year cycle through the 100-nm MPU gate-length in 2001, but is expected to return to a 3-year cycle thereafter. There is some optimism that DRAM half-pitch nodes could undergo an additional 1-year pull-in. This possibility will be re-evaluated during the year 2000 ITRS update. To reflect the variety of cycles and to allow for closer monitoring of future roadmap shifts, it was agreed to publish annual technology requirements from 1999 through 2005, called the "Near-Term Years," and at three-year (node) intervals thereafter, called the "Long-Term Years" (2008, 2011, 2014).

It should be observed that 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 tends to quickly 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.

BACKGROUND

PRODUCT GENERATIONS AND CHIP-SIZE MODEL

In this section, we will discuss "product generations" and their relationship to the technology nodes. In the past, these terms have often been used interchangeably, but 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. In this edition of the ITRS, "technology node" is still linked to an anticipated DRAM feature size (minimum metal half-pitch), but the implications of this connection are diminishing as the product evolution/shrink path becomes more complex. Thus, "technology node" is now not much more than a simple label for still somewhat convenient "tick-marks" along this path.

Historically, DRAM products have been recognized as the technology drivers for the whole 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. As anticipated, microprocessor products have now not only closed the technology gap with DRAM, but are now actually driving the most leading-edge lithography tools. It is now recognized that DRAM and microprocessor products share the technology leadership role.

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 still maximizing performance, which is dominated by the length of the transistor gate and by the number of interconnect layers. As a result, teams of both regional and international technical and business analysts worked to develop and reach consensus on models of the required functionality, chip size, cell area, and density. Additional line items were added to communicate the model consensus, and the underlying model assumptions are included in notations. It was agreed that the key ITRS technology node identifier would continue to be the DRAM half-pitch. Table 1 also includes the aggressive MPU gate-length performance-driven feature size and, for completeness, also tracks the MPU/ASIC product metal half-pitch that will trail slightly behind or equal to the DRAM half-pitch.

For each product generation, both the leading-edge ("at introduction") and the high-volume ("at production") DRAM products are indicated. As anticipated during the 1997 NTRS renewal, the recent availability of 193-nm wavelength exposure tools, working with complementary mask and photoresist technologies, should enable the 130 nm half-pitch node by 2002. The 150 nm half-pitch capability (now an annualized shrink target, rather than a formal node target) is still expected to become available by 2001, and possibly sooner. It is still anticipated that "non-optical" exposure techniques may become viable contenders to optical lithography with the advent of the 100 nm technology node, presently targeted in 2005 for DRAM metal half-pitch. Printed gate-length of 100 nm for MPU is targeted in 2001.

In comparing the 1999 edition of the ITRS with the 1997 Roadmap, 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 1995–1999. The overall schedule for introduction of a new product generation has once again been accelerated by one additional year, and a best-case opportunity exists for the industry to repeat the performance at the year 2001 Roadmap renewal.

Table 1a Product Generations and Chip Size Model—Near Term Years

YEAR TECHNOLOGY NODE	1999 180 nm	2000	2001	2002 130 nm	2003	2004	2005 100 nm	DRIVER
DRAM ½ Pitch (nm)	180	165	150	130	120	110	100	D ½
MPU Gate Length (nm) ††	140	120	100	85-90	80	70	65	M GATE
MPU/ASIC ½ Pitch (nm)	230	210	180	160	145	130	115	M AND A ½
ASIC Gate Length (nm)	180	165	150	130	120	110	100	A GATE
Memory								
Generation at introduction §	1G	—	2G	—	4G	—	8G	—
Functions per chip (Gbits)	1.07	—	2.15	—	4.29	—	8.59	Market — Moore's Law
Cell area factor	8.0	7.3	6.6	6.0	5.4	4.9	4.4	Market — Cost/Timing
Cell area (μm^2)	0.26	0.20	0.15	0.10	0.08	0.059	0.044	Market — Cost/Timing
Chip size at introduction (mm^2) §	400	—	438	—	480	—	526	Market — Cost/Timing
Cell array area at introduction (% of chip size) §	70%	—	72%	—	70%	—	72%	Market — Cost/Timing
Gbits/cm^2 at introduction §	0.27	—	0.49	—	0.89	—	1.63	Market — Cost/Timing
Generation at production §	256M	—	(512)	—	1G	—	2G	Market — Cost/Timing
Chip size at production (mm^2) §	132	—	145	—	159	—	174	Market — Cost/Timing
Cell array area at production (% of chip size) §	53%	—	55%	—	53%	—	54%	Market — Cost/Timing
Gbits/cm^2 at production §	0.20	—	0.37	—	0.68	—	1.23	Market — Cost/Timing
Logic (High-volume Microprocessor) Cost-performance *								
Generation at introduction †	p99c	—	p01c	—	p03c	—	p05c	—
Functions per chip (million transistors [Mtransistors])	23.8	—	47.6	—	95.2	—	190	Market — Moore's Law
Process/design annual improvement factor ++	0.90	0.90	0.90	0.91	0.92	0.93	0.93	Market — Cost/Timing
Transistor density SRAM at introduction (Mtransistors/cm^2)	35	50	70	95	128	173	234	Market — Cost/Timing
Transistor density logic at introduction (Mtransistors/cm^2)	6.6	9.4	13	18	24	33	44	Market — Cost/Timing
Chip size at introduction (mm^2) ***	340	—	340	—	372	—	408	Market — Cost/Timing

++ *The MPU Process/design improvement factor is an estimate of the additional annual functional area reduction required beyond the area reduction contributed by the MPU metal half-pitch reduction. Note that this additional area reduction for transistor density plays a role generally analogous to the "cell area factor" for DRAMs. It 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.*

Table 1a Product Generations and Chip Size Model—Near Term Years (continued)

YEAR TECHNOLOGY NODE	1999 180 nm	2000	2001	2002 130 nm	2003	2004	2005 100 nm	DRIVER
DRAM ½ Pitch (nm)	180	165	150	130	120	110	100	D ½
MPU Gate Length (nm) ††	140	120	100	85-90	80	70	65	M GATE
MPU/ASIC ½ Pitch (nm)	230	210	180	160	145	130	115	M AND A ½
ASIC Gate Length (nm)	180	165	150	130	120	110	100	A GATE
Logic (High-volume Microprocessor) Cost-performance *(continued)								
Cost performance MPU (Mtransistors/cm^2 at introduction) (including on-chip SRAM) ***	7	—	14	—	26	—	47	M Gate and M and A ½
Generation at ramp †	p97c	—	p99c	—	p01c	—	P03c	—
Chip size at ramp (mm^2) ***	170	—	170	—	214	—	235	Market — Cost/Timing
Cost performance MPU (Mtransistors/cm^2 at ramp, including on-chip SRAM) ***	7	—	14	—	22	—	41	M Gate and M and A ½
Logic (Low-volume Microprocessor) High-performance **								
Generation at ramp ‡	p99h	—	p01h	—	p03h	—	p05h	—
Functions per chip (million transistors)	110	—	220	—	441	—	882	Market — Moore's Law
Chip size at ramp (mm^2) ***	450	—	450	—	567	—	622	Market — Cost/Timing
High-performance MPU Mtransistors/cm^2 at ramp (including on-chip SRAM) ***	24	—	49	—	78	—	142	M Gate and M and A ½
ASIC								
ASIC usable Mtransistors/cm^2 (auto layout)	20	28	40	54	73	99	133	M Gate and M and A ½
ASIC max chip size at ramp (mm^2) (maximum lithographic field size)	800	800	800	800	800	800	800	Lithographic Field Size
ASIC maximum functions per chip at ramp (Mtransistors/chip) (fit in maximum lithographic field size)	160	224	320	432	584	800	1064	Market — Performance/ Timing

