INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

2003 EDITION

EXECUTIVE SUMMARY

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

Jointly Sponsored
by

European Semiconductor Industry Association

Japan Electronics and Information Technology Industries Association

Korea Semiconductor Industry Association

Taiwan Semiconductor Industry Association

Semiconductor Industry Association

ACKNOWLEDGMENTS

INTERNATIONAL ROADMAP COMMITTEE

❖*Europe*—Wolfgang Arden, Joop Bruines, Gerhard Goltz
❖*Japan*—Toshitaka Fukushima, Toshiaki Masuhara, Toshihiko Osada, Makoto Yoshimi
❖*Korea*—HoKyu Kang, YoungJin Park, JooTae Moon, HanKu Sho, Hyunchl Sohn
❖*Taiwan*—Chengming Hu, Chen-Hsi Lin, C.Y. Lu, Water Lur, Jack Sun, Ping Yang
❖*U.S.A.*—Bob Doering, Paolo Gargini, Juri Matisoo

TECHNOLOGY WORKING GROUP KEY CONTRIBUTORS

Chip-Size/Technology Node Study Group—Alan Allan, Chris Case, Chi-Shih Chang, Walter Class, Bob Doering, Donald Edenfeld, Denis Fandel, David Jensen, Andrew Kahng, Fred Lakani, Harry Levinson, Chris Long, Jeff Pettinato, Shizu Sawada, Luan C. Tran, Makoto Yoshimi, Peter Zeitzoff

System Drivers and Design—Lawrence Arledge, Yoshimi Asada, Tom Bednar, Kerry Bernstein, Ralf Brederlow, Juan-Antonio Carballo, Liang-Gee Chen, John Cohn, Don Cottrell, John Darringer, Narbeh Der Hacobian, Yoshiharu Furui, Shinji Furuichi, Shinji Furuno, Hideyuki Hamada, Hiroyuki Hara, Justin E. Harlow III, David Hathaway, Tamotsu Hiwatashi, Masaharu Imai, Koichiro Ishibashi, William H. Joyner, Jr., Andrew B. Kahng, Masaru Kakimoto, Masahito Kanaya, Haruhisa Kashiwagi, Yoshihiko Kinoshita, Tokinori Kozawa, Don MacMillen, Grant Martin, Joseph Morrell, Swamy V. Muddu, Ed Nowak, Tomoji Nukiyama, Nobuhiro Okano, Nobuto Ono, Shinji Ono, Bill Read, Minoru Saitoh, Prashant Saxena, Jean-Pierre Schoellkopf, Tetsuyoshi Shiota, Gary Smith, Tsutomu Someya, Leon Stok, Dennis Sylvester, Hirofumi Taguchi, Tah-Kang Ting, Hiroki Tomoshige, Kunio Uchiyama, Takuzo Uemura, Maarten Vertregt, Akihiro Yamada, Ichiro Yamamoto, Takayuki Yamanouchi, Yervant Zorian, Paul Zuchowski

Test—Takashi Aikyo, Rob Aitken, Kenichi Anzou, Davide Appello, Roger Barth, Jay Bedsole, Laroussi Bouzaida, Carl Buck, Phil Burlison, Yi Cai, Brett Casey, Jwu E. Chen, Chip Cotton, Gordon Cowan, Herve Deshayes, Dennis Eaton, Don Edenfeld, Kouichi Eguchi, Stefan Eichenberger, John Ferrario, Ichiro Fujishiro, Satoshi Fukumoto, Rudy Garcia, Anne Gattiker, Atul Goel, Mike Green, John Hartstein, Kazumi Hatayama, Ken Heiman, Camelia Hora, Rafiq Hussain, Toshinori Inoshita, Sunil Jain, John Johnson, Rich Karr, Bernd Koenemann, Bernd Laquai, Michael Lee, Marc Loranger, Maurice Lousberg, Ted Lundquist, Bob Madge, John Matthias, Peter Maxwell, Dick McClelland, Shinji Mori, Peter Muhmenthaler, Udaya Natarajan, Paul Nesrsta, Phil Nigh, Yasumasa Nishimura, Dave Noddin, Mitsuyasu Ohta, Toshinobu Ono, Akira Ooishi, Bill Ortner, Bill Price, Rochit Rajsuman, Jim Rhodes, Paul Roddy, Mike Rodgers, Charles Ross, Masayuki Sato, Lynn Schmidt, Dan Sech, Rene Segers, Lee Song, Chauchin Su, Fred Taber, Hidefumi Toda, Bob Totorica, Shih Tseng, Masanori Ushikubo, Dave Vallett, Jody Van Horn, Don Van Overloop, Larry Wagner, Hank Walker, Hsin Po Wang, Dan Weinstein, Burnie West, Don Wheater, Tom Williams, Chad Wren, Wen Ching Wu, Osamu Yamada, Bob Zacharis, Yervant Zorian

Process Integration, Devices, and Structures—David Blackburn, Joe E. Brewer, Joop Bruines, James Chung, Simon Deleonibus, Theodore A. (Ted) Dellin, Carlos Diaz, J.C. Guo, Toshiro Hiramoto, Digh Hisamoto, Atsushi Hori, Tadahiko Horiuchi, Margaret Huang, Jim Hutchby, Jiro Ida, Kiyotaka Imai, Yasuo Inoue, Naoki Kasai, Tzu-Jae King, Rich Liu, Naoki Nagashima, Takashi Nakamura, Tak H. Ning, Mark Rodder, Shizuo Sawada, Klaus Schruefer, Yee-Chaung See, Kentaro Shibahara, Thomas Skotnicki, Shoji Sudo, Toshihiro Sugii, Yoshitaka Tadaki, Shuichi Tahara, Scott Thompson, Luan C. Tran, Jason Woo, Qi Xiang, Geoffrey Yeap, Peter Zeitzoff, Bin Zhao

Radio Frequency and Analog/Mixed-signal Technologies for Wireless Communications—Herbert S. Bennett, Ralf Brederlow, Julio Costa, Peter E. Cottrell, Ding Day, Ronald Grundbacher, Ali Hajimiri, Dave Halchin, W. Margaret Huang, James A. Hutchby, Anthony A. Immorlica, Rik Jos, Tom Kazior, Minh Le, Jan-Erik Mueller, Ali M. Niknejad, Ernesto Perea, Marco Racanelli, Jeff Shealy, Hisashi (Sam) Shichijo, Roger Van Art, Chuck Weitzel, Bin Zhao, John Zolper

Emerging Research Devices—George Bourinaoff, Joe Brewer, Joop Bruines, John Carruthers, Jim Chung, Ramon Compano, Kristin De Meyer, Simon Deleonibus, Peng Fang, Mike Forshaw, Byong Gook Park, Dae Gwan Kang, Toshiro Hiramoto, Jim Hutchby, Tsu-Jae King, Tak Ning, Kentaro Shibahara, Thomas Skotnicki, Luan C. Tran, Rainer Waser, Philip Wong, In Yoo, Makoto Yoshimi, Yuegang Zhang, Victor Zhirnov

Front End Processes—Amir Al-bayati, Mauro Alessandri, Greg Allen, Michael Alles, Olli Anttila, Luis Aparicio, Leo Archer, Mayank Balsara, Sanjay Banerjee, Joel Barnett, Twan Bearda, Larry Beckwith, Enrico Bellandi, Ivan Berry, Roberto Bez, George Brown, W. Murray Bullis, Stephanie Butler, Jeff Butterbaugh, Emilio Camerlenghi, Paolo Caprara, George Celler, Juanita Chambers, S.C. Chen , Henry Cho, Phil Clark, Walter Class, Phil Crabtree, Michael Current, Charles J.J. Dachs, Tim Dalton, Suman Datta, Simon Deleonibus, Carlos Diaz, Anthony Dip, Roxanne Dulas, Yuri Erokhin, Laszlo Fabry, Sue Felch, Nobuo Fujiwara, Glenn Gale, Jeff Gelpey, Jeff Glick, Hans Gossmann, Alessandro Grossi, Dinesh Gupta, Martin Gutsche, Kevin Han, Zach Hatcher, Richard Hockett, Frank Holsteyns, Harry Hovel, Brad Howard, Howard Huff, Bill Hughes, Makarem Hussein, Shuji Ikeda, Sally Iwanski, Mike Jackson, Amitabh Jain, Craig Jasper, Ukyo Jeong, Bob Johnston, Erin C. Jones, Marilyn Jones, Mototaka Kamoshida, Kazuhiro Kaneda, Hi-Kyu Kang, Seiichiro Kawamura, Yung Kim, Brian Kirkpatrick, Hiroshi Kitajima, Tomohisa Kitano, Martin Knotter, Robert Kraft, Michitaka Kubota, Stephan Kudelka, Claudia Kupfer, Paul Langer, Larry Larson, Kam Lee, Didier Levy, Thomas Lii, Wen Lin, Jerry Liu, Shih Hsin Lo, Joern Luetzen, Christopher Lyons, Pat Lysaght, Don McCormack, Sandeep Mehta, Doug Mercer, Paul Mertens, Troy Messina, Doug Meyer, Fred Meyer, Michihiko Mifuji, Ichiro Mizushima, Alberto Modelli, Steve Moffatt, Stephane Monfray, Jim Moreland, Naim Moumen, Anthony Muscat, David Myers, Toshiro Nakanishi, Sadao Nakashima, Masaaki Niwa, Farad Nouri, Toshihide Ohgata, Carlton Osborn, Paul Packan, Gerd Pfeiffer, Francesco Pipia, Noel Poduje, Jagdish Prasad, Krishnaswamy Ramkumar, K.V. Ravi, Adriana Rebora, Rick Reidy, Karen Reinhardt, Leonard Reuben, Deborah Riley, Jae Sung Roh, Rob Rovito, Lenny Rubin, Ed Rutter, Devendra Sadana, Shizu Sawada, Uwe Schroeder, Tom Seidel, Kenneth Settlemeyer, Gregory Smith, Chris Sparks, Bob Standley, Bill Taylor, Jack Thomas, Carl Treadwell, H.H. Tsai, Hidetsugu Uchida, Steven Verhaverbeke, Eric Vogel, Bernd Vollmer, Rita Vos, Dennis Wagner, Peter Wagner, Mike Walden, Masaharu Watanabe, Neal Weaver, Gregory Wilson, Rick Wise, Kikuo Yoshikawa, Peter Zeitzoff

Lithography—Ted Fedynyshyn, Hiroshi Fukuda, Jan Willem Gemmink, Isamu Hanyu, Naoya Hayashi, Scott Hector, Yasuhiro Horiike, Koki Inoue, Rainer Kaesmaier, Masaomi Kameyama, Hokuto Kasahara, Yoshio Kawai, Beth Kells, Kurt Kimmel, Masanori Komuro, Kouichi Kuhara, Keishiro Kurihara, Harry Levinson, Burn J. Lin, Ichiro Mori, Susumu Mori, Shigeru Moriya, Makoto Nakase, Shinji Okazaki, Connie Reed, Norio Saito, Masaru Sasago, Kazuyuki Suko, Hiroaki Sumitani, Hiroki Tabuchi, Yoichi Takehana, Tsuneo Terasawa, Takayuki Uchiyama, Mauro Vasconi, Grant Wilson , Masaki Yamabe, Yuichi Yamada, Tadayuki Yamaguchi

Interconnect—Nobuo Aoi, Junichi Aoyama, Steven Avanzino, Hans-Joachim Barth, Ivan Berry, Christopher Case, Paul Feeney, Bob Geffken, Narishi Gonohe, Dirk Gravesteijn, Harold Hosack, Calvin Hsueh, Shin-Puu Jeng, Marilyn Jones, Makiko Kageyama, Hyeon-Deok Lee, J.D. Luttmer, Karen Maex, Mike Mills, Ken Monnig, Tomoji Nakamura, Yuichi Nakao, Shinichi Ogawa, Akihiko Osaki, Gary Ray, Karen Reinhardt, Hideki Shibata, Hyun Chul Sohn, Michele Stucchi, Shinichi Tanimoto, Manabu Tsujimura, Kazuyoshi Ueno, Jeff Wetzel, Osamu Yamazaki, C. H. Yu

Factory Integration—Junichi Akeuchi, Hiroyuki Akimori, Josef Bichlmeier, Ray Bunkofske, Jonathan Chang, Al Chasey, Allan Chen, Thomas Chen, Gilbert Chiang, Eric Christensen, Ron Denison, Klaus Eberhardt, Anne Elder, Neil Fisher, Len Foster, Masazumi Fukushima, Ashwin Ghatalia, Barbara Goldstein, Arieh Greenberg, Sven Hahn, Chung Soo Han, Yoshio Hayashi, Michio Honma, Yasutaka Horii, George Horn, Giichi Inoue, Junji Iwasaki, Mani Janakiram, Melvin Jung, Kazuhiro Kakizaki, Franklin Kalk, Atsuya Kano, Atsuhiko Kato, Shigeru Kobayashi, Shoichi Kodama, Ya-Shian Li, Don Martin, Dave Miller, Akira Mitsui, Ryouji Mori, Eckhard Müller, Seiya Nakagawa, Hideki Nakajima, Seiichi Nakazawa, Richard Oechsner, Shuzo Ohshio, Mikio Otani, Jeff Pettinato, Dev Pillai, Lisa Pivin, Scott Pugh, Adrian Pyke, Claus Schneider, Mike Schwartz, Steve Seall, Marlin Shopbell, Arnie Steinman, Dan Stevens, Junichi Takeuchi, Keisuke Tanimoto, Toshiyuki Uchino, K.R. Vadivazhagu, Joost van Herk, Phillippe Vialletelle, Tsutomu Wada, Harvey Wohlwend, Hiromi Yajima, Makoto Yamamoto, William W.L. Yuan

Assembly and Packaging—Joe Adam, Steve Adamson, Mark Bird, Bill Bottoms, Chi-Shih Chang, Bill Chen, Carl Chen, P.K. Chiang, Jack Fisher, Ed Fulcher, Sageo Hamano, George Harman, Shuya Haruguchi, Ryo Haruta, Ted Ho, Greg Hotchkiss, S.P. Hsu, Mahandevan Iyer, Hisao Kasuga, Michitaka Kimura, Rainer Kyburz, R.S. Lee, Shelton Lu, Jean-Pierre Moscicki, Wayne Mulholland, Alex Oscilowski, Bernd Roemer, Yuji Shirai, Michio Sono, Coen Tak, Takashi Takata, Rika Tanaka, Ho-Ming Tong, Shoji Uegaki, Henry Utsunomiya, Mike Varnau, Juergan Wolf

Environment, *Safety*, *and Health*—Hans-Peter Bipp, Shinya Hashimoto, Francesca Illuzzi, Jim Jewett, Shou-Nan Li, C.N. Lin, Ray Lin, Michael Mocella, Coleen (Miller) Regan, Dawn Speranza, Harry Thewissen, Walter Worth

Yield Enhancement—Makoto Akizuki, Mark Camenzind, Jeff Chapman, Dirk de Vries, John DeGenova, Hans-Martin Dudenhausen, Masakazu Ichikawa, Masahiko Ikeno, Masami Ikota, Sadayuki Imanishi, Rick Jarvis, Keith Kerwin, Hiroshi Kitajima, Isao Kojima, Sumio Kuwabara, Fred Lakhani, Christopher Long, James McAndrew, Mike McIntyre, Len Mei, Fumio Mizuno, Andreas Neuber, Kazuo Nishihagi, Andreas Nutsch, Hideo Ohta, Akira Okamoto, Toshihiko Osada, Joe O'Sullivan, Kevin Pate, Mike Patterson, Lothar Pfitzner, Ron Remke, Mike Retersdorf, Ralph Richardson, Eric Rouchouze, Val Stradzs, Ines Thurner, Ken Tobin, Dick Verkleij, Hank Walker, Tings Wang, Charles Weber, Yuichiro Yamazaki

Metrology—Makoto Akizuki, John Allgair, Chas Archie, Peter Borden, Dick Brundle, Michele Buckley, Ben Bunday, Alain Deleporte, Alain Diebold, Dick Hockett, Masakazu Ichikawa, Masahiko Ikeno, Masami Ikota, Sadayuki Imanishi, David Joy, Hiroshi Kitajima, Steve Knight, Isao Kojima, Sumio Kuwabara, Ulrich Mantz, Jack Martinez, Joaquin Martinez, Cecilia Martner, Fumio Mizuno, Kevin Monahan, Kazuo Nishihagi, Hideo Ohta, Akira Okamoto, Toshihiko Osada, Noel Poduje, Michael Postek, Gary Rubloff, Jyu-Horng Shieh, Vladimir Ukraintsev, Brad van Eck, Mauro Vascone, Dick T.A.P. Verkleij, Andras Vladar, Yuichiro Yamazakk, Stefan Zollner

Modeling and Simulation—Susumu Asada, Vivek Bakshi, Andreas Erdmann, Martin Giles, Ed Hall, Anco Heringa, Helen H. Hwang, Mitsunori Kimura, Herve Jaouen, Giles Le Carval, Jürgen Lorenz, Tuschar Merchant, Wolfgang Molzer, Meyya Meyyappan, Marius Orlowski, Concetta Riccobene, Shigeo Satoh, Wim Schoenmaker, Robert Szalapski, Walt Trybula, Hiroyuki Umimoto, Peter L.G. Ventzek, Tetsunori Wada, Tiger Wang, Reinout Woltjer, Jeff Wu, Tony Yen

REGIONAL SUPPORT TEAMS

We acknowledge and thank the regional teams for their support:

- *Europe*—Hetty Kaerts, Janneke Pels, Louis Poncin
- *Japan*—Yasuaki Hokari
- *Korea*—Du Su An
- *Taiwan*—Judy Chang
- *USA*—Yumiko Takamori, *Intel meeting support team*—Sally Yanez, Barbara Pebworth

International SEMATECH—Dave Anderson, Debra Elley, Sarah Mangum, Delores Stoner, Donna Towery, Linda Wilson
SEMI programs support—Pat Horwath, Jody Sullivan, *SIA*—Judy Ajifu Rodgers

FOREWORD

The International Technology Roadmap for Semiconductors (ITRS) is the result of a worldwide consensus building process. This document predicts the main trends in the semiconductor industry spanning across 15 years into the future. The participation of experts from Europe, Japan, Korea, and Taiwan as well as the U.S.A. ensures that the *ITRS* is a valid source of guidance for the semiconductor industry as we strive to extend the historical advancement of semiconductor technology and the worldwide integrated circuit (IC) market. These five regions jointly sponsor the *ITRS*.

The Semiconductor Industry Association (SIA) coordinated the first efforts of producing what was originally *The National Technology Roadmap for Semiconductors* (*NTRS*). The semiconductor industry became a global industry in the 1990s, as many semiconductor chip manufacturers established manufacturing or assembly facilities in multiple regions of the world. Similarly, the suppliers to the semiconductor industry have established worldwide operations. Furthermore, alliances, joint ventures, and many forms of cooperation have been established among semiconductor manufacturers as well as among equipment, materials, and software suppliers.