†† *Range of node targets indicates the acknowledgment of the difficulty of projecting the impact of the return to the 3-year technology node cycle starting in 2001 and the uncertainty of the long term years of the Roadmap timeframe.*

§ *DRAM Model—Generations 4× bits/chip every four years with interim 2× bits/chip generations; InTER-generation chip size growth rate model is 1.2× every four years; InTRA-generation chip size shrink model is 0.5× every three years beginning 1999.*

† *p is processor, numerals reflect year of introduction, c is cost-performance product.*

‡ *p is processor, numerals reflect year at ramp, h is high-performance product.*

* *MPU Cost-performance Model—Cost-performance MPU includes small level 1 (L1) on-chip SRAM (32Kbyte/1999), but consists primarily of logic transistor functionality; both SRAM and Logic functionality doubles every two years.*

** *MPU High-performance Model—High-performance MPU includes large level 2 (L2) on-chip SRAM (2MByte/1999) added to ramp-level cost-performance core functionality shrunk from 2-year-prior generation (P99h = 11.9M transistor (Mtransistors) (shrunk P97 core) + 98Mtransistors (2048 bytes × 8 bits/byte × 6 transistors/bit) L2 SRAM = 110Mtransistors/1999); both SRAM and Logic functionality doubles every two years.*

*** *MPU Chip Size Model—Both the cost-performance and high-performance MPUs target for InTER-generation chip size growth rate model is flat through 2001, then 1.2× growth every four years after 2001; InTRA-generation chip size shrink model is 0.5× every two years through 2001, then 0.5× every three years after 2001.*

Table 1b Product Generations and Chip Size Model—Long Term Years

YEAR TECHNOLOGY NODE	2008 70 nm	2011 50 nm	2014 35 nm
DRAM ½ Pitch (nm)	70	50	35
MPU Gate Length (nm) ††	45	30-32	20-22
MPU/ASIC ½ Pitch (nm)	80	55	40
ASIC Gate Length (nm)	70	50	35
Memory			
Generation at introduction §	—	64G	—
Functions per chip (billion bits (Gbits))	24.3	68.7	194
Cell area factor	3.5	3.0	2.5
Cell area (μm^2)	0.017	0.008	0.003
Chip size at introduction (mm^2) §	603	691	792
Cell area efficiency at introduction (% of chip size) §	69%	75%	75%
Gbits/cm^2 at introduction §	4.03	9.94	24.5
Generation at production §	—	16G	—
Chip Size at production (mm^2) §	199	229	262
Cell area efficiency at production (% of chip size) §	52%	56%	57%
Gbits/cm^2 at production §	3.05	7.51	18.5
Logic (High-volume Microprocessor) Cost-performance *			
Generation at introduction †	—	p11c	—
Functions per chip (million transistors (Mtransistors))	539	1,523	4,308
Process/design improvement factor	0.93	0.93	0.93
Transistor density SRAM at introduction (Mtransistors/cm^2)	577	1,423	3,510
Transistor density logic at introduction (Mtransistors/cm^2)	109	269	664
Chip size at introduction (mm^2) ***	468	536	615

Table 1b Product Generations and Chip Size Model—Long Term Years (continued)

YEAR TECHNOLOGY NODE	2008 70 nm	2011 50 nm	2014 35 nm
DRAM ½ Pitch (nm)	70	50	35
MPU Gate Length (nm)††	45	30–32	20–22
MPU/ASIC ½ Pitch (nm)	80	55	40
ASIC Gate Length (nm)	70	50	35
Logic (High-volume Microprocessor) Cost-performance * (continued)			
Cost-performance MPU Mtransistors/cm^2 at introduction (including on-chip SRAM) ***	115	284	701
Generation at ramp †	—	p09c	—
Chip size at ramp (mm^2) ***	269	308	354
Cost performance MPU Mtransistors/cm^2 at ramp (including on-chip SRAM) ***	100	247	609
**Logic (Low-volume Microprocessor) High-performance ** **			
Generation at ramp ‡	—	p11h	—
Functions per chip (million transistors)	2,494	7,053	19,949
Chip size at ramp (mm^2) ***	713	817	937
High-performance MPU Mtransistors/cm^2 at ramp (including on-chip SRAM) ***	350	863	2,130
ASIC			
ASIC usable Mtransistors/cm^2 (auto layout)	328	811	2,000
ASIC maximum chip size at ramp (mm^2) (maximum lithographic field size)	800	800	800
ASIC maximum functions per chip at ramp (Mtransistors/chip) (fit in maximum lithographic field size)	2,624	6,488	16,000

Since only the 2011 odd-year product generation data column is available in the Long Term table format, interpolated numbers were calculated and included in the 2008 and 2014 node columns. The extended market-need-based product trends for the product generation two-year-cycle years (1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013) are forecast to follow patterns established in Near Term Table 1a.

†† *Range of node targets indicates the acknowledgment of the difficulty of projecting the impact of the return to the 3-year technology node cycle starting in 2001 and the uncertainty of the long term years of the Roadmap timeframe.*

§ *DRAM Model—Generations 4× bits/chip every four years with interim 2× bits/chip generations; InTER-generation chip size growth rate model is 1.2× every four years; InTRA-generation chip size shrink model is 0.5× every three years beginning 1999.*

† *p is processor, numerals reflect year of introduction, c is cost-performance product.*

‡ *p is processor, numerals reflect year at ramp, h is high-performance product.*

* *MPU Cost-performance Model—Cost-performance MPU includes small level 1 (L1) on-chip SRAM (32Kbyte/1999), but consists primarily of logic transistor functionality; both SRAM and Logic functionality doubles every two years.*

** *MPU High-performance Model—High-performance MPU includes large level 2 (L2) on-chip SRAM (2Mbyte/1999) added to ramp-level cost-performance core functionality shrunk from 2-year-prior generation (P99h = 11.9M transistor (Mtransistors) (shrunk P97 core) + 98Mtransistors (2048 bytes × 8 bits/byte × 6 transistors/bit) L2 SRAM = 110Mtransistors/1999); both SRAM and Logic functionality doubles every two years.*

*** *MPU Chip Size Model—Both the cost-performance and high-performance MPUs target for InTER-generation chip size growth rate model is flat through 2001, then 1.2× growth every four years after 2001; InTRA-generation chip size shrink model is 0.5× every two years through 2001, then 0.5× every three years 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 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 consensus 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 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× lithography reduction rate) during every technology node period. Due to an additional 2-year cycle through the year 2001, the MPU products will be able to target a flat die size. However, after 2001, the intra-generation chip size of MPUs will also grow at a 20%-per-four-years rate due to the return to a 3-year technology node cycle. Doubling the on-chip functionality (transistors) and growing only 20% every four years will require MPU chip and process designers to add to lithography improvements an additional design/process chip-size reduction of 7% annually after 2001.

DRAM products must meet the requirements of the limited intra-generation chip-size growth and also maintain a cell area ratio of less than 70% of total die area. Therefore, DRAM products require aggressive cell area factors (cell area in units of minimum-feature-size-squared). The Front-End Processes 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 have been added to ORTC Table 1. 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 determined by the field size of the lithographic tool and the size and aspect ratio of the chips being printed on the wafer. The roadmap for this useful parameter is included in Table 2.

In the present ITRS chip-size model for DRAMs, the introduction chip size is smaller than the existing 800-mm^2 capability of large step-and-scan fields. Even the large high-performance MPU chip sizes are not forecast to grow larger than 800 mm^2 until the 2011 generation. However, the models depend upon not only meeting the present lithography targets, but are also dependent on 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.

Historically, another major productivity increase has resulted from the industry's conversion to wafer sizes of progressively larger diameter. It is projected that the number of available chips will increase by a factor of 2.4–2.5 on a 300-mm wafer compared to the number of chips available on a 200-mm wafer. However, it is presently forecast that conversion to high-volume production using 300-mm diameter wafers will not begin until 2001, even though a full-flow pilot line was demonstrated in 1998 and actually produced commercial 64-Mbit DRAMs. The delay of 300-mm ramp was due to a combination of: 1) overbuilt 200-mm fabs; 2) a significant slow-down in the Asian economies; and 3) the rapid deployment of accelerated-technology (primarily lithography) upgrade equipment into existing factories to meet the demand for increased unit shipments of leading-edge products in an environment of rapidly falling market prices.

The semiconductor industry shows signs of a recovery having begun 1999. So, the need for the 300-mm productivity boost should increase in urgency, especially for leading-edge manufacturers who have little or no over-capacity. The 1999 Wafer-Diameter Generation roadmap was adapted to be consistent with the delayed start of 300 mm. Therefore, the first pilot capability for the next 1.5× wafer size conversion to 450-

mm diameter is not anticipated to be required until 2009–10. Likewise, consistent with the 300-mm wafer generation experience, 450-mm wafer production will not begin volume ramp (20K wafer start per week capacity) until the 2012–2014 timeframe. 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.

Table 2a Chip-Size, Lithographic-Field and Wafer-Size Trends—Near Term Years

(Note: 1999 Lithographic field sizes represent current capability)

YEAR TECHNOLOGY NODE	1999 180 nm	2000	2001	2002 130 nm	2003	2004	2005 100 nm
DRAM Chip Size							
Cell area factor	8.0	7.3	6.6	6.0	5.4	4.9	4.4
Cell area (μm^2)	0.26	0.20	0.15	0.10	0.08	0.059	0.044
Cell array area at introduction (% of chip size) §	70%	—	72%	—	70%	—	72%
DRAM generation at introduction §	1G	—	(2G)	—	4G	—	(8G)
Chip size at introduction (mm^2) §	400	—	438	—	480	—	526
Chip height at 2:1 aspect ratio (mm)	14.1	—	14.8	—	15.5	—	16.2
Chip length at 2:1 aspect ratio (mm)	28.3	—	29.6	—	31.0	—	32.4
Cell array area at sample (% of chip size) §	60%	—	63%	—	61%	—	62%
DRAM generation at sample §	(512M)	—	1G	—	(2G)	—	4G
Chip size at sample (mm^2) §	230	—	252	—	276	—	302
Chip height at 2:1 aspect ratio (mm)	10.7	—	11.22	—	11.8	—	12.3
Chip length at 2:1 aspect ratio (mm)	21.5	—	22.5	—	23.5	—	24.6
Cell array area at production (% of chip size) §	53%	—	55%	—	53%	—	54%
Generation at production §	256M	—	(512)	—	1G	—	2G
Chip size at production (mm^2) §	132	—	145	—	159	—	174
Chip height at 2:1 aspect ratio (mm)	8.1	—	8.5	—	8.9	—	9.33
Chip length at 2:1 aspect ratio (mm)	16.3	—	17.0	—	17.8	—	18.7
Cell array area at ramp (% of chip size) §	.48%	—	49%	—	45%	—	47%
Generation at ramp §	(128)	—	256M	—	(512)	—	1G
Chip size at ramp (mm^2) §	74	—	83	—	91	—	100
Chip height at 2:1 aspect ratio (mm)	4.2	—	6.4	—	8.9	—	9.3
Chip length at 2:1 aspect ratio (mm)	8.4	—	12.7	—	17.9	—	18.6
MPU Chip Size							
*High-performance MPU generation at ramp ** ‡*	p99h	—	p01h	—	p03h	—	P05h
*Chip size at ramp (mm^2) *****	450	—	450	—	567	—	622
Maximum lithographic field size — area (mm^2)	800	800	800	800	800	800	800
Maximum lithographic field size — length (mm)	32	32	32	32	32	32	32
Maximum lithographic field size — width (mm)	25	25	25	25	25	25	25
Minimum lithographic field size — area (mm^2)	484	506	529	552	576	600	625
Minimum lithographic field size — length (mm)	22	22.5	23	23.5	24	24.5	25
Minimum lithographic field size — width (mm)	22	22.5	23	23.5	24	24.5	25
Maximum Substrate Diameter (mm) — High-volume Production (>20K wafer starts per month)							
Bulk or epitaxial or SOI wafer	200	200	300	300	300	300	300