The above considerations led to the realization that a Roadmap that provides guidance for the whole industry would benefit from inputs from all regions of the world that have leadership activities in the field of semiconductors. This realization led to the creation of the *International Technology Roadmap for Semiconductors* (*ITRS*). The invitation to cooperate on the *ITRS* was extended by the SIA at the World Semiconductor Council in April of 1998 to Europe, Japan, Korea, and Taiwan.

The most recent full revision of the Roadmap occurred in 2001. The *2001 ITRS* presented the latest prediction of the industry needs and potential solutions through 2016. Additionally, it also began to address the limits of traditional scaling, the opportunity for “equivalent scaling” [the extension of the device scaling approach by improving electrical performance with new or improved materials], the challenges of the industry to maintain historical trends in productivity, and the introduction of the investigation and feasibility of new device candidates within the timeframe of the Roadmap (2013-2016). In 2002, the *ITRS* international technology working groups (ITWGs) reviewed the tables of the *2001 ITRS* edition. The resulting update and clarification produced the *2002 ITRS Update*. Over 100 tables were revised.

The *2003 ITRS* represents a full revision of all the tables as well as a revision of the text. This new edition of the ITRS extends to the year 2018. The *2003 ITRS* does not predict a further acceleration in the timing of introduction of new technologies as the industry struggled through the worst recession of its history. It is projected that the half-pitch of DRAM (hp90 nm) will be introduced in 2004.

The Emerging Research Devices portion of the PIDS chapter has been substantially updated to expand the non-classical CMOS subject. In addition, analysis of potential post-CMOS devices has been substantially expanded to include considerations on alternate logic-state-variable nanoscale devices. This paves the way to a complete technological revolution looming ahead towards the end of the next decade.

Another addition to the PIDS section addresses the technology needs of wireless applications. For the first time a relative comparison of the ultimate performance of silicon devices and the performance of devices produced by using other semiconductors is introduced into the *ITRS*.

It is the purpose of the *ITRS* documents to provide a reference of requirements, potential solutions, and their timing for the semiconductor industry. This objective has been accomplished by providing a forum for international discussion, cooperation, and agreement among the leading semiconductor manufacturers and the leading suppliers of equipment, materials, and software, as well as researchers from university, consortia, and government labs.

In the last few years, the ITRS documents have become a truly common reference for the entire semiconductor industry. Indeed, the cooperative efforts of the ITRS participants have fostered cooperation among international consortia, universities, and research institutions around the world. It is hoped that the *2003 ITRS* will further contribute to stimulate cooperative R&D investments so that the financial burden can be more uniformly shared by the whole industry. It is also hoped that the *2003 ITRS* will continue to stimulate the fundamental elements that encourage innovation in individual companies.

TABLE OF CONTENTS

Foreword i

Introduction 1

- Overview 1
  - 2003 ITRS Special Topics 2
- Meaning of ITRS Technology Requirements 4
  - Position on Potential Solutions 4
- Overall Roadmap Process and Structure 5
  - Technology Characteristics / Requirements Tables 6
  - Technology Nodes 9
  - Drivers for ITWG Technology Requirements 10

Grand Challenges 11

- Overview 11
- In the Near Term (through ~2009) 11
  - Enhancing Performance 11
  - Cost-effective Manufacturing 13
- In The Long Term (~2010 through 2018) 16
  - Enhancing Performance 16
  - Cost-effective Manufacturing 17

Difficult Challenges 18

- System Drivers and Design 18
- Test and Test Equipment 19
- Process Integration, Devices, and Structures (including RF and Mixed-signal Technologies for Wireless Communications and Emerging Research Devices) 21
- Front End Processes 23
- Lithography 25
- Interconnect 26
- Factory Integration 27
- Assembly and Packaging 29
- Environment, Safety, and Health 31
- Yield Enhancement 33
- Metrology 34
- Modeling and Simulation 35

Overall Roadmap Technology Characteristics 36
Background 36
Overview of 2003 Revisions 36
Definitions 36
Roadmap Timeline 37
Product Generations and Chip-Size Model 40
Chip-Size, Lithographic-Field, and Wafer-Size Trends 48
Performance of Packaged Chips 50
Electrical Defect Density 55
Power Supply and Power Dissipation 56
Cost 58
Glossary 61
Characteristics of Major Markets 61
Chip and Package—Physical and Electrical Attributes 64
Fabrication Attributes and Methods 64
Maximum Substrate Diameter (mm) 64
Electrical Design and Test Metrics 65
Design and Test 65

LIST OF FIGURES

Figure 1 Composition of the Technology Working Group—936 Global Participants 6
Figure 2 A Typical Production “Ramp” Curve 8
Figure 3 Technology Node Compared to Actual Production Capacity Technology Distribution ... 9
Figure 4 Definition of Metal Half Pitch 10
Figure 5 MOS Transistor Scaling—1974 to present 37
Figure 6 Scaling Calculator 37
Figure 7 2003 ITRS—Half Pitch Trends 39
Figure 8 2003 ITRS—Gate Length Trends 40

LIST OF TABLES

Table A Improvement Trends for ICs Enabled by Feature Scaling 1
Table B ITRS Table Structure—
Key Lithography-related Characteristics by Product Type 7
Table C Rounded versus Actual Trend Numbers 38
Table 1a Product Generations and Chip Size Model Technology Nodes—
Near-term Years 41
Table 1b Product Generations and Chip Size Model Technology Nodes—
Long-term Years 41
Table 1c DRAM Production Product Generations and Chip Size Model—
Near-term Years 42
Table 1d DRAM Production Product Generations and Chip Size Model—
Long-term Years 42
Table 1e DRAM Introduction Product Generations and Chip Size Model—
Near-term Years 43
Table 1f DRAM Introduction Product Generations and Chip Size Model—
Long-term Years 43
Table 1g MPU (High-volume Microprocessor) Cost-Performance
Product Generations and Chip Size Model—Near-term Years 44
Table 1h MPU (High-volume Microprocessor) Cost-Performance
Product Generations and Chip Size Model—Long-term Years 45
Table 1i High-Performance MPU and ASIC Product Generations and
Chip Size Model—Near-term Years 46
Table 1j High-Performance MPU and ASIC Product Generations and
Chip Size Model—Long-term Years 47
Table 2a Lithographic-Field and Wafer-Size Trends—Near-term Years 49
Table 2b Lithographic-Field and Wafer Size Trends—Long-term Years 50
Table 3a Performance of Packaged Chips: Number of Pads and Pins—
Near-term Years 51
Table 3b Performance of Packaged Chips: Number of Pads and Pins—
Long-term Years 51
Table 4a Performance and Package Chips: Pads, Cost—Near-term Years 53
Table 4b Performance and Package Chips: Pads, Cost—Long-term Years 53
Table 4c Performance and Package Chips: Frequency On-chip Wiring Levels—
Near-term Years 54
Table 4d Performance and Package Chips: Frequency, On-chip Wiring Levels—
Long-term Years 54
Table 5a Electrical Defects—Near-term Years 55
Table 5b Electrical Defects—Long-term Years 55
Table 6a Power Supply and Power Dissipation—Near-term Years 57
Table 6b Power Supply and Power Dissipation—Long-term Years 57
Table 7a Cost—Near-term Years 59
Table 7b Cost—Long-term Years 60

INTRODUCTION

OVERVIEW

For four decades, the semiconductor industry has distinguished itself by the rapid pace of improvement in its products. The principal categories of improvement trends are shown in Table A with examples of each. Most of these trends have resulted principally from the industry's ability to exponentially decrease the minimum feature sizes used to fabricate integrated circuits. Of course, the most frequently cited trend is in integration level, which is usually expressed as Moore's Law in December 1975 (i.e., the number of components per chip doubles every 24 months). The most significant trend for society is the decreasing cost-per-function, which has led to significant improvements of productivity and quality of life through proliferation of computers, electronic communication, and consumer electronics.

Table A Improvement Trends for ICs Enabled by Feature Scaling

TREND	*EXAMPLE*
Integration Level	**Components/chip, Moore's Law**
Cost	**Cost per function**
Speed	**Microprocessor clock rate, GHz**
Power	**Laptop or cell phone battery life**
Compactness	**Small and light-weight products**
Functionality	**Nonvolatile memory, imager**

All of these improvement trends, sometimes called "scaling" trends, have been enabled by large R&D investments. In the last two decades, the growing size of the required investments has motivated industry collaboration and spawned many R&D partnerships, consortia, and other cooperative ventures. *The International Technology Roadmap for Semiconductors (ITRS)* has been an especially successful worldwide cooperation. It presents an industry-wide consensus on the "best current estimate" of the industry's research and development needs out to a 15-year horizon. As such, it provides a guide to the efforts of companies, research organizations, and governments. The *ITRS* has improved the quality of R&D investment decisions made at all levels and has helped channel research efforts to areas that truly need research breakthroughs.

The 2003 edition of the *ITRS* is the result of the continued worldwide consensus building process. The participation of semiconductor experts from Europe, Japan, Korea, Taiwan, and U.S.A. ensures that the *2003 ITRS* continues to be the definitive source of guidance for semiconductor research as we strive to extend the historical advancement of semiconductor technology and the integrated circuit market. This is the third edition of the fully revised *ITRS* produced by worldwide participation throughout. The diverse expertise and dedicated efforts that this international effort mobilized have brought the Roadmap to a new level of worldwide consensus about future semiconductor technology requirements.

The complete *2003 ITRS* and past editions of the ITRS editions are available for viewing and printing as an electronic document at the Internet web site *http://public.itrs.net*.

2003 ITRS SPECIAL TOPICS

WIRELESS TECHNOLOGY

Wireless applications have quickly grown to become an important driver for semiconductor products and technologies. The International Roadmap Committee (IRC) responded to this fact by extending the scope of the already existing analog and mixed-signal (AMS) working group to include radio frequency (RF) technologies for wireless communication. The IRC requested this new extended working group to include III-V compound semiconductors in the *2003 ITRS* Roadmap as a section in the Process Integration chapter.

The new RF and AMS working group is organized into four sub-groups:

1. Analog and mixed-signal (0.8–10 GHz frequency range),
2. RF transceivers (0.8–10 GHz),
3. Power amplifiers and power management (0.8–10 GHz), and
4. Millimeter wave (10–100 GHz).

It investigates the challenges that RF and AMS technologies have in meeting the demands of wireless applications. The technology requirements for meeting the demands of wireless systems are manifold, often conflicting and very different from digital requirements. Thus we see today in wireless systems often a combination of specialized analog and RF technologies such as Si CMOS, SiGe, Si BiCMOS, Si LDMOS, GaAs MESFET, GaAs PEMT, GaAs HBT, InP HEMT, and InP HBT. Cost and performance drive integration. Depending on requirements either monolithic system on chip (SoC) or system in package (SiP) integration may be preferred. When required, the SiP approach is especially suited to bring the specialized RF and AMS technologies together in a highly integrated, high-performance unit.

The drivers for wireless communications systems are cost, time to market, and adequate performance. Additional (technical and/or regulatory) drivers are available frequency bands, power consumption, functionality, size of mobile units, high volumes of product, standards, and protocols. RF technologies often require added tolerances for the values of performance parameters because several conflicting or competing requirements must be met simultaneously and therefore design compromises must be made. For example, the effective bit resolution of analog-to-digital converters should be greater than what is needed for converting in order to perform signal error correction in “real time” and to keep latency to a minimum. These include noise figure, power added efficiency (PAE), linearity, high output power, low current, and low voltage. Increased RF performance for silicon is usually achieved by geometrical scaling. Increased RF performance for III-V compound semiconductors is achieved by optimizing carrier transport properties through materials and bandgap engineering.

During the last two decades, technologies based on III-V compounds have established new business opportunities for wireless communications systems. When high volumes of product are expected, silicon and more recently silicon-germanium replace the III-Vs in those markets for which these group IVs can deliver appropriate performance at low cost. Developing RF and AMS technology roadmaps is not straightforward due to the complex requirements of wireless products.

In addition to the many technologies already mentioned above additional metrics than those usually associated with CMOS processes are considered. For III-V compound semiconductors these other metrics include carrier frequency for wireless applications and the printed gate length. The location of boundaries between the kinds of RF semiconductors (e.g., Si, SiGe, GaAs, and InP) is broad, diffuse, and change with time. The boundary between the group IV semiconductors Si and SiGe and the III-V semiconductor GaAs has been moving to higher frequencies with time and for other applications the boundary between GaAs and InP is tending to shift to lower frequencies. In future years, carrier frequency is expected to lose its significance in defining the boundaries among technologies for some of the applications listed therein. This is expected because most of the RF technologies can provide very high operating frequencies.

Future boundaries will be dominated more by such parameters as noise figure, output power, PAE, and linearity. Performance tends to increase in the following order: Si CMOS, SiGe, GaAs, InP, and metamorphic transistors. Two or more technologies may coexist with one another for certain applications such as cellular transceivers, modules for terminal or base station power amplifiers, and mm-wave receivers. Currently BiCMOS in cellular transceivers has the biggest share in terms of volume compared to CMOS. But the opposite may occur in the future. Today, both GaAs HBT and discrete LDMOS devices in modules for terminal power amplifiers have big market shares compared to GaAs PHEMT and GaAs MESFET. In the future, silicon based technologies having higher integration capabilities will gain importance. Today GaAs PHEMT and InP HEMT are present in mm-wave receivers. In the future competition may come

from SiGe HBT, GaAs MHEMT, and InP HEMT technologies. Nevertheless, while SiGe has shown capability in the 10–40 Gbit/s range, it is unlikely to replace III-Vs in applications where either high output power or ultra low noise is required.

EMERGING RESEARCH DEVICES

The *2001 ITRS* document marked the additional investigation of the limits of traditional scaling, its extension by improving electrical performance with new or improved materials, and of the introduction and feasibility of new device architectures. This new section on Emerging Research Devices (ERD), in 2001 highly coordinated with Process Integration, Devices, and Structures (PIDS), has evolved considerably since then. This evolution is seen in the two primary goals for Emerging Research Devices for 2003.

The first and original goal is to stimulate invention and research leading to proof of concept and feasibility demonstration for one or more Roadmap-extending concepts. A new and important corollary goal this year is to provide a balanced view of many of the exciting new approaches to information processing, articulating their potential contributions balanced with brief discussion of their limiting challenges. In addressing these goals, this section serves two purposes. First is to introduce advanced non-classical CMOS structures and memory technologies aimed at extending microelectronic technologies to the end of this Roadmap timeframe. Second is to introduce and critique (without endorsement) new technological concepts for logic and architectures aimed at extending information and signal processing beyond the end of the Roadmap timeframe.

The quickening pace of MOSFET scaling is accelerating introduction of new technologies to extend CMOS beyond the 65 nm node. These technologies include both new materials and advanced MOSFET structures. The Front End Processes chapter discusses new materials required for the gate stack and the PIDS chapter identifies technology requirements for CMOS structures. In a complementary fashion, the new ERD chapter serves as the bridge between bulk and non-classical CMOS and the realm of microelectronics beyond the end of the Roadmap.

The ERD section is divided into four categories: Non-classical CMOS, memory devices, logic devices and new information processing architectures. The discussions of non-classical CMOS, memory devices, and logic devices provide some detail regarding their operation principles, advantages, disadvantages, and maturity. The non-classical CMOS, memory and logic subsections also provide some estimates of performance upon their introduction to manufacturing and a few speculative estimates for ultimate performance. The discussion of architectures, given its "beyond-the-Roadmap" time frame, focuses on principles of operation, major advantages and challenges, and its maturity or state of exploration and development. This section ends with a preliminary but interesting comparison of the performance projections and cost attributes for several speculative new approaches to information and signal processing.

A new feature this year relates to scaling non-classical CMOS devices to the end of the Roadmap, perhaps even to a new 16 nm node, for high-performance, low operating power, and low standby power applications. Based on first-principle physics, the cumulative efficacy is assessed for extending CMOS through the 22 nm node using a wide variety of proposed "Technology Boosters" (e.g., strained silicon, ultra-thin body SOI, metal gate electrode, double gate structures, etc.). This analysis (performed assuming a high-κ gate dielectric at the 65 nm node) clearly shows that the Roadmap driven increase of 17%/year in intrinsic transistor speed together with manageable increase in leakage currents can be achieved using all the proposed boosters. It also illustrates the unprecedented need perhaps to introduce more than one Technology Booster per node.

Accepting the risk of including ideas that eventually will be shown as non-functional or impractical, the intent of the new chapter is to "cast a broad net." That is to gather in one place substantive, alternative concepts for memory, logic and information processing architectures that would, if successful, substantially extend the Roadmap beyond CMOS. As such, it will provide a window into candidate approaches. Also, the intent is to provide a balanced perspective on these new approaches, discussing their exciting possibilities in the light of their most difficult challenges. The inclusion of a particular concept in the ERD chapter does not in any way constitute advocacy or endorsement. Rather, inclusion does point out that existing research efforts are exploring a variety of basic technology and architectural concepts for information and signal processing.

MEANING OF *ITRS* TECHNOLOGY REQUIREMENTS

Since its inception in 1992, a basic premise of the Roadmap has been that continued scaling of microelectronics would further reduce the cost per function (historically, ~25% per year) and promote market growth for integrated circuits (averaging ~17% per year). Thus, the Roadmap has been put together in the spirit of a challenge—essentially, "What technical capabilities need to be developed for us to stay on Moore's Law and the other trends?" This challenge has become so formidable that more and more of the semiconductor industry's research effort, including consortia and collaboration with suppliers, has been shared in a precompetitive environment. In this process, the ITRS identifies the principal technology needs to guide the shared research. It does this in the three following ways: (1) showing the "targets" that need to be met by "technology solutions" currently under development, (2) recognizing Interim Solutions are Known," i.e., limitations of available solutions will not delay the start of production, and (3) indicating where there are no "known manufacturable solutions" (of reasonable confidence) to continued scaling in some aspect of the semiconductor technology.

Note that the new (the second) situation "Interim Solutions are Known" means work-arounds will be initially employed in these cases. Subsequent improvement is expected to close any gaps for production performance in areas such as process control, yield, and productivity. The third and last situation is highlighted as "red" on the Roadmap technology requirements tables, and has been referred to as the "Red Brick Wall" since the beginning of ITRS. (The "red" is officially on the Roadmap to clearly warn where progress might end if tangible breakthroughs are not achieved in the future.) Numbers in the red regime, however, are only meant as warnings and should not be interpreted as "targets" on the Roadmap. For some Roadmap readers, the "red" designation may not have adequately served its *sole* purpose of highlighting significant and exciting challenges. There can be a tendency to view *any* number in the Roadmap as "on the road to sure implementation" regardless of its color. To do so would be a serious mistake.