Table 2b Chip-Size, Lithographic-Field and Wafer Size Trends—Long Term Years

YEAR TECHNOLOGY NODE	2008 70 nm	2011 50 nm	2014 35 nm
DRAM Chip Size			
Cell area factor	3.5	3.0	2.5
Cell area (μm^2)	0.017	.008	0.003
Cell array area at introduction (% of chip size) §	69%	75%	75%
DRAM generation at introduction §	—	64G	—
Chip size at introduction (mm^2) §	603	691	792
Chip height at 2:1 aspect ratio (mm)	17.4	18.6	19.9
Chip length at 2:1 aspect ratio (mm)	34.7	37.2	39.8
Cell array area at sample (% of chip size) §	60%	65%	65%
DRAM generation at sample §	—	(32G)	—
Chip size at sample (mm^2) §	347	398	456
Chip height at 2:1 aspect ratio (mm)	13.2	14.1	15.1
Chip length at 2:1 aspect ratio (mm)	26.3	28.2	30.2
Cell array area at production (% of chip size) §	52%	56%	57%
Generation at production §	—	16G	—
Chip size at production (mm^2) §	199	229	262
Chip height at 2:1 aspect ratio (mm)	10.0	10.7	11.4
Chip length at 2:1 aspect ratio (mm)	20.2	21.4	22.9
Cell array area at ramp (% of chip size) §	45%	49%	49%
Generation at ramp §	—	(8G)	—
Chip size at ramp (mm^2) §	115	131	151
Chip height at 2:1 aspect ratio (mm)	7.6	8.1	8.7
Chip length at 2:1 aspect ratio (mm)	15.2	16.2	17.4
MPU Chip Size			
High-performance MPU generation at ramp ** ‡	—	p11h	—
Chip size at ramp (mm^2) ***	713	817	937
Maximum lithographic field size—area (mm^2)	800	800	800
Maximum lithographic field size—length (mm)	32	32	32
Maximum lithographic field size—width (mm)	25	25	25
Minimum lithographic field size—area (mm^2)	625	625	625
Minimum lithographic field size—length (mm)	25	25	25
Minimum lithographic field size—width (mm)	25	25	25
Maximum Substrate Diameter (mm)—High-volume Production (>20K wafer starts per month)			
Bulk or epitaxial or SOI wafer	300	300	450

SOI—Silicon on Insulator

§ *DRAM Model—Generations 4× bits/chip every four years with interim 2× bits/chip generations; InTER-generation chip size growth rate model is 1.2× every four years; InTRA-generation chip size shrink model is 0.5× every three years beginning 1999.*

‡ *p is processor, numerals reflect year at ramp, h is high-performance product.*

** *MPU High-performance Model—High-performance MPU includes large level 2 (L2) on-chip SRAM (2MByte/1999) added to ramp-level cost-performance core functionality shrunk from 2-year-prior generation (P99h = 11.9M transistor (Mtransistors) (shrunk P97 core) + 98Mtransistors (2048 bytes ×8 bits/byte ×6 transistors/bit) L2 SRAM = 110Mtransistors/1999); both SRAM and Logic functionality doubles every two years.*

*** *MPU Chip Size Model—Both the cost-performance and high-performance MPUs target for InTER-generation chip size growth rate model is flat through 2001, then 1.2× growth every four years after 2001; InTRA-generation chip size shrink model is 0.5× every two years through 2001, then 0.5× every three years after 2001.*

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 Table 3.

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 TWG, logic products (MPUs and high-performance ASICs) both approach a maximum of about 4K pads over the ITRS period. The MPU products are forecast to almost double the total number of pads through this period, whereas the ASICs nearly triple 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 have a typical ratio of 1/3 signal I/O pads to 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 only one power/ground pad for each signal I/O pad.

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

YEAR / *TECHNOLOGY NODE*	*1999* / *180 nm*	*2000*	*2001*	*2002* / *130 nm*	*2003*	*2004*	*2005* / *100 nm*
Number of Chip I/Os (Number of Total Chip Pads) — Maximum							
Total pads—MPU	2,304	2,560	3,042	3,042	3,042	3,042	3,042
Signal I/O—MPU (1/3 of total pads)	768	1,024	1,024	1,024	1,024	1,024	1,024
Power and ground pads—MPU (2/3 of total pads)	1,536	1,536	2,018	2,018	2,018	2,018	2,018
Total pads—ASIC high-performance	1,400	1,800	2,200	2,600	3,000	3,400	3,800
Signal I/O pads—ASIC high-performance (½ of total pads)	700	900	1,100	1,300	1,500	1,700	1,900
Power and ground pads—ASIC high-performance (½ of total pads)	700	900	1,100	1,300	1,500	1,700	1,900
Chip-to-package pads (Peripheral)	368	397	429	464	501	541	584
Number of Total Package Pins/Balls—Maximum							
Microprocessor/controller, cost-performance	740	821	912	1,012	1,123	1,247	1,384
ASIC (high-performance)	1,600	1,792	2,007	2,248	2,518	2,820	3,158

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

YEAR / *TECHNOLOGY NODE*	*2008* / *70 nm*	*2011* / *50 nm*	*2014* / *35 nm*
Number of Chip I/Os (Number of Total Chip Pads)—Maximum			
Total pads—MPU	3,840	4,224	4,416
Signal I/O pads—MPU (1/3 of total pads)	1,280	1,408	1,472
Power and ground pads—MPU (2/3 of total pads)	2,560	2,816	2,944
Total pads—ASIC high-performance	4,600	5,400	6,000
Signal I/O pads—ASIC high-performance (½ of total pads)	2,300	2,700	3,000
Power and ground pads—ASIC high-performance (½ of total pads)	2,300	2,700	3,000
Chip-to-package pads (Peripheral)	736	927	1,167
Number of Total Package Pins/Balls—Maximum			
Microprocessor/controller, cost-performance	1,893	2,589	3,541
ASIC (high-performance)	4,437	6,234	8,758

Package pin-count (Table 3) and cost-per-pin (Table 4), provided by the Assembly and Package TWG, 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 11%, 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 (1.11 cost/pin × .95 pins/year = 1.05 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× functionality/chip at flat price every two years = 29%/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 11% 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 seemingly insatiable market demand for higher-performance, cost-effective products. Just as "Moore's Law" predicts that functions-per-chip will double every 1.5 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 increasing average chip size will continue to demand the development of novel process, design, and packaging techniques.

These considerations are reflected in Table 4, which includes line items contributed by the Design TWG to forecast the multiple categories of 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 (across-chip clock) tends to become progressively larger in the future 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 (κ~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, and Frequency—Near Term Years

YEAR TECHNOLOGY NODE	1999 180 nm	2000	2001	2002 130 nm	2003	2004	2005 100 nm
Chip Pad Pitch (micron)							
Pad pitch—ball bond	50	48	47	45	43	42	40
Pad pitch—wedge bond	45	43	42	40	39	38	35
Pad pitch—area array	200	200	200	200	182	165	150
Cost-Per-Pin							
Package cost (cents/pin) (cost-performance)—maximum	1.90	1.81	1.71	1.63	1.55	1.47	1.40
Package cost (cents/pin) (cost-performance)—minimum	0.90	0.86	0.81	0.77	0.73	0.70	0.66
Package cost (cents/pin) (Memory)—maximum	1.90	1.71	1.54	1.39	1.25	1.12	1.01
Package cost (cents/pin) (Memory)—minimum	0.40	0.38	0.36	0.34	0.33	0.31	0.29
Chip Frequency (MHz)							
On-chip local clock, (high-performance)	1,250	1,486	1,767	2,100	2,490	2,952	3,500
On-chip, across-chip clock (high-performance)	1,200	1,321	1,454	1,600	1,724	1,857	2,000
On-chip, across-chip clock, high-performance ASIC	500	559	626	700	761	828	900
On-chip, across-chip clock (cost-performance)	600	660	727	800	890	989	1,100
Chip-to-board (off-chip) speed (high-performance, reduced-width, multiplexed bus)	1,200	1,321	1,454	1,600	1,724	1,857	2,000
Chip-to-board (off-chip) speed (high-performance, for peripheral buses)	480	589	722	885	932	982	1,035
Maximum number wiring levels—maximum	7	7	7	8	8	8	9
Maximum number wiring levels—minimum	6	6	7	7	8	8	8