An analysis of "red" usage might classify the "red" parameters into two categories:

1. where the consensus is that the particular value will ultimately be achieved (perhaps late), but for which the industry don't have much confidence in any currently proposed solution(s), or

2. where the consensus is that the value will never be achieved (for example, some "work-around" will render it irrelevant or progress will indeed end)

To achieve the red parameters of the first category, breakthroughs in research are needed. It is hoped that such breakthroughs would result in the "red" turning to "yellow" (defined as "manufacturable solutions are known") and, ultimately "white" (defined as "manufacturable solutions are known and are being optimized") in future editions of *ITRS*.

Traditionally, the ITRS has focused on the continued scaling of CMOS (Complementary Metal-Oxide-Silicon) technology. However, since 2001, we have reached the point where the horizon of the Roadmap challenges the most optimistic projections for continued scaling of CMOS (for example, MOSFET channel lengths below 9 nm). It is also difficult for most people in the semiconductor industry to imagine how we could continue to afford the historic trends of increase in process equipment and factory costs for another 15 years! Thus, the *ITRS* must address post-CMOS devices. The Roadmap is necessarily more diverse for these devices, ranging from more familiar non-planar CMOS devices to exotic new devices such as spintronics. Whether extensions of CMOS or radical new approaches, post-CMOS technologies must further reduce the cost-per-function and increase the performance of integrated circuits. Thus new technologies may involve not only new devices, but also new manufacturing paradigms.

The scope of the *2003 ITRS* specifically includes detailed technology requirements for all CMOS integrated circuits, including wireless communication and computing products. This group constitutes over 75% of the world's semiconductor consumption. Of course, many of the same technologies used to design and manufacture CMOS ICs are also used for other products such as compound semiconductor, discrete, optical, and micro-electromechanical systems (MEMS) devices. Thus, to a large extent, the Roadmap covers many common technology requirements for most IC-technology-based micro/nanotechnologies, even though that is not the explicit purpose of the Roadmap.

POSITION ON POTENTIAL SOLUTIONS

The *ITRS* strives to avoid prematurely identifying definite solutions to the future technology challenges. This is difficult, since guidance on the research needs is intended. Despite this need to provide guidance, the Roadmap participants are continually pursuing new ways to prevent the Roadmap itself from being interpreted as limiting the range of creative approaches to further advance microelectronics technology. One of the resulting compromises has been to only present

illustrative examples of potential solutions to selected challenges in the *ITRS*. These are not to be construed even as complete lists of all solutions suggested to date, much less exhaustive lists of what should be explored. A few of the potential technical solutions are listed, where known, only to inform the readers of current thinking and efforts. Furthermore, the listing of a particular potential solution does not constitute an endorsement by the Roadmap process.

It *is* the intent of this document to identify the technological barriers and when the industry will likely run into them. It is *not* the intent of this document to identify the most likely solutions to be adopted, nor to focus attention on those potential solutions currently known at the expense of other new concepts. In fact, it is eagerly hoped that this Roadmap will inspire additional innovative solutions. *The semiconductor industry's future success continues to depend on new ideas.*

OVERALL ROADMAP PROCESS AND STRUCTURE

Overall coordination of the *ITRS* process is the responsibility of the International Roadmap Committee (IRC), which has two-to-four members from each sponsoring region (Europe, Japan, Korea, Taiwan, and the U.S.A.). The principal IRC functions include the following:

- Providing guidance/coordination for the ITWGs
- Hosting the *ITRS* Workshops
- Editing the *ITRS*

International Technology Working Groups (ITWGs) write the corresponding technology-area chapters of the ITRS. The ITWGs are of two types: *Focus* ITWGs and *Crosscut* ITWGs. The Focus ITWGs correspond to typical sub-activities that sequentially span the Design/Process/Test/Package product flow for integrated circuits. The Crosscut ITWGs represent important supporting activities that tend to individually overlap with the "product flow" at multiple critical points.

For the *2003 ITRS*, the Focus ITWGs are the following:

- Design
- System Drivers
- Test and Test Equipment
- Process Integration, Devices, and Structures
    - The PIDS chapter includes a new section, RF and Analog/mixed-signal Technologies for Wireless Communications
    - Expands the Emerging Research Devices section
- Front End Processes
- Lithography
- Interconnect
- Factory Integration
- Assembly and Packaging

Crosscut ITWGs are the following:

- Environment, Safety, and Health
- Yield Enhancement
- Metrology
- Modeling and Simulation

Each ITWG receives inputs from the regional Technology Working Groups (TWGs). Regional TWGs send one to two representatives to participate on their corresponding ITWG and to attend ITRS meetings. The regional TWGs are composed of experts from industry (chip-makers as well as their equipment and materials suppliers), government research organizations, and universities. In 2003, a total of 936 experts volunteered their support and expertise (an increase from 2001 of 839 participants). The composition of the total TWG membership is analyzed in Figure 1.

It is important to note that per region, particular sectors of the industry are more predominant and the demographics for that region indicate this in the composition by affiliation. For example, there are not many supplier companies in Taiwan, therefore the percentage of participants from suppliers is low. In the United States and Japan, the supplier participation reflects the fact there are more supplier companies in those regions. Likewise, the demographics per ITWG also reflect the affiliations that populate the technology domains. For Emerging Research Devices, a longer-term focus area, the percentage of research participants is 42%, while suppliers is only 8%. In the process technologies of Front End Processes (43%), Lithography (27%), and Interconnect (58%), the percentages of suppliers reflect the equipment/materials suppliers' participation as much higher due to the near-term requirements that must be addressed.

The ITRS 15-year projections are assessed by the TWG teams, based on current industry indicators, research and development progress, historical trends, and data interpolations based on the Overall Roadmap Technology Characteristics. The TWGs review the draft roadmap information by holding quarterly meetings. These meetings serve to review and incorporate feedback gathered from an even larger community through "sub-TWG meetings."

For the 2003 edition, three *ITRS* meetings were also held world-wide as follows: Amsterdam, the Netherlands (sponsored by the ESIA and hosted by Philips Semiconductor); San Francisco, U.S.A., sponsored by the SIA and organized by International SEMATECH; and Hsinchu, Taiwan (sponsored and hosted by the TSIA). These meetings provided the main forums for face-to-face discussions among the members of each ITWG and coordination among the different ITWGs. In addition, the ITRS teams hold public ITRS conferences bi-annually to present the latest ITRS information and to solicit feedback from the semiconductor industry at-large.

The ITRS is released annually, with updates and corrections to data tables each even numbered year (such as 2000, 2002, 2004) while complete editions are released each odd-numbered year (2001, 2003, 2005). This ITRS process thus ensures continual assessment of the semiconductor industry's near and long-term needs. It also allows the teams to correlate in a timely fashion the ITRS projections to most recent research and development breakthroughs that may provide solutions to those needs.

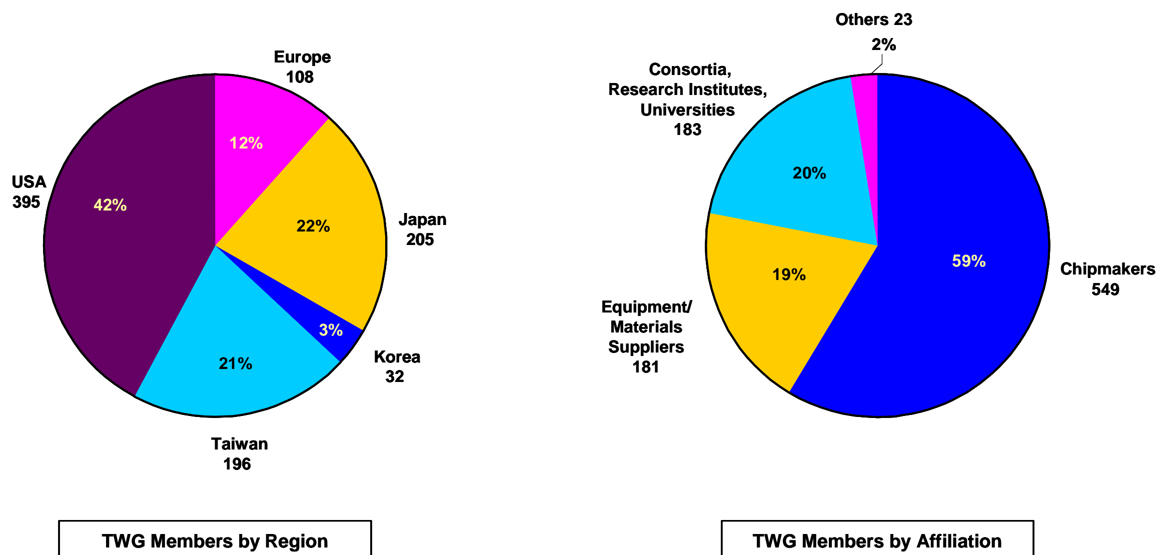


Figure 1 Composition of the Technology Working Group—936 Global Participants

TECHNOLOGY CHARACTERISTICS/REQUIREMENTS TABLES

A central part of the IRC guidance and coordination is provided through the initial creation (as well as continued updating) of a set of Overall Roadmap Technology Characteristics (ORTC) tables. These tables summarize key high-level technology requirements, which define the future "technology nodes" and generally establish some common reference points to maintain consistency among the chapters written by individual ITWGs. The high-level targets expressed in the ORTC tables are based in part on the compelling economic strategy of maintaining the historical high rate of advancement in integrated circuit technologies. Thus, the ORTC provide a "top-down business incentive" to balance the tendency for the ITWGs to become conservative in expressing their individual, detailed future requirements.

Each ITWG chapter contains several principal tables. They are individual ITWGs' technology requirements tables patterned after the ORTC tables. For the *2003 ITRS*, the ORTC and technology requirements tables are separated into "Near-term Years" (2001, 2002… through 2009) and "Long-term Years" (2010, 2012, 2013, 2015, 2016, and 2018). This format is illustrated in Table B, which contains a few key rows from lithography-related ORTC tables.

The ITRS technology node is defined as the minimum metal pitch used on any product, for example, either DRAM half pitch or Metal 1 (M1) half pitch in Logic/MPU. In 2003, DRAMs continue to have the smallest metal half pitch; thus it continues to represent the technology node. Commercially used numbers for the technology generations typically differ from the ITRS technology node numbers. However, the most reliable technology standard in the semiconductor industry is provided by the above definition, which is quite clear in that the patterning and processing (etching, etc.) capability of the technology is represented as the pitch of the minimum metal lines. The above definition is maintained not only for the 2003 version, but also as a continuation from previous editions. Therefore, the official *2003 ITRS* metal hpXX node indicator has been added to differentiate the ITRS definition from commercial technology generation numbers. Interim shrink-level node trend numbers are calculated and included for convenience of monitoring the inter-node progress of the industry.

Table B ITRS Table Structure—Key Lithography-related Characteristics by Product Type

Near-term Years

YEAR OF PRODUCTION	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		**hp90**			**hp65**		
DRAM ½ Pitch (nm)	**100**	**90**	**80**	**70**	**65**	**57**	**50**
MPU/ASIC M1 ½ Pitch (nm)	**120**	**107**	**95**	**85**	**75**	**67**	**60**
MPU/ASIC Poly Si ½ Pitch (nm)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm)	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**

Long-term Years

YEAR OF PRODUCTION	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	**hp45**		**hp32**		**hp22**	
DRAM ½ Pitch (nm)	**45**	**35**	**32**	**25**	**22**	**18**
MPU/ASIC M1 ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC Poly Si ½ Pitch (nm)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm)	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**

The ORTC and technology requirements tables are intended to indicate current best estimates of introduction timing for specific technology requirements. Please refer to the Glossary for detailed definitions for Year of Introduction and Year of Production.

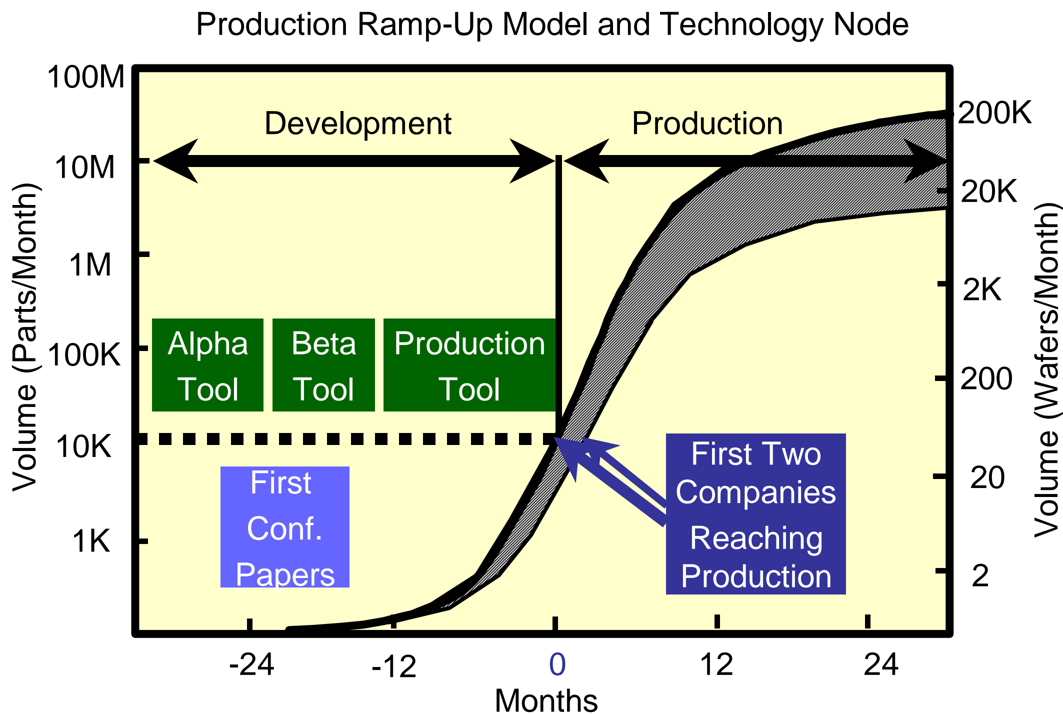


Figure 2 A Typical Production "Ramp" Curve

The ORTC and technology requirements tables are intended to indicate current best estimates of introduction time points for specific technology requirements. Ideally, the Roadmap might show multiple time points along the "research-development-prototyping-manufacturing" cycle for each requirement. However, in the interests of simplicity, usually only one point in time is estimated. The default "Time of Introduction" in the *ITRS* is the "Year of Production," which is defined in Figure 2.

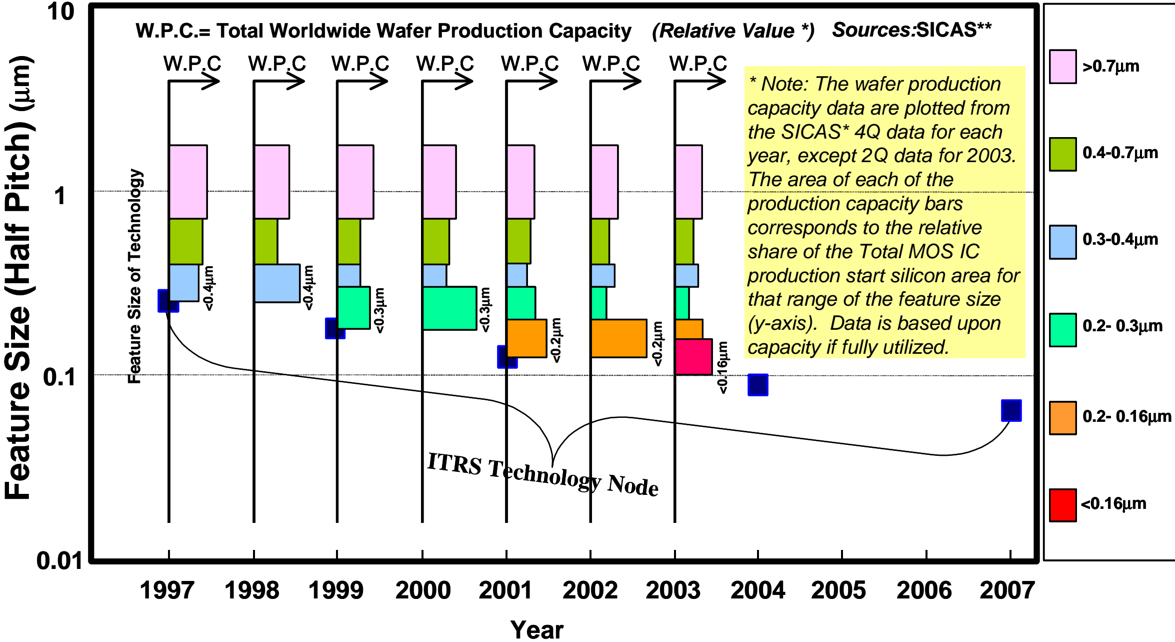
The "Production" time in ITRS refers to the time when the first company brings a technology to production and a second company follows within three months. Production means the completion of both process and product qualification. The product qualification means the approval by customers to ship products, which may take one to twelve months to complete after product qualification samples are received by the customer. Preceding the production, process qualifications and tool development need to be completed. Production tools are developed typically 12 to 24 months prior to production. This means that alpha and succeeding beta tools are developed preceding the production tool.

Also note that the Production "time zero (0)" in Figure 2 can be viewed as the time of the beginning of the ramp to full production wafer starts. For a fab designed for 20K wafer-starts-per-month (WSPM) capacity, the time to ramp from 20 wspm to full capacity can take nine to twelve months. This time would correspond to the same time for ramping device unit volume capacity from 6K units to 6M units per month if the chip size were 140 mm^2 (430 gross die per 300 mm wafer × 20K wspm × 70% total yield from wafer starts to finished product = 6M units/month).

It is noted that the ITRS, by its definition, focuses on forecasting the earliest introduction of the leading-edge technologies in respective fields for producing semiconductors. Therefore, it is obvious that many companies, for a variety of reasons, may choose to introduce a technology node later than the earliest introductions, hence there is a wide variation of the technologies in actual production status from leading edge to trailing edge. Figure 3 shows, in horizontal bar graph (normalized by bar area to total MOS IC industry silicon processing capacity), the actual, annual worldwide wafer production capacity distributions over different process feature sizes. The distributions are quite widespread while the ITRS technology nodes, shown in small blue marks, are located exactly on the leading edges of each of them.

The data for the graphical analysis were supplied by the Semiconductor Industry Association (SIA) from their Semiconductor Industry Capacity Supply statistics (SICAS). The detailed data are available to the public online at the SIA website, http://www.sia-online.org/pre_statistics.cfm. It can be noted from analysis of the SICAS data that the

relative percentage of the most leading-edge technology capacity has been rapidly growing. The combined capacity of the most recent two technology nodes rapidly grows to nearly half the capacity of the industry within two to three years after their introduction.



*** Source: Semiconductor Industry Capacity Statistics (SICAS) – collected from worldwide semiconductor manufacturers (estimated >90% of Total MOS Capacity) and published by the Semiconductor Industry Association (SIA), as of July 2003.*

Figure 3 Technology Node Compared to Actual Production Capacity Technology Distribution

Note that some rows in the ORTC and technology requirements tables refer to timing points other than production and are defined for each case (e.g., “at sample”). Also, for the long-term years, for which the table intervals are three years, it is possible for the best-estimate year of production to fall in between the selected three-year intervals for some technology requirements. It was decided, for convenience to ITRS readers, to retain the *2001 ITRS* Long-term year columns (2010, 2013, 2016) to facilitate comparison, since the new *2003 ITRS* long-term year columns (2012, 2015, 2018) fall between the long-term technology nodes timing.

TECHNOLOGY NODES

The concept of “technology nodes” used to be quite straightforward to understand as it has historically been linked to the introduction of new Dynamic Random Access Memory (DRAM) generations with a 4× increase in bits/chip between generations. For as long as this cycle strictly followed Moore’s Law (three-year cycle for 4×), the technology nodes and DRAM generations were essentially synonymous. However, in recent years, a greater diversity of products serving as technology drivers, faster introduction/optimization of product-specific technology, and the general increase in business and technology complexity are all tending to de-couple the many technology parameters that have traditionally characterized “advance to the next technology node.” For example, microprocessor unit (MPU) and ASIC or Logic products have recently driven the reduction of gate length at a faster pace than lithography half-pitch. While DRAM continues to drive the lithography half-pitch, MPU/ASIC drives the gate length. Even the choice of basic lithography technology has tended to become more product specific (such as “pushing the wavelength as fast as possible” versus “using phase-shift masks”). Following the practice of the *2001 ITRS*, the *2003 ITRS* ORTC tables list the DRAM half-pitch, the MPU /ASIC half-pitch, and MPU gate length, as shown in Table B. These technology parameters are defined in Figure 4, which now also includes the MPU/ASIC Metal 1 (M1) half-pitch line item. Any of these five parameters (rows) in Table B may be chosen as the driver for a given technology requirements table of a given ITWG. Nevertheless, for a point of reference, the DRAM half-pitch is still used in the *2003 ITRS* as the designation of the technology node (DRAM still requires the smallest half pitch among all products). For example, according to Table B, 2004 will be the year of production of the 90 nm node. Again, the “node designation” is defined by DRAM half pitch, not by the transistor gate length or minimum feature size characteristic of that node. Additional (and, in some cases, more precise) definitions related to the *ITRS* tables may be found in the *Glossary*.

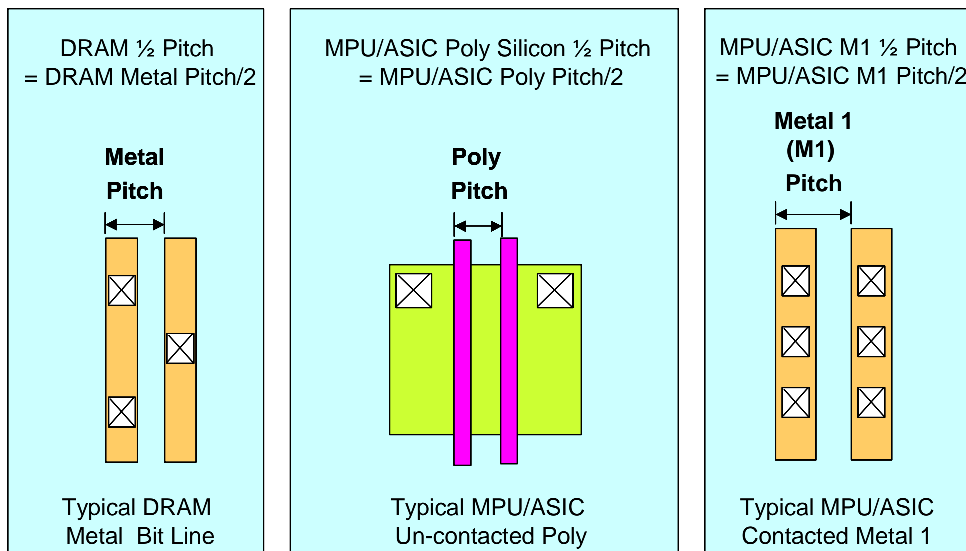


Figure 4 Definition of Metal Half Pitch

In recent years, the scaling ratio of the minimum feature size of one technology generation to the previous generation may no longer adhere to the historical value of 0.7. For example, 100 nm is not 0.7× of 130 nm. In addition, some companies may choose to introduce a "half node" (for example 150 nm may be considered a half node between the 180 nm node and the 130 nm node) with the intention of introducing the next "full" node at a later time. The *2003 ITRS* acknowledges the validity of these practices and follows the *2001 ITRS* in listing the near-term (2003–2009) technology requirements on yearly intervals and the long-term (2010–2018) requirements on three-year intervals as shown in Table B. Thus, we can say that 2003 is the year of production of the 100 nm node and 2004 is the year of production of the 90 nm node. One may also interpret the long-term three-year interval table as a "challenge" to extend the trend of one major new technology node every three years with scaling ratio of 0.7 between nodes through the 22 nm node (which include 9 nm transistor gate lengths) in 2016.

DRIVERS FOR ITWG TECHNOLOGY REQUIREMENTS

The particular lithography-related rows selected for Table B from the ORTC tables are special in that any one of them may be designated by an ITWG as a "driver" for any specific row in one of their technology requirements tables. For example, the physical gate length may be the appropriate driver for the gate CD control and the source/drain junction depth. The designation of drivers for technology requirements provides some traceable and supportable assumptions for constructing the ITWG tables. It also provides useful links between the ORTC tables and the ITWG tables. Thus, as the Roadmap is updated in subsequent editions, these links will be used for creating a first-pass version of the new tables. For example, if the requirements in one of these driver rows of the ORTC tables were subsequently pulled-in by one year, it would be assumed that rows in the ITWG technology requirements tables would shift by default along with their designated ORTC driver row.

GRAND CHALLENGES

IN THE NEAR- (THROUGH 2009) AND LONG-TERM (2010 AND BEYOND)

OVERVIEW

The *2001* edition of *ITRS* reported on the presence of a potential "Red Brick Wall" circa 2005 to 2006 (as indicated by the red-colored cells in the technology requirements) that could block the further scaling envisaged by Moore's Law. However, the continued research and development efforts in our industry have succeeded in pushing some of the difficulties into the future. In the chapter on lithography, immersion technology, a potential solution that has been added as a candidate to extend the limit of photolithography, is a good example. In the chapters on PIDS and FEP, on the other hand, it is observed that careful optimization of device parameters by "sharing the pain" has been proven to still enable continuing performance improvement in MOSFETs. Nevertheless, there remain many technological challenges to be overcome to achieve further scaling and continuing growth of the LSI industry.

In this chapter, difficult challenges identified by each ITWG in the *Difficult Challenges* sections are gathered and summarized under the rubric of "Grand Challenges." This chapter is intended to help readers grasp the overall picture concerning major technological issues. These Grand Challenges are classified into two categories: Enhancing Performance and Cost-effective Manufacturing. They are also described according to the "near term" (2003 through 2009) and the "long term" (2010 through 2018) timeframes of the Roadmap.

The reader should realize, by looking back at Table B that the Near-term years span from 2003 to 2009 while the Long-term years span from 2010 to 2018. The reader will notice however that the year 2010 represents the year in which the hp45 technology node will occur. Several ITWGs have included any consideration related to this technology node in their Near-term challenges since it was easier for many of them to concentrate their attention on a well-defined technology node. Normally the technology nodes are studied very attentively while the data shown in the tables for the intermediate years are obtained by interpolation.

IN THE NEAR TERM (THROUGH ~2009)

ENHANCING PERFORMANCE

PERFORMANCE AND POWER DISSIPATION IN HIGH-PERFORMANCE APPLICATIONS [PROCESS INTEGRATION, DEVICES, AND STRUCTURES]

Aggressive scaling of the gate length in high-performance applications makes device parameter optimization quite difficult. Ultra-shallow junction formation for suppressing the short channel effect cannot be achieved without incurring a significant increase of parasitic resistance. Doping requires quite precise profile design and process control, whereas increasing channel doping concentration degrades carrier mobility, lowering the drain current. Moreover, statistical fluctuation of channel dopants causes increasing variation of the threshold voltage, posing difficulty in circuit design while scaling the supply voltage. Gate insulator, on the other hand, becomes thinner and thinner due to the requirement for rapid CV/I improvement. Excessive gate leakage current due to excessively thin oxynitride films will necessitate the introduction of high-κ material circa 2007. Introduction of various technology innovations ("technology boosters"), such as strained-Si, metal gate, ultra-thin body (UTB) SOI MOSFETs, and multiple-gate MOSFETs including FinFETs, should also be considered to meet the drain current requirement and to control short-channel effects with scaling. The solutions for these issues should be pursued concurrently with circuit design and architecture improvements, particularly to manage power dissipation.

PERFORMANCE AND LEAKAGE IN LOW POWER APPLICATIONS [PROCESS INTEGRATION, DEVICES, AND STRUCTURES]

In low-power applications (mainly portable products), suppression of the leakage current is strictly required in low stand-by power (LSTP) applications, whereas maximization of device performance under low supply voltage becomes more important in low operating power (LOP) applications. In both LSTP and LOP applications, the gate leakage of oxynitride

films will soon reach an unacceptable level, driving introduction of high-κ material as early as 2006. In LSTP applications, very slow scaling of the supply voltage makes overall device optimization difficult. Introduction of "technology boosters" mentioned in the previous paragraph should also be considered, and power management using circuit design and architecture techniques becomes important.

NEW GATE STACK PROCESSES AND MATERIALS [FRONT END PROCESSES]

Continued reduction of the gate length accompanies approximately proportional reduction of the gate oxide thickness in the region where tunneling leakage current becomes dominant. In MPU applications, for example, it is envisaged that the gate physical thickness will reach 1 nm as early as 2006. Although extension of oxynitride to less than 1 nm may be consistent with the device reliability requirement for MPU, it will no longer meet the strict leakage current requirement in low standby power (LSTP) applications. Therefore, introduction of a higher dielectric constant (high-κ) material in which tunneling current can be suppressed while maintaining the drain current will be necessary in LSTP applications first (circa 2006), followed by MPU applications (circa 2007). In either case, the gate electrode material and process should be optimized so that the depletion width in the gate electrode may be minimized and the boron-diffusion prevented. The former necessitates the introduction of metal gates having appropriate work function after the conventional poly Si ceases to work. These material changes pose a great challenge in MOSFET technology, where silicon dioxide/poly Si has long played a central role as the most reliable gate stack system.

CMOS INTEGRATION OF NEW MEMORY MATERIALS AND PROCESSES [FRONT END PROCESSES]

Continued DRAM scaling requires construction of memory capacitors in ever-smaller cell area, while maintaining the memory capacitance of 25–35 fF to ensure reliability of stored data. This has resulted in the introduction of dielectric materials with a high dielectric constant (high κ), such as aluminum oxide, aluminates (for example, HfAlOx) and tantalum oxide, along with a 3D memory structure. The capacitor structures are shifting from metal-insulator-silicon (MIS) to metal-insulator-metal (MIM) to avoid problems associated with capacitor dielectric thickness. For further scaling, however, it will be necessary to address process construction by using a thinner dielectric film and/or a higher dielectric constant material. In flash memory devices, on the other hand, continuous scaling and the reduction in write voltage requires the use of a thinner inter-poly and tunnel oxide, suggesting the need to introduce high-κ material into flash memory process. Along with scaling issue of ferroelectric material in FeRAM, process integration of these materials will continue to pose major challenges in the development of memory applications.

CD AND L_{EFF} CONTROL [FRONT END PROCESSES AND LITHOGRAPHY]

With aggressive scaling of gate length, control of critical dimension (CD) has been one of the most difficult issues in lithography and etching. In particular, resist slimming and profile-control of the sidewall, which are both commonly utilized to minimize the dimension of effective gate length (L_{eff}), have made CD control far more difficult. Although the acceptable 3-sigma scattering of the gate length is shared by lithography and etching at an optimum ratio, the tolerances in both technologies are approaching their limits. In addition, it is becoming very difficult to suppress line edge roughness (LER) even by the optimum control of resist printing and etching, because it is affected by polymer characteristics of resist material. CD control and LER measurement also pose challenges to metrology in terms of accuracy and efficiency.

OPTICAL AND POST-OPTICAL MASK FABRICATION [LITHOGRAPHY]

Accuracy and fabrication cost of optical masks continue to be major concerns in lithography. To realize aggressively scaled MPU gate length, gate electrode formation utilizes post-etch in addition to resist slimming. Since part of linewidth tolerance (CD control) in the completed gate electrode is shared by the etch process, the margin in lithography process is reduced. Moreover, utilization of low κ1 process enhances the mask error factor (MEF), resulting in the requirement for more stringent accuracy in mask writing. Thus, it is becoming hard for the current mask writer to meet the accuracy required by the Roadmap. From the viewpoint of mask cost, increasing difficulty in inspections, e.g., optical proximity correction (OPC), rising machine cost of drawing, and lowering mask yield with aggressive scaling, raises the mask cost greatly. Regarding mask technology in next-generation lithography (NGL), it is not yet clear how to construct pellicles and defect free masks.

INTRODUCTION OF NEW MATERIALS TO MEET CONDUCTIVITY AND DIELECTRIC PERMITTIVITY [INTERCONNECT]

To minimize the interconnect delay, development of low dielectric constant (low κ) material together with low-resistivity metal system is critical. Low-κ material should have sufficient mechanical integrity to survive harsh integration processes, such as chemical mechanical planarization (CMP), etching, and assembly/packaging. Since resistivity of

narrow-lined Cu interconnect is predicted to start to increase below 100 nm linewidth due to electron scattering at the Cu/barrier-metal interface, care should be taken for intermediate wiring at hp65 nm node (circa 2007). "Barrier engineering" including construction of very thin and low-resistivity barrier metal, as well as efficient "pore shield" for low-κ material, is essential to achieve high-reliability of interconnect system.

ENGINEERING MANUFACTURABLE INTERCONNECT STRUCTURES [INTERCONNECT]

Introduction of new materials and technologies for interconnect has raised additional issues due to their combinations and interactions. These include adhesion at the interfaces, contamination, diffusion, and leakage concerns. Mechanical damage by CMP has significantly retarded implementation of low-κ material. Complexity in interconnect structure also makes the effective dielectric constant deviate from its intrinsic value. Failure mechanisms in the Cu/low-κ systems should be clarified, along with establishment of detection metrology and predictive models. In regard to assembly and packaging technology, lack of optimization tools for interconnect/packaging architecture design makes total optimization of interconnect system difficult.

POWER MANAGEMENT [DESIGN]

Even off-currents in low-power devices increase by a factor of 10 per node, and so design technology must maintain constant static power. On the other hand, while power dissipation for high-performance MPU will exceed package limits by 25× in 15 years, design technology must achieve power limits. As a result, efficient power management requires highly complex controllability across the entire large-scale integrated (LSI) circuit. Additionally, any power optimization must simultaneously and fully exploit varying degrees of freedom, for example, by switching the operating power state of circuits using multi-V_t, multi-T_{ox}, multi-V_{dd} in the LSI core block while guiding the architecture, operating system, and software.

HIGH-FREQUENCY CIRCUIT MODELING FOR 5–40 GHZ APPLICATIONS [MODELING AND SIMULATION]

Accurate and efficient modeling of interconnect parasitics and delays is of prime importance. 2D and 3D effects on interconnects must be considered with their statistical variations. Partitioning is needed for distributed R-C-L extractions. Efficient simulation techniques should handle multi-layer dielectrics. Compact models for active devices are needed for e.g. HBTs, CMOS and LDMOSTs. These include non-quasi-static effects and surrounding parasitics. Compact models for passive devices are needed for e.g. varactors, inductors, high-density capacitors, transformers and transmission lines. The parameter extraction for RF compact models preferably tries to minimize RF measurements. Parameters should be extracted from standard I-V and C-V measurements with supporting simulations, if needed. Third harmonic distortion for 40 GHz applications implies modeling of harmonics up to 120 GHz. Modeling of effects that have a more global influence gains in importance. Examples are cross talk, substrate return path, substrate coupling, EM radiation, and heating. For these global effects accurate and efficient (layout) extraction techniques are needed. If possible, models should be physics-based to enable efficient modeling of statistics and variations.

FRONT-END PROCESS MODELING FOR NANOMETER STRUCTURES [MODELING AND SIMULATION]

Front-end process modeling for nanometer structures is the key challenge for the prediction of result from device fabrication. It overlaps to some extent with the Difficult Challenge "Ultimate nanoscale CMOS simulation capability", which however also includes materials and device simulation. Most important and challenging in the area of front-end process modeling is the modeling of ultra-shallow junction formation, which starts from very low energy implant and especially focuses on the thermal annealing and diffusion of dopants. Due to the strongly reduced thermal budgets needed for shallow junctions that process is highly transient and is governed by the diffusion and reaction of dopant atoms and defects, and especially by the dynamics of clusters of these two. Implantation damage, amorphization, recrystallisation, and silicidation must be accurately simulated. In view of the need to increase carrier mobilities in the channel, the modeling of stress and strain and their influence on diffusion and activation has become vital, especially for strained silicon, SiGe, and for SOI structures. Model development, calibration and evaluation as well as process characterization require numerous experimental activities and large progress in the metrology for dopants, defects and stress, especially regarding two- and three-dimensional measurements.

COST-EFFECTIVE MANUFACTURING

SCALING OF MAXIMUM QUALITY DESIGN IMPLEMENTATION PRODUCTIVITY [DESIGN]

The number of available transistors scales by a factor of two every two years (DRAM) or a factor of two at each technology node (MPU), increasing design complexity as well. In order to maintain design quality even after process

technologies advance, design implementation productivity must be scaled to the same degree as design complexity is scaled. Improving design productivity and reusing the design are the key considerations for this issue. Namely, overall design productivity of quality- (difficulty-) normalized functions on-chip must scale at the rate of 2× per node. However, analog and mixed-signal design traditionally suffers from difficulty in improving design productivity and reusing the design along with process migration. There is a pressing need to develop a new design methodology to ameliorate those problems by implementing analog and mixed-signal synthesis, verification, and testing. Embedded software productivity also needs to be improved on a similar scale since the on-chip memory size is also growing and some functions are built into such embedded software rather than the hardware.