Table 4b Performance and Package Chips: Pads, Cost, and Frequency—Long Term Years

YEAR TECHNOLOGY NODE	2008 70 nm	2011 50 nm	2014 35 nm
Chip Pad Pitch (micron)			
Pad pitch—ball bond	40	40	40
Pad Pitch—wedge bond	35	35	35
Pad Pitch—area array	150	150	150
Cost-Per-Pin			
Package cost (cents/pin) (cost-performance)—maximum	1.20	1.03	0.88
Package cost (cents/pin) (cost-performance)—minimum	0.57	0.49	0.42
Package cost (cents/pin) (memory)—maximum	0.74	0.54	0.39
Package cost (cents/pin) (memory)—minimum	0.25	0.22	0.19
Chip Frequency (MHz)			
On-chip local clock, (high-performance)	6,000	10,000	13,500
On-chip, across-chip clock (high-performance)	2,500	3,000	3,600
On-chip, across-chip clock (high-performance ASIC)	1,200	1,500	1,800
On-chip, across-chip clock (cost-performance)	1,400	1,800	2,200
Chip-to-board (off-chip) speed (high-performance, reduced-width, multiplexed bus)	2,500	3,000	3,600
Chip-to-board (off-chip) speed (high-performance, for peripheral buses)	1,285	1,540	1,800
Maximum number wiring levels—maximum	9	10	10
Maximum number wiring levels—minimum	9	9	10

ELECTRICAL DEFECT DENSITY

The latest targets for electrical defect density of DRAM, MPU, and ASIC (necessary to achieve 65–85% chip yield in the year of volume production) are shown in Table 5. 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 and 2 for DRAM and microprocessors. The maximum chip area of ASIC products is assumed equal to the maximum available field size of the exposure tool. In addition, the data in the table are now reported at the production life-cycle point. Other defect densities may be calculated at different chip sizes at the same technology node by using the formula found in the *Defect Reduction* 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 TECHNOLOGY NODE	1999 180 nm	2000	2001	2002 130 nm	2003	2004	2005 100 nm
Defect Reduction							
DRAM at production electrical D_0 chip size at 85% yield (d/m^2) §	1,249	1,193	1,140	1,089	1,040	994	950
*MPU at ramp electrical D_0 chip size at 75% yield (d/m^2) ****	1,742	1,742	1,742	1,552	1,383	1,321	1,262
ASIC first year electrical D_0 at 65% yield (d/m^2)	562	562	562	562	562	562	562
Minimum, mask count—maximum	24	24	24	24	25	25	26
Minimum, mask count—minimum	22	23	23	24	24	24	24

Table 5b Electrical Defects—Long Term Years

YEAR TECHNOLOGY NODE	2008 70 nm	2011 50 nm	2014 35 nm
Defect Reduction			
DRAM at production electrical D_0 chip size at 85% yield (d/m^2) §	828	723	630
*MPU at ramp electrical D_0 chip size at 75% yield (d/m^2) ****	1,101	960	837
ASIC first year electrical D_0 at 65% yield (d/m^2)	562	562	562
Minimum, mask count—maximum	28	28	30
Minimum, mask count—minimum	26	28	29

D_0—defect density

§ *DRAM Model—Generations 4× bits/chip every four years with interim 2× bits/chip generations; InTER-generation chip size growth rate model is 1.2× every four years; InTRA-generation chip size shrink model is 0.5× every three years beginning 1999.*

*** *MPU Chip Size Model—Both the cost-performance and high-performance MPUs target for InTER-generation chip size growth rate model is flat through 2001, then 1.2× growth every four years after 2001; InTRA-generation chip size shrink model is 0.5× every two years through 2001, then 0.5× every three years after 2001.*

POWER SUPPLY AND POWER DISSIPATION

Reduction of power supply voltage is driven by several factors: reduction of power dissipation, reduced transistor channel length, and reliability of gate dielectrics. The way in which the value of the power supply voltage is represented in the 1999 ITRS is the same as that used in the 1997 Roadmap. As seen in Table 5, 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 by 2011 are still considered viable, and the target to go to .3 Volts by 2014 has been added to the Roadmap.

Maximum power trends (e.g., for MPUs) are still presented in the two categories: 1) high-performance desktop applications, for which a heat sink on the package is permitted, and 2) portable battery operations. In both cases, total power consumption continues to increase, despite the use of a lower supply voltage. The increased power consumption is driven by higher operating frequency and the higher overall capacitance and resistance of larger chips with more on-chip functions.

Table 6a Power Supply and Power Dissipation—Near Term Years

YEAR *TECHNOLOGY NODE*	*1999* *180 nm*	*2000*	*2001*	*2002* *130 nm*	*2003*	*2004*	*2005* *100 nm*
Power Supply Voltage (V)							
Minimum logic V_{dd} (V)—maximum (for maximum performance)	1.8	1.8	1.5	1.5	1.5	1.2	1.2
Minimum logic V_{dd} (V)—minimum (for lowest power))	1.5	1.5	1.2	1.2	1.2	0.9	0.9
Maximum Power							
High-performance with heatsink (W)	90	100	115	130	140	150	160
Battery (W)—(hand-held)	1.4	1.6	1.7	2.0	2.1	2.3	2.4

Table 6b Power Supply and Power Dissipation—Long Term Years

YEAR *TECHNOLOGY NODE*	*2008* *70 nm*	*2011* *50 nm*	*2014* *35 nm*
Power Supply Voltage (V)			
Minimum logic V_{dd} (V)—maximum (for maximum performance)	0.9	0.6	0.60
Minimum logic V_{dd} (V)—minimum (for lowest power))	0.6	0.5	0.30
Maximum Power			
High-performance with heatsink (W)	170	174	183
Battery (W)—(hand-held)	2.0	2.2	2.4

COST

Table 7 is 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 delivering twice the functionality on-chip every 1.5 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 four times as many functions per chip every three years with only a moderate 1.4× increase in chip size and 1.4× in cost (approximately constant cost per cm^2 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. 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 participants have proposed a new model for achieving the required reduction: provide the customer twice the functionality every two years at a constant cost per chip. This new model 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 new model has been used to set the trends for the affordable cost/bit and cost/transistor for DRAM and microprocessors, respectively. These cost targets, together with the technical targets, should guide the activity of the engineering community in planning and executing each technology-development program. One illustration of the new model, shown in Table 7, is that DRAM suppliers, to keep maximum average selling prices (ASPs) constant between (inter) generations, must introduce generations with 2× on-chip bits every two years and set targets to decrease the cost/bit at an average rate of 29%/year.