HIGH-SPEED SERIAL DEVICE INTERFACES [TEST AND TEST EQUIPMENT]

Penetration of high-speed interfaces into new designs is increasing dramatically. Loop-back alone may not be sufficient to achieve the needed product quality. Therefore test and design-for-test (DFT) methods must be developed to enable development and production testing. Drastically increasing pincount and number of test sockets are also major roadblocks and it is necessary to find ways of coping with them. Research and development is urgently required in order to overcome these obstacles and find cost-effective solutions. The trend toward faster high-speed serial interfaces and an increased port count will continuously drive the need for high-speed analog source/capture and jitter analysis instrument capability during characterization. Device interface circuitry must not degrade equipment bandwidth and accuracy, especially in the case of high-frequency differential I/O and analog circuits. Otherwise, it would introduce noise.

HIGHLY INTEGRATED DESIGNS, TEST IN SOCS AND SIPS [TEST AND TEST EQUIPMENT] [TEST AND TEST EQUIPMENT]

Customer requirements for form factor and power consumption are driving a significant increase in design integration levels. Test complexity will increase dramatically with the combination of different classes of circuits on a single die or within a single package. Disciplined, structured design-for-test (DFT) is required to reduce test complexity. In particular for SIP, increased focus on known good die (KGD) and sub-assembly test will be driven by the cost issue. Test of MEMS, optical devices, and other emerging or new devices to be integrated into SIP will also be an essential requirement. Manufacturing repair may be another requirement for non-stacked die.

FAILURE ANALYSIS AND DIAGNOSIS [TEST AND TEST EQUIPMENT]

Enhanced automated software diagnostic capabilities are required to improve physical failure analysis return-of-investment (ROI). Characterization capabilities must identify, locate, and distinguish individual defect types. Moreover accuracy and throughput must be improved. For example, throughput is expected to increase from days to hours. Failure analysis methods for analog devices constitute another critical issue. Design-for-test (DFT) is essential to localize failures because it can improve efficiency by reducing design complexities associated with testing. Defect types and behavior will continue to evolve with advances in fabrication process technology, and so fundamental research on existing and novel fault models to address emerging defects will be required.

TOOL COST AND R&D COST [LITHOGRAPHY]

Lithography has long accounted for a significant portion of overall semiconductor manufacturing costs, and this situation will become more marked. Shortening of the wavelength of the light source obliged us to use a new and costly lens material, CaF_2. The combination of lengthening R&D periods and shorter technology lifetimes is also pushing up the cost of lithography tools. On the other hand, possible extension of optical lithography down to hp45 nm, node with the use of immersion technology, will make the requirement for mask accuracy significantly difficult. Moreover, existence of multiple candidates for lithography tools in future generations leads to fragmentation of development resources. All these factors are predicted to significantly degrade ROI of lithography in future generations.

RESPONDING TO RAPIDLY CHANGING COMPLEX BUSINESS REQUIREMENTS [FACTORY INTEGRATION]

To correspond to customers' rapidly changing complex business requirements, various types of business models, such as integrated device manufacturer (IDM), fabless, foundry, joint venture, and out-sourcing, have emerged and become widespread. Factories now must integrate an even larger number of new and different items of equipment and software applications in a much shorter time while realizing high process reliability and volume productivity, which poses great challenges in factory integration. Construction of information exchange/control systems covering all the relevant fields, extending from design, mask, front-end-of-line (FEOL), and back-end-of-line (BEOL) to testing, packaging, etc., is crucial. The ability to model factory performance using various metrics is essential for optimizing the production output. Also, the achievement of higher visibility in manufacturing processes is important to find solutions that meet the increasing expectations of customers.

Meeting Process Requirements at hp65 nm and hp45 nm Nodes [Factory Integration]

How to construct a manufacturing system that guarantees product reliability at the hp65 nm and hp45 nm nodes is not clear. With reduced process windows and increasingly difficult targets, process control will becomes quite difficult in many process modules. Integration of next-generation lithography into a factory leads to unwieldy complexity in the factory design. A novel streamlined system that controls various parameters for stabilizing fabrication processes and achieving satisfactory product quality, while realizing shortening of the cycle time of products, should be pursued.

Tools and Methodologies to Address Chip and Package Co-Design [Assembly and Packaging]

Assembly and packaging have become critically important and thus are factors influencing the competitiveness of LSI technology, because they have significant effects on operating frequency, power consumption, complexity, form factor, reliability, and cost of final products. Concurrent design at both the chip and package levels is essential to satisfy stringent requirements for the system, including design cycle time. To achieve this, establishment of simulation tools and methodologies, which can accurately predict electrical characteristics, thermal dissipation, and thermo-mechanical stress while simultaneously considering physical layout, cost, and environmental impact, is required. This modeling and simulation can save time and cost significantly by dispensing with the need for extensive experiments. This requirement is especially strong in RF and mixed-signal applications, where inductance and capacitance of interconnections are important design parameters, and low-cost and high bandwidth products are emerging one after another with extremely short time-to-market requirements. Support from commercial electronic design automation (EDA) suppliers is indispensable.

Chemical and Material Assessments [ESH]

The rapid introduction of new chemicals/materials/processes requires new rapid assessment methodologies to ensure that new chemicals/materials can be utilized in manufacturing without inducing new hazardous impacts on human health, safety, and the environment. Although methodologies are needed to meet the evaluation and quantification demands for ESH impacts, the focus is currently on expediting process implementation.

Resource Conservation [ESH]

As the industry grows and its technology advances toward finer patterning and larger wafer sizes, the natural tendency is toward increased use of water, energy, chemicals, and materials. Resource conservation is becoming a major concern with respect to availability, cost reduction, manufacturing location, sustainability, and waste disposal. Thus, it is necessary to develop diverse process equipment capable of utilizing resources efficiently.

Design for Manufacture and Test and Systematic Yield [Yield Enhancement]

IC designs must be optimized for a given process capability and must be testable and diagnosable. Understanding Systematic Mechanism Limited Yield (SMLY) is mandatory for achieving historic yield ramps in the future. Design to process compatibility, design for manufacturability, design for test, and design for diagnosability would be perfected through systematic mechanisms limited yield (SMLY) model development.

High-Aspect-Ratio Inspection [Metrology and Yield Enhancement]

Control of high-aspect-ratio technologies such as Damascene challenges all metrology methods. Key requirements are void detection in copper lines and pore size distribution in patterned low κ. The need is to have a rapid, in-line observation of a very small number of voids/larger pores. Critical dimension measurements are also required for very high-aspect-ratio structures that are made from porous dielectric materials and require 3D information for trench and via/contact sidewalls. These measurements will be further complicated by the underlying multi film complexity. The detection of via defects near/at the bottom of a Damascene trench will also continue to be a grand challenge. However, the challenge is complicated by the simultaneous need for high sensitivity and high throughput. High-speed, cost-effective detection tools that satisfy both demands are therefore needed.

Non-Visual Defect Sourcing and Manufacture and Test Oriented Design [Yield Enhancement]

Fault isolation complexity is expected to grow exponentially, combining the difficult tasks of defining fault dimensions in the horizontal plane and vertical layers. It is especially difficult to analyze circuit failures that leave no detectable physical remnant. Accordingly, new analysis tools and techniques that can isolate those non-visual failures are needed. Although IC current design is optimized for a given process capability and is testable/diagnosable, many defects that cause electrical faults are still not detectable in-line. Tools are needed that enable design and process matching so that optimum yields can be achieved.

Factory Level and Company Wide Metrology Integration [Metrology]

Real time *in situ*, integrated, and in-line metrology is required for manufacturing complicated devices. Continued development of robust sensors, process controllers, and data management will allow integration of add-on sensors. Standards for process controllers and data management must be agreed upon. Massive quantities of raw data should be converted to information useful for enhancing the yield of a semiconductor manufacturing process. Better sensors must be developed for trench etch end point, ion species/energy/dosage (current), and wafer temperature during RTA.

Starting Materials Manufacturing Metrology [Metrology]

The introduction of new substrates such as SOI affects impurity detection (especially particles) at levels of interest and edge exclusion for metrology tools. In particular, CD, film thickness, and defect detection are influenced by thin SOI optical properties and charging by electron and ion beams. SiGe and strained silicon layers add to the complexity of this challenge. Existing capabilities will not meet Roadmap specifications. Very small particles must be detected and properly sized. Capabilities for SOI wafers, SiGe, and strained silicon need enhancement. Challenges originate from the extra optical reflection in SOI and the surface.

In The Long Term (~2010 through 2018)

Enhancing Performance

Implementation of Advanced, Non-Classical CMOS Device with Enhanced Drive Current [Process Integration, Devices, and Structures]

To continue MOSFET scaling to less than Lg = 20 nm, it is quite likely that the device structure will change to advanced non-classical CMOS such as multiple-gate, ultrathin body (UTB) MOSFETs. In these devices, various "technology boosters," such as mobility enhancement by strained Si, elevated source source/drain, high-κ gate dielectric, and metal gate electrode, will likely be simultaneously implemented with the new device structure. In UTB MOSFETs having less than 10 nm Si thickness, various quantum effects will impact the electric characteristics. Toward the end of the Roadmap, devices will increasingly be operated in the quasi-ballistic mode, where the current gain will be enhanced by parameters different from those currently known. Choice of the optimum device structures, their physical characterization, and construction of cost-effective processing flows will become very important along with construction of their circuit architecture.

Identification, Selection, and Implementation of Advanced, Beyond-CMOS Devices [Process Integration, Devices, and Structures]

It is not at all clear what novel device structure will emerge in information processing beyond the end of the CMOS Roadmap. However, it is forecasted that CMOS architecture will continue to be used as a technology platform, and the beyond-CMOS devices will be integrated either functionally or technologically with this platform. This prediction is based on the observation that no known beyond-CMOS device can completely replace CMOS in terms of speed, power, and cost performance. This observation is especially true if we confine the discussion to room temperature operation. Therefore, the novel devices are expected to significantly extend system performance beyond that attainable with CMOS alone.

Post-Optical Lithography and Mask Technology [Lithography]

Though immersion technology shows possibility of extending optical lithography down to the hp45 nm node, lithography for nodes beyond that is controversial. Candidates for the post-optical lithography are extreme ultraviolet lithography (EUV), electron projection lithography (EPL), mask-less lithography (ML2), and imprint technology that is newly added in the 2003 version. Since all these technologies require development of a totally new infrastructure, the R&D costs will boost the overall manufacturing costs. Regarding mask technology, development of defect-free masks, overlay metrology, defect inspection, and repair will become critical.

Identify Solutions Which Address Global Wiring Scaling Issues [Interconnect]

Traditional interconnect scaling will no longer satisfy performance requirements. Defining and finding solutions beyond copper and low κ will require material innovation, combined with accelerated design, packaging, and unconventional interconnect. Novel interconnect schemes include 3D interconnect, RF/microwave, optical interconnects, etc.

NOISE MANAGEMENT [DESIGN]

Since the operating voltage decreases 20% per technology node, increasing noise sensitivity is becoming an important issue in the design of functional devices (e.g., bits, transistors, gates) and products (such as DRAMs or MPUs). This sensitivity is becoming more evident due to lower noise headroom in low-power devices, coupled interconnects, IR drop and ground bounce in the supply voltage, thermal impact on device off-currents and interconnect resistivities, mutual inductance, substrate coupling, single-event upset (alpha particle), and increased use of dynamic logic families. Consequently, modeling, analysis, and estimation must be performed at all design levels.

COST-EFFECTIVE MANUFACTURING

ERROR-TOLERANT DESIGN [DESIGN]

The scaling of the design complexity and the increasing transistor count will greatly increase the potential for failures to occur. In this case, relaxing the requirement for 100% correctness in both transient and permanent failures of signals, logic values, devices, or interconnects may reduce the cost of manufacturing, verification and testing. Potential solutions are adaptive and self-correcting/self-repairing circuits and the use of on-chip re-configurability.

STARTING MATERIALS ALTERNATIVES GREATER THAN 300 MM [FRONT END PROCESSES]

The requirement of continued productivity enhancement dictates the need for a next-generation, large-area starting substrate material. Historical trends suggest that the new starting material should have nominally twice the area of current-generation substrates, such as 450 mm. However, it is questionable whether or not the conventional Czochralski crystal pulling, wafer slicing, and polishing process can be employed beyond 300 mm. It is possible that the alternative substrate will not be bulk silicon but silicon-on-insulator (SOI). In any event, research is required on cost-effective 450 mm diameter wafers. If they are to be available for production in 2011 as currently forecast, wafer-manufacturing development should be implemented eight years earlier, in 2003.

CHEMICAL AND MATERIAL MANAGEMENT BY ESH DESIGN AND MEASUREMENT METHODS [ESH]

Equipment design engineers and equipment users require timely information regarding ESH characteristics of potential new process chemicals and materials. This information is essential to the proper selection of optimal chemicals and materials for function and ESH impact with respect to reaction product emissions, health and safety properties, compatibility of materials with both equipment and other chemical components, flammability, and reactivity. It must be possible to do so while minimizing unnecessary impacts on business after processes are developed and are in production. For integrated ESH design and measurement methods, a methodology for determining the lowest ESH impact of materials and processes needs to be developed.

YIELD MODEL DEVELOPMENT FOR NEW MATERIALS AND INTEGRATION [YIELD ENHANCEMENT]

Defect budgets will require frequent revalidation and updates as information about future processing technologies becomes available. Yield models need to consider complex integration issues with respect to random defect-limited yield as well as systematic limited yield (such as parametric yield loss or circuit yield loss) for future technology nodes. As a result, the models must take into account greater parametric sensitivities, complex integration issues, ultra-thin film integrity, impact of circuit design, and greater transistor packing density.

NON-DESTRUCTIVE PRODUCTION MEASUREMENTS [METROLOGY]

As is well known, surface charging and contamination interfere with electron beam imaging. CD measurements must account for the sidewall shape. In addition, CD for the Damascene process may require measurement of trench structures. On the other hand, process control such as focus exposure and etch bias will require greater precision and 3D capability. As a result, non-destructive (without charging or contaminating the surface) wafer/mask level microscopy for measuring the critical dimensions of 3D structures, overlay, and defect detection is required. Furthermore, analysis of 3D structures such as tapered- or undercut-gate electrodes, trenches, high-aspect-ratio capacitors, and contacts is needed.

MODELING OF PROCESSING AND ELECTRICAL PROPERTIES OF NEW MATERIALS [MODELING AND SIMULATION]

Increasingly new materials need to be introduced in technology development due to physical limits which otherwise would prevent further scaling. This is required especially for gate stacks, interconnect structures, and photoresists. In consequence, equipment, process, device and circuit models must be extended to include these new materials. Furthermore, computational material science needs to be developed and applied to contribute to the assessment and selection of new materials in order to reduce the experimental effort. This Grand Challenge crosscuts most of the Difficult Challenges in the Modeling and Simulation chapter.

DIFFICULT CHALLENGES

OF THE INTERNATIONAL TECHNOLOGY WORKING GROUPS

SYSTEM DRIVERS AND DESIGN

Challenges ≥50 nm/Through 2009	*Summary Of Issues*
Silicon Complexity: devices and interconnects	All—Exponential increase in leakage power S, P, A—Power density and distribution All—Technology and library characterization S, P, A—High-frequency noise analysis S, P, A—Transmission-line interconnects All—eDRAM, eFPGA, SiGe, optical, MEMS
System Complexity: number of states, design diversity	S, P—Verifying systems with exploding number of states S, P, A—Concurrent multi-factor analysis and optimization S, P—Scalable algorithms S, A—Design and test of mixed analog and digital designs All—Complex package analysis S, A—Integrating A/D tools
Design Productivity	S—Integrating third party components S, P—Design tool interoperability S, P—Early analysis and verification methods
Time-to-Market	S, A—Support for platform-based design S, P, A—Exploiting parallel processing
Challenges <50 nm/Beyond 2009	
Manufacturability	All—Cross-chip variability All—Sub-wavelength mask correction (AltPSM, OPC, RET) All—Design for yield and manufacturability All—Standards for sharing design and manufacturing data
System-level Design	S—Common hardware/software (HW/SW) representation S—SW synthesis, HW/SW optimization

This table summarizes challenges to the design process advances implied by the above four trends. Each challenge is labeled with a list of the most relevant system drivers (S—system on chip, P—microprocessor, A—analog/mixed-signal, M—memory).

TEST AND TEST EQUIPMENT

Five Difficult Challenges ≥45 nm/Through 2010	*Summary of Issues*
High-speed Device Interfaces	A major roadblock will be the need for high frequency, high pin count probes and test sockets; research and development is urgently required to enable cost-effective solutions with reduced parasitic impedance. High-speed serial interface speed and port count trends will continue to drive high-speed analog source/capture and jitter analysis instrument capability for characterization. DFT/DFM techniques must be developed for manufacturing. Device interface circuitry must not degrade equipment bandwidth and accuracy, or introduce noise; especially for high-frequency differential I/O and analog circuits.
Highly Integrated Designs	Highly structured DFT approaches are required to enable test access to embedded cores. Individual cores require special attention when using DFT and BIST to enable test. Analog DFT and BIST techniques must mature to simplify test interface requirements and slow ever increasing instrument capability trends. Testing chips containing RF and audio circuits will be a major challenge if they also contain large numbers of noisy digital circuits. DFT must enable test reuse for reusable design cores to reduce test development time for highly complex designs.
Reliability Screens	Existing methodologies are limited (burn-in versus thermal runaway, IDDQ versus background current increases). Research is required to identify novel infant mortality defect acceleration stress conditions
Manufacturing Test Cost	Test cell throughput enhancements are needed to reduce manufacturing test cost. Opportunities include massively parallel test, wafer-level test, wafer-level burn-in, and others. Challenges include device interfacing/contacting, power and thermal management. Device test needs must be managed through DFT to enable low-cost manufacturing test solutions; including reduced pin count test, equipment reuse, and reduced test time. Automatic test program generators are needed to reduce test development time. Test standards are required to enable test content reuse and manufacturing agility.
Modeling and Simulation	Logic and timing accurate simulation of the ATE, device interface, and DUT is needed to enable pre-silicon test development and minimize costly post-silicon test content development/debug on expensive ATE. High-performance digital and analog I/O and power requirements require significant improvements to test environment simulation capability to ensure signal accuracy and power quality at the die. Equipment suppliers must provide accurate simulation models for pin electronics, power supplies, and device interfaces to enable interface design.

Test and Test Equipment Difficult Challenges (continued)

Five Difficult Challenges <45 nm/Beyond 2010	*Summary of Issues*
DUT to ATE interface	Probing capability for optical and other disruptive technologies Support for massively parallel test—including full wafer contacting Decreasing die size and increasing circuit density are driving dramatic increases in die thermal density. This problem is further magnified by the desire to enable parallel test to maximize manufacturing throughput. New thermal control techniques will be needed for wafer probe and component test. DFT to enable test of device pins not contacted by the interface and test equipment.
Test Methodologies	New DFT techniques (SCAN and BIST have been the mainstay for over 20 years). New test methods for control and observation are needed. Tests will need to be developed utilizing the design hierarchy. Analog DFT and BIST techniques must mature to simplify test interface requirements and slow ever increasing instrument capability trends Logic BIST techniques must evolve to support new fault models, failure analysis, and deterministic test. EDA tools for DFT insertion must support DFT selection with considerations for functionality, coverage, cost, circuit performance and ATPG performance.
Defect Analysis	Defect types and behavior will continue to evolve with advances in fabrication process technology. Fundamental research in existing and novel fault models to address emerging defects will be required. Significant advances in EDA tools for ATPG capacity and performance for advanced fault models and DFT insertion are required to improve efficiency and reduce design complexities associated with test.
Failure Analysis	Realtime analysis of defects in multi-layer metal processes is needed. Failure analysis methods analog devices must be developed and automated. Transition from a destructive physical inspection process to a primarily non-destructive diagnostic capability. Characterization capabilities must identify, locate, and distinguish individual defect types.
Disruptive Device Technologies	Develop new test methods for MEMS and sensors. Develop new fault models for advanced/disruptive transistor structures.

PROCESS INTEGRATION, DEVICES, AND STRUCTURES (INCLUDING RF AND MIXED-SIGNAL TECHNOLOGIES FOR WIRELESS COMMUNICATIONS AND EMERGING RESEARCH DEVICES)

Difficult Challenges ≥45 nm/Through 2010	*Summary of Issues*
1. High-performance applications: meeting performance and power dissipation requirements for highly scaled MOSFETs	Cost effectiveness, process control, and reliability of very thin oxy-nitride gate dielectrics, especially considering the high gate leakage Difficulty in controlling short-channel effects for highly scaled devices Negative impact of high channel doping needed for highly scaled devices. Also, the difficulty in controlling threshold voltage due to statistical fluctuations in the doping Need to reduce series S/D parasitic resistance Controlling static power dissipation in the face of rapidly increasing leakage: architecture and circuit design improvement and innovation will be needed.
2. Low-power applications: meeting performance and leakage requirements for highly scaled MOSFETs	Early availability of manufacturing-worthy high-κ gate dielectrics is necessary to meet stringent gate leakage and performance requirements. Slow scaling of V_{dd} for low standby power logic will make overall device scaling difficult. Rapid scaling of V_{dd} for low operating power logic will make overall device scaling difficult.
3. Assuring the reliability and implementing into manufacturing of multiple material, process, and structural changes in a relatively short period of time	Multiple material changes projected: high-κ gate dielectric, metal gate electrodes, strained Si, nickel silicide by 2008 or so Elevated S/D (selective epi) Ultra-thin body (UTB) SOI by 2008 or so, followed by multiple-gate structures. Near mid-gap metal gate electrodes will be desirable to set the threshold voltage for UTB SOI. Difficulty in ensuring reliability of all these new materials, processes, and structures in a timely manner
4. Implementation of DRAM, SRAM, and high-density nonvolatile memory (NVM) for scaled technologies	DRAM main issues—adequate storage capacitance for devices with reduced feature size, including difficulties in implementing high-κ storage dielectrics; access device design; holding the overall leakage to acceptably low levels; and deploying low sheet resistance materials for bit and word lines to ensure desired speed for scaled DRAMs SRAM—Difficulties with maintaining adequate noise margin and controlling key instabilities with scaling. Also, difficult lithography and etch issues with scaling NVM, Flash—Scaling of tunnel dielectric and interpoly dielectric involves many complex tradeoffs. Dielectric material properties and dimensional control are key issues NVM, FeRAM—Ferroelectric material properties and dimensional control. Sensitivity to IC processing temperatures and conditions NVM, SONOS—ONO stack dimensions and material properties, including nitride layer trap distribution in space and energy NVM, MRAM—Magnetic material properties and dimensional control. Sensitivity to IC processing temperatures and conditions
5. High-performance and low-cost RF and analog/mixed-signal solutions	Signal isolation Optimizing RF/analog CMOS devices with scaled technologies: mismatch, 1/f noise, and leakage with high-κ gate dielectrics High-density integrated passive element scaling and use of new materials: Q-factor value for inductors; matching and linearity for capacitors Reduced power supply voltages: degradation in SNR (signal-to-noise ratio) and signal distortion performance Reduced device breakdown voltage in scaled technologies High-frequency devices with increased operating voltage for base station applications Compound semiconductor substrates with good thermal dissipation and process equipment for fabrication at low cost See section on RF and A/MS Technologies for Wireless Communications for detailed discussion of these issues.

Process Integration Difficult Challenges (continued)

Difficult Challenges <45 nm/Beyond 2010	*Summary of Issues*
6. Implementation of advanced, non-classical CMOS with enhanced drive current and acceptable control of short channel effects for highly scaled MOSFETs	Advanced non-classical CMOS (e.g., multiple-gate, ultra-thin body [UTB] MOSFETs) with lightly doped body will be needed to effectively scale MOSFETs to well under 20 nm gate length (L_g). Most likely, advanced material solutions such as strained Si (enhanced mobility) channels, elevated source/drain, high-κ gate dielectric, metal gate electrode, etc., will be utilized along with the advanced non-classical CMOS Particularly for the *highly scaled* UTB MOSFETs required towards the end of the Roadmap, with body thickness well under 10 nm, electrical performance and the impact of quantum effects are not well understood To attain adequate drive current for the highly scaled MOSFETs, quasi-ballistic operation with enhanced carrier saturation velocity appears to be needed See Emerging Research Devices section for more detail.
7. Dealing with atomic-level fluctuations and statistical process variations in sub-20 nm MOSFETs	Fundamental issues of atomic-level statistical fluctuations for sub-20 nm MOSFETs are not completely understood, including the impact of quantum effects.
8. Identifying, selecting, and implementing new memory structures	Highly scaled, dense, fast, non-volatile memory will become highly desirable Increasing difficulty is expected in scaling DRAMs, especially scaling down the dielectric equivalent oxide thickness, attaining the very low leakage currents that will be required, and reducing the cell area factor All of the existing forms of nonvolatile memory face limitations based on material properties. Success will hinge on finding and developing alternative materials and/or development of alternative emerging technologies. See Emerging Research Devices section for more detail.
9. Identifying, selecting, and implementing novel interconnect schemes	Eventually, it is projected that the performance of copper/low-κ interconnect will become inadequate to meet the speed and power dissipation goals of highly scaled ICs. Solutions (optical, microwave/RF, etc,) are currently unclear.
10 Toward the end of the Roadmap or beyond, identification, selection, and implementation of advanced, beyond-CMOS devices and architectures for advanced information processing	Will drive major changes in process, materials, device physics, design, etc. Performance, power dissipation, etc., of beyond-CMOS devices need to extend well beyond CMOS limits. Beyond-CMOS devices need to integrate into a CMOS platform. Integration of the two may be difficult, especially for mixed signal. See Emerging Research Devices sections for more discussion and detail.

FRONT END PROCESSES

MPU/ASIC Physical Gate Length ≥20 nm/ Through 2009	*Summary of Issues*
New gate stack processes and materials	Extension of oxynitride gate dielectric materials to < 1.0 nm E.O.T for high-performance MOSFETs, consistent with device reliability requirements Introduction and process integration of high-κ gate stack materials and processes for high-performance, low operating and low standby power MOSFETs CMOS integration of enhanced channel mobility, e.g., strained layers Control of boron penetration from doped polysilicon gate electrode Minimized depletion of dual-doped polysilicon electrodes Introduction of dual metal gate electrodes with appropriate work function (toward end of period) Metrology issues associated with gate dielectric film thickness and gate stack electrical and materials characterization
Critical dimension and effective channel length (L_{eff}) control	Control of gate etch processes that yield a physical gate length that is smaller than the feature size printed in the resist, while maintaining <10% overall 3-sigma control of the combined lithography and etch processes Control of profile shape, edge roughness, line and space width for isolated as well as closely-spaced fine line patterns Control of self-aligned doping processes and thermal activation budgets to achieve L_{eff} control Maintenance of CD and profile control throughout the transition to new gate stack materials and processes CD and etch metrology Site flatness to ensure effective lithographic printing
Introduction and CMOS integration of new memory materials and processes	Development and introduction of very high-κ DRAM capacitor dielectric layers Migration of DRAM capacitor structures from silicon-insulator-metal to metal-insulator-metal Integration and scaling of FeRAM ferroelectric materials Scaling of Flash inter-poly and tunnel dielectric layers may require high κ Limited temperature stability of high-κ and ferroelectric materials challenges CMOS Integration
Surfaces and interfaces—structure, composition, and contamination control	Contamination, composition, and structure control of channel/gate dielectric interface as well as gate dielectric/gate electrode interface Interface control for DRAM capacitor structures Maintenance of surface and interface integrity through full-flow CMOS processing Statistically significant characterization of surfaces having extremely low defect concentrations for starting materials and pre-gate clean surfaces
Scaled MOSFET dopant introduction and control	Doping and activation processes to achieve shallow source/drain regions having parasitic resistance that is less than ~17–33% of ideal channel resistance ($=V_{dd}/I_{on}$) Control of parasitic capacitance to achieve less than ~23–29% of gate capacitance, consistent with acceptable Ion and minimum short channel effect Achievement of activated dopant concentration greater than solid solubility in dual-doped polysilicon gate electrodes Formation of continuous self-aligned silicide contacts over shallow source and drain regions Metrology issues associated with 2D dopant profiling

Front End Processes Difficult Challenges (continued)

MPU/ASIC physical Gate Length <20 nm/Beyond 2009	*Summary of Issues*
Continued scaling of planar CMOS devices	Higher κ gate dielectric materials including temperature constraints Metal gate electrodes with appropriate work function Sheet resistance of clad junctions CD and L_{eff} control Chemical, electrical, and structural characterization
Introduction and CMOS integration of non-standard, double gate MOSFET devices	Devices are needed starting from 2011 and may be needed as early as 2007 (this is a backup for high-κ materials and metal gates on standard CMOS) Selection and characterization of optimum device types CMOS integration with other devices, including planar MOSFETs Introduction, characterization, and production hardening of new FEP unit processes Device and FEP process metrology Increased funding of long term research Role of SOI utilization (and structural configuration) for advanced non-classical CMOS
Starting silicon material alternatives greater than 300 mm diameter require the start of wafer manufacturing development in year 2003	Need for future productivity enhancement dictates the requirement for a next generation, large silicon substrate material Historical trends suggest that the new starting material have nominally twice the area of present generation substrates, e.g., 450 mm Economies of the incumbent Czochralski crystal pulling, wafer slicing, and polishing processes are questionable beyond 300 mm; research is required for a cost-effective substrate alternative to bulk silicon If 450 mm wafers are to become available for production in 2011 as currently forecasted, wafer-manufacturing development should be implemented eight years earlier, e.g., 2003.
New memory storage cells, storage devices, and memory architectures	Scaling of DRAM storage capacitor beyond $6F^2$ Further scaling of Flash memory interpoly and tunnel oxide thickness FeRAM storage cell scaling Introduction of new memory types and storage concepts (Candidates—MRAM, phase-change memory for 2010, and single electron, molecular, nano-floating products beyond 2010)
Surface and interface structural, contamination, and compositional control	Achievement and maintenance of structural, chemical, and contamination control of surfaces and interfaces that may be horizontally or vertically oriented relative to the chip surface Metrology and characterization of surfaces that may be horizontally or vertically oriented relative to the chip surface Achievement of statistically significant characterization of surfaces and interfaces that may be horizontally or vertically oriented relative to the chip surface

LITHOGRAPHY

Five Difficult Challenges ≥50 nm/Through 2009	*Summary of Issues*
Optical Masks with Features for Resolution Enhancement and Post-optical Mask Fabrication	Registration, CD control, defectivity, and 157 nm pellicles; defect free multi-layer EUV substrates or EPL membrane masks Equipment infrastructure (writers, inspection, repair
Cost Control and Return on Investment (ROI)	Achieving constant/improved ratio of tool cost to throughput over time Cost-effective resolution enhanced optical masks and post-optical masks Sufficient lifetimes for the technologies Resources for developing multiple technologies at the same time High-output, cost-effective, EUV light source
Process Control	Processes to control gate CDs to less than 1.8 nm (3 sigma) New and improved alignment and overlay control methods independent of technology option to < 19 nm overlay Accuracy of OPC
Resists for ArF, Immersion Lithography, and F2	Outgassing, LER, SEM-induced CD changes, defects ≥30 nm.
CaF^2	Yield, cost, quality
Five Difficult Challenges <45 nm/Beyond 2010	
Mask Fabrication and Process Control	Defect-free NGL masks Equipment infrastructure (writers, inspection, repair) Mask process control methods
Metrology and Defect Inspection	Capability for critical dimensions down to 7 nm and metrology for overlay down to 7.2 nm, and patterned wafer defect inspection for defects <30 nm
Cost Control and ROI	Achieving constant/improved ratio of tool cost to throughput Development of cost-effective post-optical masks Achieving ROI for industry with sufficient lifetimes for the technologies
Gate CD Control Improvements, Process Control, and Resist Materials	Development of processes to control gate CDs <1 nm (3 sigma) with appropriate line-edge roughness Development of new and improved alignment and overlay control methods independent of technology option to <7.2 nm overlay
Tools for Mass Production	Post optical exposure tools capable of meeting requirements of the Roadmap

SEM—scanning electron microscope

INTERCONNECT

Five Difficult Challenges ≥45 nm/Through 2009	*Summary of Issues*
Introduction of new materials to meet conductivity requirements and reduce the dielectric permittivity*	The rapid introductions of new materials/processes that are necessary to meet conductivity requirements and reduce the dielectric permittivity create integration and material characterization challenges.
Engineering manufacturable interconnect structures compatible with new materials and processes*	Integration complexity, CMP damage, resist poisoning, dielectric constant degradation. Lack of interconnect/packaging architecture design optimization tool
Achieving necessary reliability	New materials, structures, and processes create new chip reliability (electrical, thermal, and mechanical) exposure. Detecting, testing, modeling and control of failure mechanisms will be key.
Three-dimensional control (3D CD) of interconnect features (with it's associated metrology) is required to achieve necessary circuit performance and reliability.	Line edge roughness, trench depth and profile, via shape, etch bias, thinning due to cleaning, CMP effects. The multiplicity of levels combined with new materials, reduced feature size, and pattern dependent processes create this challenge.
Manufacturability and defect management that meet overall cost/performance requirements	As feature sizes shrink, interconnect processes must be compatible with device roadmaps and meet manufacturing targets at the specified wafer size. Plasma damage, contamination, thermal budgets, cleaning of high A/R features, defect tolerant processes, elimination/reduction of control wafers are key concerns. Where appropriate, global wiring and packaging concerns will be addressed in an integrated fashion.
Five Difficult Challenges <45 nm/Beyond 2009	*Summary of Issues*
Mitigate impact of size effects in interconnect structures	Line and via sidewall roughness, intersection of porous low−κ voids with sidewall, barrier roughness, and copper surface roughness will all adversely affect electron scattering in copper lines and cause increases in resistivity
Three-dimensional control (3D CD) of interconnect features (with it's associated metrology) is required	Line edge roughness, trench depth and profile, via shape, etch bias, thinning due to cleaning, CMP effects. The multiplicity of levels, combined with new materials, reduced feature size and pattern dependent processes, use of alternative memories, optical and RF interconnect, continues to challenge.
Patterning, cleaning, and filling at nano dimensions	As features shrink, etching, cleaning, and filling high aspect ratio structures will be challenging, especially for low-κ dual-Damascene metal structures and DRAM at nano dimensions.
Integration of new processes and structures, including interconnects for emerging devices	Combinations of materials and processes used to fabricate new structures create integration complexity. The increased number of levels exacerbate thermomechanical effects. Novel/active devices may be incorporated into the interconnect.
Identify solutions which address global wiring scaling issues*	Traditional interconnect scaling will no longer satisfy performance requirements. Defining and finding solutions beyond copper and low κ will require material innovation, combined with accelerated design, packaging and unconventional interconnect.