Extrapolation of historical trends would indicate an "at introduction" affordable cost/bit of approximately 42 microcents for 1-Gbit DRAMs in 1999. In addition, the historical trends indicate that, within a DRAM generation, a 45%/year reduction in cost/bit should be expected.[1] A corresponding analysis conducted from published data for microprocessors yields similar results.[2] In the case of MPUs, generations are now also targeted for 2× functionality (transistors) increase every two years. The 29%/year target for reduction in affordable cost/transistor from generation to generation is also being used in this case, along with the 45%/year reduction rate within the same generation.

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 3% and 9% per year (Table 7), the number of pins grows at 11%/year (Table 3). 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 1999 International Technology Roadmap for Semiconductors. Further discussion is detailed in the *Test* chapter.

---

[1] 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.

[2] 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 TECHNOLOGY NODE	1999 180 nm	2000	2001	2002 130 nm	2003	2004	2005 100 nm
Affordable Cost per Function ++							
DRAM cost/bit at (packaged microcents) at samples/introduction	42	—	21	—	11	—	5.3
DRAM cost/bit at (packaged microcents) at production §	15	—	7.6	—	3.8	—	1.9
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at introduction ***	1,735	—	868	—	434	—	217
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at ramp ***	1,050	—	525	—	262	—	131
High-performance MPU (microcents/transistor) (including on-chip SRAM) at ramp ***	245	—	123	—	61	—	31
Cost-Per-Pin (see Table 4)	—	—	—	—	—	—	—
Test							
Volume tester cost per high-frequency signal pin ($K/pin) (high-performance ASIC)—maximum	8	7	7	6	6	5	5
Volume tester cost per high-frequency signal pin ($K/pin) (high-performance ASIC)—minimum	4	3	3	3	3	2	2
Volume tester cost/pin ($K/pin) (cost-performance MPU)	8	8	7	7	6	6	5

Table 7b Cost—Long Term Years

YEAR TECHNOLOGY NODE	2008 70 nm	2011 50 nm	2014 35 nm
Affordable Cost per Function ++			
DRAM cost/bit (packaged microcents) at samples/introduction	—	0.66	—
DRAM cost/bit (packaged microcents) at production §	—	0.24	—
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at introduction ***	—	27	—
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at ramp ***	—	16	—
High-performance MPU (microcents/transistor) (including on-chip SRAM) at ramp ***	—	3.8	—
Cost-Per-Pin (see Table 4)	—	—	—
Test			
Volume tester cost per high-frequency signal pin ($K/pin) (high-performance ASIC)—maximum	5	5	5
Volume tester cost per high-frequency signal pin ($K/pin) (high-performance ASIC)—minimum	N/A	N/A	N/A
Volume tester cost/pin ($K/pin) (cost-performance MPU)	4	2	2

++ *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 7–8 years after introduction; MPU unit volume life-cycle peak occurs typically after four years, when the next generation processor enters its ramp phase (typically two years after introduction).*

§ *DRAM Model—Generations 4× bits/chip every four years with interim 2× bits/chip generations; InTER-generation chip size growth rate model is 1.2× every four years; InTRA-generation chip size shrink model is 0.5× every three years beginning 1999.*

*** *MPU Chip Size Model—Both the cost-performance and high-performance MPUs target for InTER-generation chip size growth rate model is flat through 2001, then 1.2× growth every four years after 2001; InTRA-generation chip size shrink model is 0.5× every two years through 2001, then 0.5× every three years after 2001.*

GLOSSARY

KEY OVERALL ROADMAP TECHNOLOGY CHARACTERISTICS (ORTC) TERMINOLOGY (WITH OBSERVATIONS AND ANALYSIS)

CHARACTERISTICS OF MAJOR MARKETS

TECHNOLOGY NODE (nm)—The ground rules of process governed by the smallest feature printed. The half-pitch of first-level interconnect dense lines is most representative of the DRAM technology level required for the smallest economical chip size. For logic, such as microprocessors (MPUs), gate length is most representative of the leading-edge technology level required for maximum performance. MPU and ASIC logic interconnect half-pitch processing requirements typically lag behind DRAM half-pitch. For cost reasons, high-volume, low-cost ASIC gate-length requirements will typically match DRAM half-pitch targets, but the low-volume leading-edge high-performance ASIC gate-length requirements will track closely with MPUs.

"MOORE'S LAW"—An historical observation by Intel executive, Gordon Moore, that the market demand (and semiconductor industry response) for functionality per chip (bits, transistors) doubles every 1.5 to 2 years. He also observed that MPU performance [clock frequency (MHz) × instructions per clock = millions of instructions per second (MIPS)] also doubles every 1.5 to 2 years. Although viewed by some as a "self-fulfilling" prophecy, "Moore's Law" has been a consistent macro trend, and key indicator of successful leading-edge semiconductor products and companies, for the past 30 years.

"COST-PER-FUNCTION" MANUFACTURING PRODUCTIVITY IMPROVEMENT DRIVER—In addition to "Moore's Law", there is a historically-based "corollary" to the "law," which suggests that, to be competitive, manufacturing productivity improvements must also enable the cost-per-function (microcents per bit or transistor) to decrease by -29% per year. Historically, when functionality doubled every 1.5 years, then cost-per-chip (packaged unit) could double every six years and still meet the cost-per-function reduction requirement. If functionality doubles only every two years, as suggested by consensus DRAM and MPU models of the 1999 ITRS, then the manufacturing cost per chip (packaged unit) must remain flat.

"Affordable" Packaged Unit Cost/Function—Final cost in microcents of the cost of a tested and packaged chip divided by ***Functions/Chip.*** Affordable costs are calculated from historical trends of affordable average selling prices [gross annual revenues of a specific product generation divided by the annual unit shipments] less an estimated gross profit margin of approximately 35% for DRAMs and 60% for MPUs. The affordability per function is a guideline of future market "tops-down" needs, and as such, was generated independently from the chip size and function density. Affordability requirements are expected to be achieved through combinations of—1) smaller chip sizes from technology and design improvements; 2) increasing wafer diameters; 3) decreasing equipment cost-of-ownership (CoO); 4) increasing equipment overall equipment effectiveness; 5) reduced package and test costs; 6) improved design tool productivity; and 7) enhanced product architecture and integration.

DRAM Generation* at *(product generation life-cycle level)—The anticipated bits/chip of the DRAM product generation introduced in a given year, manufacturing technology capability, and life-cycle maturity (Demonstration, Introduction, Sample, Production, Ramp, Peak).

MPU Generation* at (*product generation life-cycle level)—The generic processor generation identifier for the anticipated Microprocessor Unit (MPU) product generation functionality (logic plus SRAM transistors per chip) introduced in a given year, manufacturing technology capability, and life-cycle maturity (Introduction, Ramp, Peak, Embedded).