**Three top challenges for Interconnect*

FACTORY INTEGRATION

Difficult Challenges ≥45 nm/Through 2009	*Summary of Issues*
Responding to Rapidly Changing, Complex Business Requirements	Many new and co-existing business models including IDM, Foundry, Joint Ventures, Collaborations, other Outsourcing, etc Increased expectations by customers for faster delivery of new and volume products Need for improve integration of the entire product design and manufacturing process Faster design -> prototype and prototype -> volume production Enhanced customer visibility into outsourced production operations Reduced time to ramp factories, products, and processes to stay competitive within the rapidly changing business environment Building 30+ mask layer System on a Chip (SoC) with long cycle times where needs are rapidly changing Rapid and frequent factory plan changes driven by changing business needs Ability to model factory performance so that output can be optimized Ability to constantly adjust equipment loading to keep the factory profitable Need to run globally disparate factories as single "virtual factory"
Achieving Growth Targets while Margins are Declining	Implications of rising wafer, packaging, and other materials cost on meeting cost targets Meeting high factory yield much faster at startup Addressing increased complexity while keeping costs in check Reducing complexity and waste across the supply chain Inefficiencies introduced by non-product wafers (NPW) competing for resources with production wafers High cost and cycle time of mask sets for manufacturers impacting affordability of new product designs Increasing dedication of masks and equipment causing manufacturing inefficiencies Challenges introduced with sharing of mask sets Difficulty maintaining the historical 0.7× transistor shrink per year for die size and cost efficiency
Managing Ever Increasing Factory Complexity	Quickly and effectively integrating rapid changes in process technologies Managing carriers with multiple lots, wafers with multiple products, or multiple package form factors Comprehending increased purity requirements for process and materials Need to run aluminum and copper back end in the same factory Increasing number of processing steps coupled with process and product complexity Need to concurrently manage new and legacy software systems and systems with increasingly high interdependencies Explosive growth of data collection/analysis requirements driven by process and modeling needs Increased requirements for wafer level tracking and die level tracking
Meeting Factory and Equipment Reliability, Capability or Productivity Requirements per the Roadmap	Process equipment not meeting availability, run rate, and utilization targets out of the box Stand alone and integrated reliability for equipment and systems to keep factories operating Increased impacts that single points of failure have on a highly integrated and complex factory Quality issues with production equipment embedded controllers Lack of good data to measure equipment and factory effectiveness for optimization and improvement programs Factory capacity planning and supply chain management systems are not continuously base lined with actual factory data creating errors Lack of migration paths which inhibit movement from old inefficient systems to new highly productive systems

Factory Integration Difficult Challenges (continued)

Difficult Challenges ≥45 nm/Through 2009	*Summary of Issues*
Meeting the Flexibility, Extendibility, and Scalability Needs of a Cost-effective, Leading-edge Factory	Need to quickly convert factories to new process technologies while reusing equipment, facilities, and skills Minimizing downtime to on-going operations while converting factories to new technologies Scalability implications to meet large 300 mm factory needs [50K WSPM] Continued need to improve both throughput and cycle time Reuse of building, production and support equipment, and factory information and control systems across multiple technology nodes Understanding up-front costs to incorporate EFS Ability to convert 200 mm facilities to 300 mm wafer size Comprehending increased purity requirements for process and materials Accelerating the pace of standardization to meet industry needs
Meeting Process Requirements at 65 nm and 45 nm Nodes Running Production Volumes	Small process windows and tight process targets at 65 nm and 45 nm nodes in many modules make process control increasingly difficult Complexity of integrating next generation lithography equipment into the factory Overall development and volume production timelines continuing to shrink Device and process complexity make the ability to trace functional problems to specific process areas more difficult Difficulty in running different process parameters for each wafer while maintaining control windows and cycle time goals Reducing the impacts of parametric variation
Increasing Global Restrictions on Environmental Issues	Need to meet regulations in different geographical areas Need to meet technology restrictions in some countries while still meeting business needs Comprehending tighter ESH/Code requirements Lead free and other chemical and materials restrictions New material introduction
Difficult Challenges <45 nm/2010 Through 2018	*Summary of Issues*
Post-conventional CMOS Manufacturing Uncertainty	Uncertainty of novel device types replacing conventional CMOS and the impact of their manufacturing requirements will have on Factory design Timing uncertainty to identify new devices, create process technologies, and design factories in time for a low risk industry transition Potential difficulty in maintaining an equivalent 0.7× transistor shrink per year for given die size and cost efficiency Need to run CMOS and post CMOS processes in the same factory
Emerging Factory Paradigm and Next Wafer Size Change	Uncertainty about the next wafer size [450 mm] and the conversion timing Traditional strategies to scale wafers and carriers for the next wafer size conversion may not work with [450 mm] 25 wafer carriers and drive significant production equipment and material handling changes Uncertainty concerning how to reuse buildings, equipment, and systems to enable the next wafer size conversion [to 450 mm] at an affordable cost

ASSEMBLY AND PACKAGING

Difficult Challenges ≥ 45 nm/Through 2010	*Summary of Issues*
Improved Organic Substrates	Tg compatible with Pb free solder processing Increased wireability at low cost Improved impedance control and lower dielectric loss to support higher frequency applications Improved planarity and low warpage at higher process temperatures Low-moisture absorption Low-cost embedded passives Substrate cost is barrier to flip chip wide spread adoption today Increased via density in substrate core Alternative plating finish to improve reliability
Improved Underfills for Flip Chip on Organic Substrates	Thermal performance and thermal coupling between parts Materials which enable integration of SMT, varying semiconductors, and substrate types reliably Thin die, stack die, very large and very small die, passives component integration, SAW, shielding interconnect process Narrowing gaps Higher bump densities
Coordinated Design Tools and Simulators to address Chip, Package, and Substrate Co-design	Mix signal co-design and simulation environment Integrated analysis tools for transient thermal analysis and integrated thermal mechanical analysis Electrical (power disturbs, EMI, signal integrity associated with higher frequency/current and lower voltage switching) Commercial EDA supplier support System level co-design is needed now. EDA support for “native” area array is required to meet the Roadmap projections. Educational programs required to train engineers in these technologies/requirements.
Impact of Cu/low κ on Packaging	Direct wirebond and bump to Cu Bump and underfill technology to assure low-κ dielectric integrity Improved mechanical strength of dielectrics Interfacial adhesion Reliability of first level interconnect with low κ Mechanisms to measure the critical properties need to be developed. Probing over copper /low κ due to damage and bonding over probe mark
High Current Density Packages	Electromigration will become a more limiting factor. It must be addressed through materials changes together with thermal/mechanical reliability modeling. Whisker growth Thermal dissipation

Assembly and Packaging Difficult Challenges (continued)

Difficult Challenges <45 nm/Beyond 2010	*Summary of Issues*
Package Cost does not follow the Die Cost Reduction Curve	Margin in packaging inadequate to support investment required to reduce cost
Small Die with High Pad Count, High Power Density, and/or High Frequency	Current density, operating temperature, etc for these devices exceed the capabilities of current assembly and packaging technology
High Frequency Die	Substrate wiring density to support >20 lines/mm Lower loss dielectrics– Skin effect above 10 GHz "Hot spot" thermal management needs to be addressed before 2007. There is a "brick wall at five-micron lines and spaces. Design TWG would like to have an upper bound on thermal management capability of future packages.
Close Gaps between Substrate Technology and the Chip	Interconnect density scaled to silicon (silicon I/O density increasing faster than the package substrate technology Production techniques will require silicon-like production and process technologies after 2005.
System-level Design Capability to Integrated Chips, Passives, and Substrates	Partitioning of system designs and manufacturing across numerous companies will make required optimization for performance, reliability, and cost of complex systems very difficult. Complex standards for information types and management of information quality along with a structure for moving this information will be required. Hardware only This is also an issue before 2007.Embedded passives may be integrated into the "bumps" as well as the substrates.
New Device Types (Organic, Nanostructures, Biological) that require New Packaging Technologies	Organic device packaging requirements not yet define (will chips grow their own packages) Biological interfaces will require new interface types
Bumpless area array technologies will be needed during this period. Face to face packages and other 3D packages are examples. High frequency, low power and low profile are driving forces	

SIP as addressed but the Roadmap does not deal with the critical issues of "systems" packaging. System designers and design tools need to contemplate these alternatives.

ENVIRONMENT, SAFETY, AND HEALTH

Five Difficult Challenges ≥ 50 nm/Through 2009	*Summary of Issues/Needs*
Chemicals, Materials, and Equipment Management	*New Chemical Assessment* Need for quality rapid assessment methodologies to ensure that new chemicals (or those carried over from previous technologies but that now face new restrictions) can be utilized in manufacturing, while protecting human health, safety, and the environment without delaying process implementation. *Chemical Data Collection* Need to document and make available environment, safety, and health characteristics of chemicals *Chemical Reduction* Need to develop processes that meet technology demands while reducing impact on human health, safety and the environment, both through replacement of hazardous materials with materials that are more benign, and by reducing chemical quantity requirements through more efficient and cost-effective process management *Environment Management* Need to develop effective management systems to address issues related to disposal of equipment, and hazardous and non-hazardous residue from the manufacturing process
Resource Conservation	*Natural Resource Conservation (Energy, Water)* Need to implement known (from supplier optimization studies, benchmarking surveys and best known methods) energy and water use reduction solutions Continue to design innovative energy and water efficient processing equipment *Chemicals and Materials Use* Need more efficient utilization of chemicals and materials *Resource Recycling* Increase resource reuse and recycling *Sustainable Growth* Continued expansion of semiconductor manufacturing with reduced impact on natural resources
Workplace Protection	*Equipment Safety* Continue to design ergonomically correct and safe equipment Minimize ergonomic stressors and health and safety risks during maintenance activities *Chemical Exposure Protection* Increase knowledge base on health and safety characteristics of chemicals, materials, and process byproducts in the manufacturing and maintenance processes and design out potential for chemical exposures and need for PPE
Climate Change Mitigation	*Reduce Energy Use of Process Equipment* Need to design energy efficient processing equipment *Reduce Energy Use of the Manufacturing Facility* Need to develop energy efficient facilities systems *Reduce High Global Warming Potential (GWP) Chemicals Emission* Need ongoing improvement in methods that reduce emissions from processes using GWP chemicals
Design for Environment, Safety, and Health (DFESH)	*Evaluate and Quantify ESH Impact* Need integrated way to evaluate and quantify ESH impact of process, chemicals, and process equipment, and to make ESH a design parameter in development of new equipment and processes

ESH Difficult Challenges (continued)

Five Difficult Challenges <50 nm/Beyond 2009	*Summary of Issues/Needs*
Chemicals, Materials and Equipment Management	*Chemical Use Information* Need to understand regulatory requirements that set chemical restrictions Need for comprehensive material life cycle analysis of semiconductor products
Resource Conservation	*Reduce Water, Energy, Chemicals and Materials Use* Need resource efficient processing and facility support equipment driving toward resource sustainability and greener fabs
Workplace Protection	*Equipment Safety* Need more emphasis on safety of fab automation systems/robotics and sub-system isolation (e.g., LOTO of components of cluster tools) for tool maintenance
Climate Change Mitigation	*Reduce Energy Use* The importance of reducing energy use to minimize/slow climate change will grow *Reduce High GWP Chemicals Emissions* The international pressures to reduce emissions of GWP chemicals will continue
Design for Environment, Safety, and Health (DFESH)	*Evaluate and Quantify ESH Impact* Need ESH integrated into the design and development of new equipment and processes

YIELD ENHANCEMENT

Difficult Challenge ≥45 nm/Through 2010	*Summary of Issues*
Design for Manufacture and Test (DFM and DFT) and Systematic Mechanisms Limited Yield (SMLY)—IC designs must be optimized for a given process capability and must be testable and diagnosable. Understanding SMLY is mandatory for achieving historic yield ramps in the future.	Design to process compatibility, design for manufacturability, design for test, design for diagnosability, systematic mechanisms limited yield (SMLY) model development.
High-Aspect-Ratio Inspection—High-speed, cost-effective tools are needed to rapidly detect defects at 1/2 X ground rule (GR) associated with high-aspect-ratio contacts, vias, and trenches and especially defects near or at the bottoms of these features	Poor transmission of energy into bottom of via and back out to detection system Large number of contacts and vias per wafer
Detection of Ever Shrinking Yield Critical Defects—High throughput, high capture rate detection tools are needed for ever shrinking critical defects of interest.	Line edge roughness, ACLV, subtle process variation. Where does process variation stop and defect start? Need to improve signal to noise to delineate defect from process variation.
Non-visual Defect Sourcing—Failure analysis tools and techniques are needed to enable localization of defects where no visual defect is detected.	Many defects that cause electrical faults are not detectable inline.
Difficult Challenge <45 nm/Beyond 2010	*Summary of Issues*
Develop Yield Models that include New Materials and Integration—Models must comprehend greater parametric sensitivities, complex integration issues, ultra-thin film integrity, impact of circuit design, greater transistor packing, etc.	Develop test structures for new technology nodes. Address complex integration issues. Model ultra-thin film integrity issues. Improve scaling methods for front-end processes including increased transistor packing density.
Defect Detection—Detection and simultaneous differentiation of multiple killer defect types is necessary at high capture rates and throughput.	Existing techniques trade-off throughput for sensitivity, but at predicted defect levels, both throughput and sensitivity are necessary for statistical validity. Ability to detect particles at critical size may not exist.
Correlation of Impurity Level to Yield—Methodology for employment and correlation of fluid/gas types to yield of a standard test structure/product.	Establish an employment methodology for each material type. Define a standard test for yield/parametric effect.
Yield Ramp and Mature Yields—With increasing process complexity and fewer yield learning cycles with each subsequent technology node, it would be impossible to achieve historic yield ramps and mature yield levels.	Long and complex process will make achieving historic yield ramps challenging. Also, new materials will introduce previously unseen yield problems. Need tools and methods to shorten yield learning cycles

METROLOGY

Five Difficult Challenges ≥45 nm/Through 2009	*Summary of Issues*
Factory level and company wide metrology integration for real time *in situ*, integrated, and inline metrology tools; continued development of robust sensors and process controllers; and data management that allows integration of add-on sensors.	Standards for process controllers and data management must be agreed upon. Conversion of massive quantities of raw data to information useful for enhancing the yield of a semiconductor manufacturing process. Better sensors must be developed for trench etch end point, ion species /energy /dosage (current), and wafer temperature during RTA.
Starting materials metrology and manufacturing metrology are impacted by the introduction of new substrates such as SOI. Impurity detection (especially particles) at levels of interest for starting materials and reduced edge exclusion for metrology tools. CD, film thickness, and defect detection are impacted by thin SOI optical properties and charging by electron and ion beams.	Existing capabilities will not meet Roadmap specifications. Very small particles must be detected and properly sized. Capability for SOI wafers needs enhancement. Challenges come from the extra optical reflection in SOI and the surface quality.
Control of high-aspect ratio technologies such as Damascene challenges all metrology methods. Key requirements are dimensional control, void detection in copper lines, and pore size distribution and detection of killer pores in patterned low-κ dielectrics.	New process control needs are not yet established. For example, 3D (CD and depth) measurements will be required for trench structures in new, low-κ dielectrics. Sidewall roughness impacts barrier integrity and the electrical properties of lines and vias.
Measurement of complex material stacks and interfacial properties including physical and electrical properties.	Reference materials and standard measurement methodology for new, high-κ gate and capacitor dielectrics with engineered thin films and interface layers as well as interconnect barrier and low-κ dielectric layers, and other process needs. Optical measurement of gate and capacitor dielectric averages over too large an area and needs to characterize interfacial layers. Carrier mobility characterization will be needed for stacks with strained silicon and SOI substrates. The same is true for measurement of barrier layers. High frequency dielectric constant measurements have shown a constant frequency response and are no longer a pressing need.
Measurement test structures and reference materials.	The area available for test structures is being reduced especially in the scribe lines. There is a concern that measurements on test structures located in scribe lines do not correlate with in die performance. Overlay and other test structures are sensitive to process variation, and test structure design must be improved to ensure correlation between measurements in the scribe line and on chip properties. Standards institutions need rapid access to state of the art development and manufacturing capability to fabricate relevant reference materials.
Five Difficult Challenges <45 nm/Beyond 2009	
Nondestructive, production worthy wafer and mask level microscopy for critical dimension measurement for 3D structures, overlay, defect detection, and analysis	Surface charging and contamination interfere with electron beam imaging. CD measurements must account for sidewall shape. CD for Damascene process may require measurement of trench structures. Process control such as focus exposure and etch bias will require greater precision and 3D capability.
New strategy for in-die metrology must reflect across chip and across wafer variation.	Correlation of test structure variations with in die properties is becoming more difficult as device shrinks.
Statistical limits of sub-45 nm process control	Controlling processes where the natural stochastic variation limits metrology will be difficult. Examples are low-dose implant, thin gate dielectrics, and edge roughness of very small structures.
Structural and elemental analysis at device dimensions.	Materials characterization and metrology methods are needed for control of interfacial layers, dopant positions, and atomic concentrations relative to device dimensions. One example is 3D dopant profiling.
Determination of manufacturing metrology when device and interconnect technology remain undefined.	The replacement devices for the transistor and structure and materials replacement for copper interconnect are being researched.

Note that statistical process control parameters are needed to replace inspection, reduce process variation, control defects, and reduce waste.

MODELING AND SIMULATION

Difficult Challenges ≥45 nm/Through 2010	*Summary of Issues*
High-Frequency Circuit Modeling for 5–40 GHz Applications	Efficient extraction and simulation of full-chip interconnect delay Accurate and yet efficient 3D interconnect models, especially for transmission lines and S-parameters High frequency circuit models including non-quasi-static, substrate noise and parasitic coupling Parameter extraction assisted by numerical electrical simulation instead of RF measurements
Front-end Process Modeling for Nanometer Structures	Diffusion/activation/damage models and parameters including low thermal budget processes in Si-based substrate, i.e., Si, SiGe:C (incl. strain), SOI, and ultra-thin body devices Characterization tools/methodologies for these ultra shallow geometries/junctions and low dopant levels Modeling hierarchy from atomistic to continuum for dopants and defects in bulk and at interfaces
Modeling of Equipment Influences on Features Generated in Deposition and Etching Processes	Fundamental physical data (e.g., rate constants, cross sections, surface chemistry); reaction mechanisms, and reduced models for complex chemistry Linked equipment/feature scale models CMP (full wafer and chip level, pattern dependent effects) MOCVD, PECVD, and ALD modeling Multi-generation equipment/wafer models
Lithography Simulation including NGL	Optical simulation of resolution enhancement techniques including mask optimization (OPC, PSM) Predictive resist models including line-edge roughness, etch resistance and mechanical stability Multi-generation lithography system models
Ultimate Nanoscale CMOS Simulation Capability	Methods and algorithms that contribute to prediction of CMOS limits Quantum based simulators Models and analysis to enable design and evaluation of devices and architectures beyond traditional planar CMOS Phenomenological gate stack models for ultra-thin dielectrics Models for device impact of statistical fluctuations in structures and dopant distributions
Thermal-mechanical-electrical Modeling for Interconnections and Packaging	Model thermal-mechanical and electronic properties of low κ, high κ and conductors and the impact of processing on these properties Model reliability of packages and interconnects, e.g. stress voiding, electromigration, piezoelectric effects; textures, fracture, adhesion
Difficult Challenges <45 nm/Beyond 2010	*Summary of Issues*
Modeling of Processing and Electrical Properties of New Materials	Computational materials science tools to understand materials properties, process options, and operating behavior for new materials applied in devices and interconnects, including especially for the following: gate stacks, predictive modeling of dielectric constant, bulk polarization charge, surface states, reliability, breakdown, and leakage currents including band structure, tunneling from process/materials and structure conditions
Compact Modeling including More Physical Models and Statistics	Computer-efficient inclusion of influences of statistics (incl. correlations) before process freeze, quantum/ballistic transport, etc., into compact modeling
Nano-scale Modeling	Process modeling tools for the development of novel nanostructure devices (nanowires, carbon nanotubes, quantum dots, molecular electronics) Device modeling tools for analysis of nanoscale device operation (quantum transport, resonant tunneling, spintronics, contact effects)
Optoelectronics Modeling	Coupling between electrical and optical systems, optical interconnect models, semiconductor laser modeling Physical design tools for integrated electrical/optical systems

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 includes 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 undergo 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 2002. Additionally, an ORTC Glossary is provided as an appendix.

OVERVIEW OF 2003 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, 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,” Furthermore, a second company must begin 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.

Due to confusion in public press announcements, especially regarding Logic technology “nodes,” additional clarification was provided this year by the *ITRS* executive International Roadmap Committee (IRC), adding a technology node designator, “hpXX” to the table header. This designator represents the most aggressive interconnect half-pitch in the industry, which at present is the DRAM cell metal half-pitch. At some point in the future it may be represented by a half-pitch of a different product. Please see the Glossary section for additional details on “Technology Node” and “Production” timing definitions.

Per previously established IRC guidelines, the *2003 ITRS* retains the definition of a technology node as the achievement of significant advancement in the process technology. To be explicit, a technology node is defined as the achievement of an approximate 0.7× reduction per node (0.5× per two nodes). Refer to Figure 5. The period of time in which a new technology node is reached is called a “technology-node cycle.” Refer to 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 *2003 ITRS* timeframe from 2009 (2012, 2015, 2018) and still include the previous ITRS 2001 columns (2010, 2013, 2016) as a reference.