Cost-Performance MPU—MPU product optimized for lowest cost by minimizing on-chip SRAM to level-one (L1) cache only (32Kbytes/1999). Logic functionality and L1 cache typically double every 2-year generation. This typically has a 6-year (introduction plus ramp plus peak) computer-market-application life-cycle before being replaced by the next generation cost-performance MPU, then continues on in embedded applications.

High-performance MPU—MPU product optimized for maximum system performance by using a shrunk cost-performance ramp-level MPU core combined with a large (2Mbyte/1999) level-two (L2) SRAM. Logic functionality and L2 cache typically double every two-year generation. Typically has only a 4-year (ramp and peak) life cycle in the relatively low-volume, higher-priced, high-performance computer market. There is no classic "embedded" application for the high-SRAM-content MPU, but that may change as future market demand develops for multiple-MPU-per box internet server and communication processor applications emerge. Those applications will provide increased demand for more cost-effective inTRA generation shrinks of the high-performance MPU, thus extending the life-cycles of future generations.

PRODUCT INTER-GENERATION—Product generation-to-generation targets for periodically increasing on-chip functionality and allowable chip size. The targets are set to maintain "Moore's Law," while preserving economical manufacturability. The 1999 ITRS consensus target for the rate of increase of DRAM and MPU inTER-generation functionality is 2×/chip every two years. The allowable inTER-generational chip size growth for DRAMs is 1.2× every four years. For MPUs, the allowable chip size growth is flat through 2001, then grows at 1.2× every four years. To add only 20% in area every four years, while quadrupling functionality, requires an inTER-generation design productivity which further reduces chip size by an additional minus 7–8 % per year. This design-related productivity reduction is in addition to the basic lithography-provided area reduction of -11% per year.

PRODUCT INTRA-GENERATION—Within a given product generation. The consensus-based targets reduce chip size (by shrinks and "cut-downs") utilizing the latest available manufacturing and design technology at every point through the roadmap. The ITRS consensus targets for both DRAM and MPU reduce chip size within a generation by minus 50% per technology node. For DRAM, this reduction of minus 50% occurs every three years, or minus 37% every two years. For MPU, the 50% reduction occurs every two years through 2001, then slows to minus 37% per two years (same as DRAM).

YEAR OF DEMONSTRATION—Year in which the leading chip manufacturer supplies an operational sample of a product as a demonstration of design and/or technology node processing feasibility and prowess. A typical venue for the demonstration is a major semiconductor industry conference, such as the International Solid State Circuits Conference (ISSCC) held by the Institute of Electrical and Electronic Engineers (IEEE). Demonstration samples are typically manufactured with early development or demonstration- level manufacturing tools and processes. Historically, DRAM products have been demonstrated at 4× bits-per-chip every three years at the leading-edge process technology node, typically 2–3 years in advance of actual market introduction. DRAM demonstration chip sizes have doubled every six years, requiring an increasing number of shrinks and delay before market introduction is economically feasible. Frequently, chip sizes are larger than the field sizes available from lithography equipment, and must be "stitched" together via multiple-exposure techniques that are feasible only for very small quantities of laboratory samples. Example: 1997/ISSCC/1Gb DRAM.

YEAR OF INTRODUCTION (DRAM)—Year in which the leading chip manufacturer supplies small quantities of engineering samples (<1K). These are provided to key customers for early evaluation, and are manufactured with qualified production tooling and processes. To balance market timeliness and economical manufacturing, DRAM products will be introduced at 2× functionality per chip every two years. In addition, manufacturers will delay introduction until a chip-size shrink or "cut-down" level is achieved which limits the inTER-generation chip-size growth to 1.2× every 4 years, or approximately 1.1× every 2-year generation. Example: 1999/1Gb DRAM.

YEAR OF SAMPLE (DRAM)—Year in which the leading chip manufacturers begin to supply volume quantities of qualification samples (10K–100K) of dynamic random access memories (DRAMs) that are manufactured with qualified production tooling and processes. The leading-edge DRAM production sample products are typically half-size-generation products which are "cut down" from the introduction-level generation design and manufacturing processes. Example: 1999/512Mb DRAM.

YEAR OF PRODUCTION (DRAM)—Year in which leading chip manufacturers begin shipping volume quantities (millions per month) of product manufactured with qualified production tooling and processes. This product typically contains one-fourth (1/4) the bits per chip of the introduction-level generation design, from which it is "cut-down." Example: 1999/256Mb DRAM.

YEAR OF RAMP (DRAM)—Year in which leading chip manufacturers begin rapid expansion of production capacity, shipping high volumes (16–80 million per month) quantities of product manufactured using production tooling and processes which are being quickly "copied" into multiple modules of manufacturing capacity. Price per bit has "crossed over" the previous generation, which experiences a unit volume demand "peak" as it is replaced in the market. The ramp-level product generation is also "cut down" from the introduction-level generation design and manufacturing processes. Example: 1999/128Mb DRAM.

YEAR OF PEAK (DRAM)—Year in which the highest-volume DRAM generation is replaced by a next-generation product which is more cost-effective on a per-bit basis. After this cost-per-bit "cross-over" occurs volumes cease to grow, and then quickly drop, as the generation reaches the end of the life-cycle begun years earlier at its own introduction level. Typically this "peak" product generation will also be "cut-down" out of the present introduction-level generation design and process to minimize cost and maximize availability. However some portion of "peak" product capacity might also remain at the "ramp" chip size and technology to balance utilization of existing mature production tooling and process capacity. Example: 1999/64Mb DRAM.

YEAR OF INTRODUCTION (MPU)—Year in which the leading chip manufacturer supplies small quantities of engineering samples (<1K). These are provided to key customers for early evaluation, and are manufactured with qualified production tooling and processes. The introduction cost-performance MPU may be combined in a multi-chip module, along with L2 cache, in low-volume computer applications which demand high performance.

YEAR OF RAMP (MPU)—The year in which leading chip manufacturers begin rapid expansion of production capacity, shipping high volumes (2–10 million per month) quantities of cost-performance MPU product manufactured using production tooling and processes. As demand increases, the tooling and processes are being quickly "copied" into multiple modules of manufacturing capacity. Lower-volume, high-performance MPUs are also ramping concurrently as its co-existing cost-performance MPU core, but the L2 cache is now included on-chip but with twice the memory as its high-performance-generation predecessor.

YEAR OF PEAK (MPU)—An MPU generation experiences quicker and sharper unit volume demand "peak" as it is rapidly replaced in its primary computer market applications. High-performance MPUs peak in the same year as the cost-performance MPU that is used as the high-performance MPU core. Both cost-performance and high-performance MPU have shorter life-cycles than DRAM, and each is replaced rapidly by its respective higher functionality/performance inTER-generation MPU.

YEAR OF EMBEDDED (MPU)—Unlike DRAM products, the ramp-level MPU cost-performance product generation is not "cut down" from the overlapping introduction-level generation, and therefore cannot easily coexist in the same computer markets. Therefore, an MPU generation which is past the peak life-cycle stage will move into high-volume, but very low-cost, "embedded" applications as it shrinks to cost-effective levels.

Functions/Chip—The number of bits (DRAMs) or logic transistors (MPUs, application-specific integrated circuits [ASICs]) that can be cost-effectively manufactured on a single monolithic chip at the available technology level. Logic functionality (transistors per chip) include both SRAM and logic transistors. DRAM functionality (bits per chip) is based only on the bits (after repair) on a single monolithic chip.

Chip Size (mm^2)—The typical area of the monolithic memory and logic chip that can be affordably manufactured in a given year based upon the best available leading-edge design and manufacturing process. (Estimates are projected based upon historical data trends and the ITRS analyst consensus models).

Functions/cm²—The density of functions in a given square centimeter = ***Functions/Chip*** on a single monolithic chip divided by the ***Chip Size***. This is an average of the density of all of the functionality on the chip, including pad area and wafer scribe area. In the case of DRAM, it includes the average of the high-density cell array and the less-dense peripheral drive circuitry. In the case of the MPU products, it includes the average of the high-density SRAM and the less-dense random logic. In the case of ASIC, it will include high-density embedded memory arrays, averaged with less dense array logic gates and functional cores. Most typical ASIC designs will be slightly less dense than the high-performance MPUs, which are mostly SRAM.

DRAM Cell Array Area Percentage—The maximum practical percentage of the total DRAM chip area that the cell array can occupy at the various stages of the generation life cycle. At the introduction chip size targets, this percentage must be typically less than 70% to allow space for the peripheral circuitry, pads, and wafer scribe area. Since the pads and scribe area do not scale with lithography, the maximum cell array percentage is reduced in other inTRA-generation shrink levels (typically less than 55% at the production level, and less than 50% at the ramp level).

DRAM Cell Area (µm²)—The measure of the maximum allowable DRAM memory bit cell area specified by the requirement to meet the target chip size and cell array area percentage requirements. May also be expressed as the cell area factor – number of equivalent units of area of a square of the DRAM half-pitch. Minimizing the area for the cell is in conflict with the desire to maximize the capacitance storage capability of the continuously shrinking cell. This creates a conflict between the technical feasibility of the cell area required to meet the economic constraints of the maximum allowable chip size.

DRAM Cell Area Factor—The measure of the maximum allowable DRAM memory bit cell area, expressed as the number of equivalent units of area of a square of the DRAM half-pitch.
Example: 1999: square of the half-pitch = $(180\ nm)^2 = .032\ \mu m^2$; maximum cell area for 1Gb DRAM to be < 70% of total chip area = $0.26\ \mu m^2$; therefore, the maximum cell area factor = 0.26/0.32 = 8. The cell factor is also often expressed by equivalent aspect ratios of the half-pitch units (2×4 = 8, 2×3 = 6, 2×2=4, 1.6×1.6=2.5, etc.).

Usable Transistors/cm² (High-performance ASIC, Auto Layout)—Number of transistors per cm^2 designed by automated layout tools for highly differentiated applications produced in low volumes. High-performance, leading-edge, embedded-array ASICs include both on-chip array logic cells, as well as dense functional cells (MPU, I/O, SRAM, etc). Density calculations include the connected (useable) transistors of the array logic cells, in addition to all of the transistors in the dense functional cells. The largest high-performance ASIC designs will fill the available production lithography field.

CHIP AND PACKAGE—PHYSICAL AND ELECTRICAL ATTRIBUTES

Number of Chip I/Os – Total (Array) Pads—The maximum number of chip signal I/O pads plus power and ground pads permanently connected to package plane for functional or test purposes, or to provide power/ground contacts (including signal conditioning). These include any direct chip-to-chip interconnections or direct chip attach connections to the board (Package plane is defined as any interconnect plane, leadframe, or other wiring technology inside a package, i.e., any wiring that is not on the chip or on the board.). MPUs typically have a ratio of signal I/O pads to power/ground pads of 1:2, whereas the high-performance ASIC ratio is typically 1:1.

Number of Chip I/Os – Total (Peripheral) Pads—The maximum number of chip signal I/O plus power and ground pads for products with contacts only around the edge of a chip.

Pad Pitch—The distance, center-to-center, between pads, whether on the peripheral edge of a chip, or in an array of pads across the chip.

Number of Package Pins/Balls—The number of pins or solder balls presented by the package for connection to the board (may be fewer than the number of chip-to-package pads because of internal power and ground planes on the package plane or multiple chips per package).

Package cost (cost-performance)—Cost of package envelope and external I/O connections (pins/balls) in cents/pin.

Chip Frequency (MHz)

On-chip, local clock, high-performance—On-chip clock frequency of high-performance, lower volume microprocessors in localized portions of the chip.

On-chip, across-chip clock—On-chip clock frequency of microprocessors and ASICs for interconnect signals that run across the full width of the chip (Typically, this is lower than the localized clock performance due to capacitance loading of the long cross-chip interconnect.).

Chip-to-board (off-chip) speed (high-performance, reduced-width, multiplexed bus)—Maximum signal I/O frequency to specialized board reduced-width, multiplexed buses.

Chip-to-board (off-chip) speed (high-performance, peripheral buses)—Maximum signal I/O frequency to board peripheral buses of high and low volume logic devices.

Other Attributes

Lithographic Field Size (mm^2)—Maximum single step or step-and-scan exposure area of a lithographic tool at the given technology node.

Maximum Number Of Wiring Levels—On-chip interconnect levels including local interconnect, local and global routing, power and ground connections, and clock distribution.

FABRICATION ATTRIBUTES AND METHODS

Electrical D_0 Defect Density (d/m^{-2})—Number of electrically significant defects per square meter at the given technology node, production life-cycle year, and target probe yield.

Minimum Mask Count—Number of masking levels for mature production process flow with maximum wiring level (Logic).

MAXIMUM SUBSTRATE DIAMETER (MM)

Bulk or Epitaxial or Silicon-on-Insulator Wafer—Silicon wafer diameter used in volume quantities by mainstream IC suppliers. The ITRS timing targets, contributed by the Factory Integration Technology Working Group, are based on the first 20K wafer-starts-per-month manufacturing facility, versus the first-pilot-line timing target of the 1997 NTRS.

ELECTRICAL DESIGN AND TEST METRICS

Power Supply Voltage (V)

Minimum Logic V_{dd}— Nominal operating voltage of chips from power source for operation at design requirements.

Maximum Power

High-performance with heat sink (W)—Maximum total power dissipated in high-performance chips with an external heat sink.

Battery (W)—Maximum total power/chip dissipated in battery operated chips.

DESIGN AND TEST

Volume Tester Cost/Pin ($K/pin)—Cost of functional (chip sort) test in high volume applications divided by number of package pins.