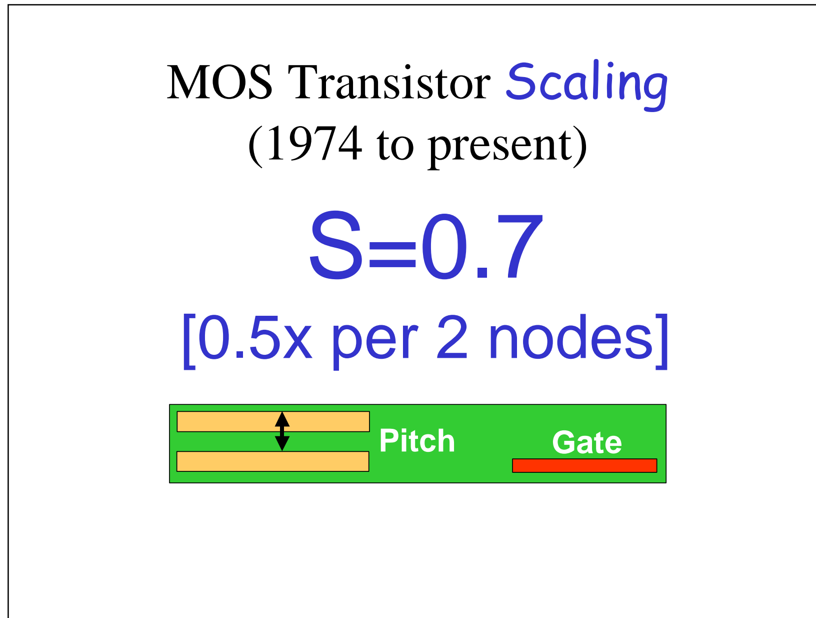


Figure 5 MOS Transistor Scaling—1974 to present

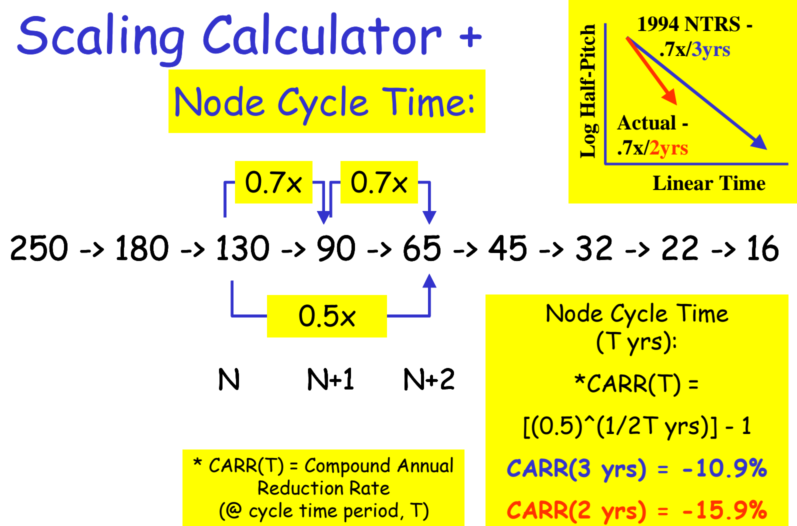


Figure 6 Scaling Calculator

ROADMAP TIMELINE

The 2003 edition of the Roadmap maintains a 15-year projection, from 2003 as a reference year and through 2018. The timing trend of future technology nodes (three years between nodes) has remained unchanged from the 2001 edition. Therefore, by international consensus, the 90 nm DRAM half-pitch node could begin production ramp between 1Q04 to 4Q04, depending on the completion of customer product qualification, which was made an explicit requirement of the "Production" definition for the *2003 ITRS*.

In the *2001 ITRS*, the 130 nm node was pulled in an additional year (from 2002 in the *1999 ITRS* to 2001), anticipating a continuation of an observed historical two-year technology-node cycle calculated from 350 nm/1995, 250 nm node in 1997, 180 nm node in 1999). Data provided by DRAM manufacturers in 2003, which was based upon the rigorous customer-product-qualified production ramp, indicated that the actual production ramp timing was as follows: 350 nm/1995, 250 nm/1998, 180 nm/2000 and 130 nm/2002. This new data indicates a two-year node cycle timing, but delayed one year from the original *2001 ITRS* timing. Data gathered on actual DRAM product ramped in 2003 will confirm if the interim node step is 100 nm, per the 2001 and *2003 ITRS*, or 110 nm, which would indicate a two-year cycle step between 130 nm/2002 and 90 nm/2004. Although there is the possibility of a continuation of this new delayed two-year-node cycle trend, the present consensus projects a three-year cycle for DRAM interconnect half-pitch nodes throughout the 2003–2018 Roadmap period, as illustrated in Figure 7.

As mentioned above, the DRAM interconnect 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, future data analysis might indicate an aggressive trend for the lagging MPU, ASIC, and Flash metal and/or polysilicon interconnect half-pitches to pass the DRAM half-pitch after 2004, and become the ITRS header node standard. See Figure 7.

ROUNDED TREND NUMBERS

As a result of the new DRAM half-pitch data inputs, and using the 180 nm node as the calculation standard for trends, the *2003 ITRS* now 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 in Table C.

Table C Rounded versus Actual Trend Numbers

YEAR OF PRODUCTION	*1995*	*1998*	*2000*	*2002*	*2003*	*2004*	*2006*	*2007*	*2009*	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
						hp90		*hp65*		*hp45*		*hp32*		*hp22*	
Calculated Trend Numbers (nm)	**360**	**255**	**180**	**127.3**	**101**	**90**	**71.4**	**63.6**	**50.5**	**45**	**35.7**	**31.8**	**25.3**	**22.5**	**17.9**
ITRS Rounded Node Numbers (nm)	**350**	**250**	**180**	**130**	**100**	**90**	**70**	**65**	**50**	**45**	**35**	**32**	**25**	**22**	**18**

Note the new rounding corrections become more critical as the industry moves into the double-digit technology nodes of the new nanotechnology (sub-100 nm) era. Please note also that some regions, for their own past publication consistency, will retain their right to continue to track the previous technology nodes beginning with 100 nm/2003. Starting from 100 nm in 2003 will result in node milestones that are targeted one year earlier than the present 2003 roadmap hpXX convention (70 nm/2006; 50 nm/2009; 35 nm/2012; 25 nm/2015). By consensus of the IRC both node number sets are available for long-term calculations, since the original *2001 ITRS* long-term columns were retained (2010/hp45/45 nm; 2013/hp32/32 nm; 2016/hp22/22 nm), and new columns (2012/35 nm; 2015/25 nm; 2018/18 nm) were added.

UPDATES TO THE ORTC

A new addition to the *2003 ITRS* ORTC technology target line items is the Logic Metal 1 (M1) half-pitch. This was added to the ORTC Table 1a and 1b in addition to the unchanged polysilicon half-pitch in order to be consistent with observed industry status and also to be consistent with the Interconnect TWG logic pitch targets, which track contacted Metal 1 rather than the *2001 ITRS* un-contacted polysilicon half-pitch.

The *printed* MPU gate length received a major correction to more an aggressive starting point in the *2001 ITRS*. 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 remain unchanged for the *2003 ITRS*, and are forecast to continue scaling by about 70% per two-year cycle through the 32 nm 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. Refer to Figure 8.

The ORTC metrics are often used by semiconductor companies as a set of targets that need to be achieved ahead of schedule to secure 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, the gathering and analysis of actual data, combined with the ITRS 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, the actual data and conference papers, along with company survey data and public announcements will be re-evaluated during the year *2004 ITRS* update process, and the possibility of a continued two-year node cycle. In addition, logic and Flash product half-pitch acceleration will be monitored for future header leadership candidates.

As mentioned above, 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 2003 through 2009, called the "Near-term Years," and at three-year (node) intervals thereafter, called the "Long-term years" (2012, 2015, 2018), while retaining the previous *2001 ITRS* long-term columns for ease of comparison and to retain the tracking of the three-year cycle nodes.

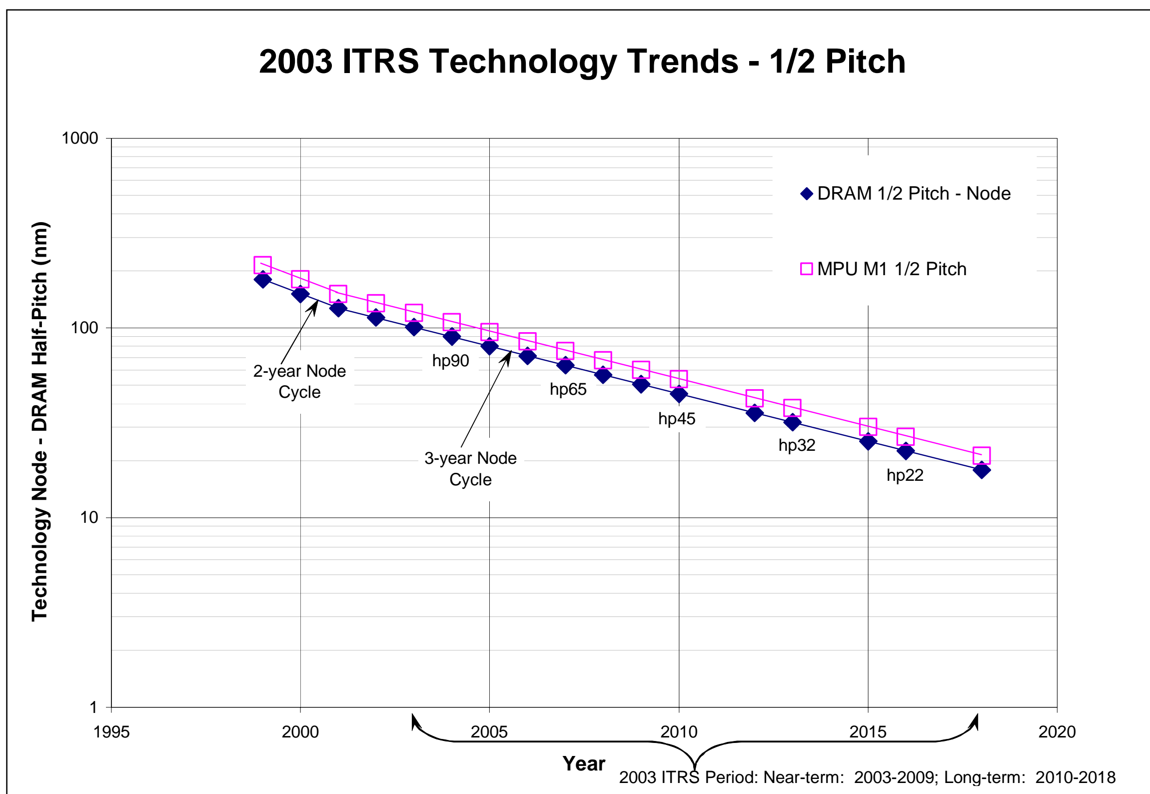


Figure 7 2003 ITRS—Half Pitch Trends

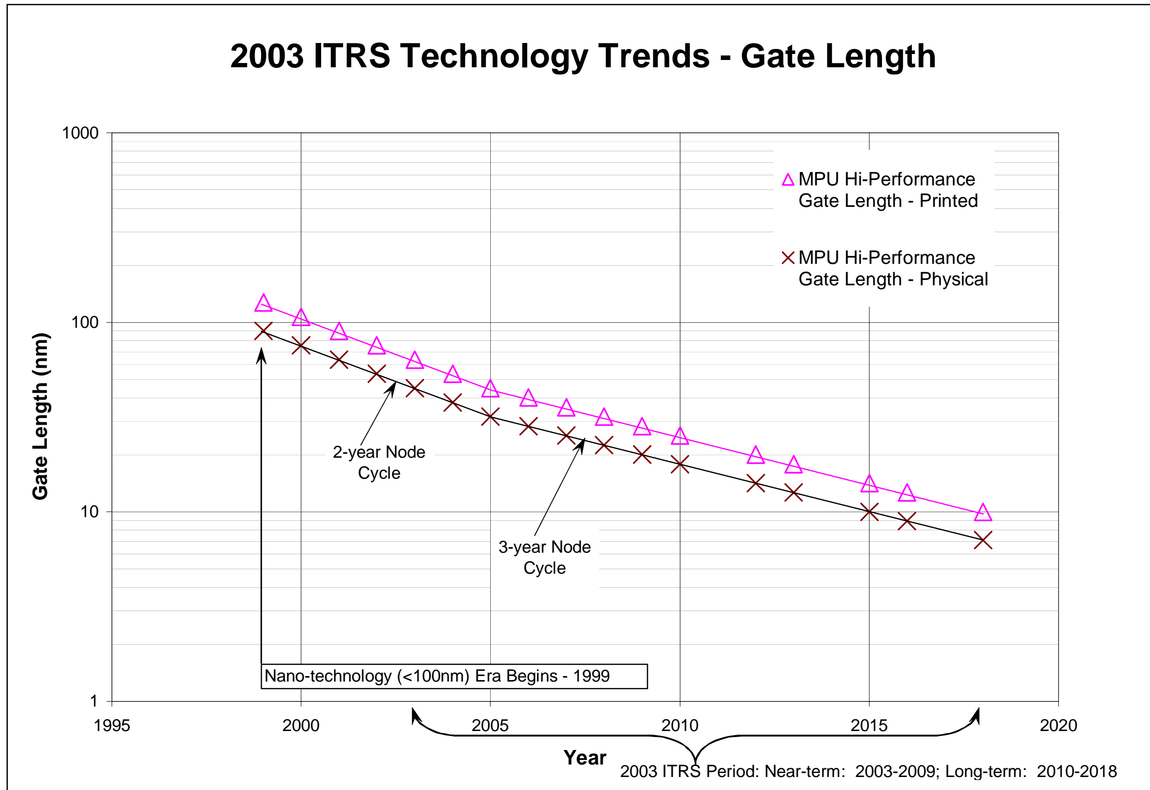


Figure 8 2003 ITRS—Gate Length Trends

PRODUCT GENERATIONS AND CHIP-SIZE MODEL

This section discusses "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 2003 ITRS 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 2003 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 2003 *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 the ORTC table notations. Table 1a and 1b summarize the near and long-term 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 included.

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	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
ASIC/Low Operating Power Printed Gate Length (nm) ††	**90**	**75**	**65**	**53**	**45**	**40**	**35**
ASIC/Low Operating Power Physical Gate Length (nm)	**65**	**53**	**45**	**37**	**32**	**28**	**25**

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

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
ASIC/Low Operating Power Printed Gate Length (nm) ††	**32**	**25**	**22**	**18**	**16**	**13**
ASIC/Low Operating Power Physical Gate Length (nm)	**22**	**18**	**16**	**13**	**11**	**9**

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.

In addition, during the 2003 ITRS development, an attempt has been made to reconcile the many published press releases by Logic manufacturers referencing “90 nm” technology node manufacturing in 2003. Since the metal 1 (M1) half-pitch of actual devices was cited at 110–120 nm, confusion arose regarding the relationship to the ITRS DRAM half-pitch-based header targets. After conversation with leading-edge manufacturers, it was determined that the public citations were in reference to an “indexed” technology node roadmap that represented the average of the half-pitch (for density) and the printed gate length (for speed performance).

The IRC has decided that the best way to minimize confusion between the ITRS and individual company public announcements is to identify the ITRS table header node with the industry's most aggressive half-pitch targets, and to label these targets as hpXX (i.e., hp90, hp65, hp45, etc.). Currently the industry's most aggressive half pitch is the DRAM cell metal half-pitch.

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

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

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Cell area factor [a]	**8**	**8**	**7.5**	**7**	**7**	**6**	**6**
Cell area [Ca = af^2] (mm^2)	**0.082**	**0.065**	**0.048**	**0.036**	**0.028**	**0.019**	**0.015**
Cell array area at production (% of chip size) §	**63.00%**	**63.00%**	**63.00%**	**63.00%**	**63.00%**	**63.00%**	**63.00%**
Generation at production §	**1G**	**1G**	**1G**	**2G**	**2G**	**4G**	**4G**
Functions per chip (Gbits)	**1.07**	**1.07**	**1.07**	**2.15**	**2.15**	**4.29**	**4.29**
Chip size at production (mm^2)§	**139**	**110**	**82**	**122**	**97**	**131**	**104**
Gbits/cm^2 at production §	**0.77**	**0.97**	**1.31**	**1.76**	**2.22**	**3.27**	**4.12**

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

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
Cell area factor [a]	**6**	**6**	**6**	**6**	**5**	**5**
Cell area [Ca = af^2] (mm^2)	**0.012**	**0.0077**	**0.0061**	**0.0038**	**0.0025**	**0.0016**
Cell array area at production (% of chip size) §	**63.00%**	**63.0%**	**63.0%**	**63.0%**	**63.0%**	**63.0%**
Generation at production §	**4G**	**8G**	**8G**	**16g**	**32G**	**32G**
Functions per chip (Gbits)	**4.29**	**8.59**	**8.59**	**17.18**	**34.36**	**34.36**
Chip size at production (mm^2)§	**83**	**104**	**83**	**104**	**138**	**87**
Gbits/cm^2 at production §	**5.19**	**8.23**	**10.37**	**16.46**	**24.89**	**39.51**

Notes for Tables 1c and 1d:

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

1999-2004/8×: 2005/7.5×; 2006-2007/7×; 2008-2015/6×; 2016–2018/5×. The delay of the "6" DRAM cell design improvement factor [a] by five years, from 2003 to 2008, requires the slowing of the addition of "Moore's Law" bits/chip from 2× every 1.5–2 years to 2× every 2.5–3 years in the 2003 ITRS DRAM Chip Size Model, which remains on a three-year DRAM half-pitch node cycle after 2004.

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 16 Gbit generation, the introduction rate is 4×/six years (2×/three years); and

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

InTER-generation chip size growth rate model target for Production-phase DRAMs is now "flat" at less than 140 mm^2, similar to the MPU model. This new flat-chip-size model target requires the bits/chip "Moore's Law" model for DRAMs to increase the time for doubling bits per chip to an average of 2×/2.5 years by alternating between 2×/2 years and 2×/3 years (see ORTC Table 1c,d). In addition, the Cell Array Efficiency (Array % of total chip area) was increased to 63% , which also assists in the achievement of the target flat-chip-size model for the Production-phase product chip size, even under the slower design improvement factor contribution (see note above).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	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
DRAM ½ Pitch (nm)	**100**	**90**	**80**	**70**	**65**	**57**	**50**
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Cell area factor [a]	**8**	**8**	**7.5**	**7**	**7**	**6**	**6**
Cell area [$Ca = af^2$] (mm^2)	**0.082**	**0.065**	**0.048**	**0.036**	**0.028**	**0.019**	**0.015**
Cell array area at introduction (% of chip size) §	**72.23%**	**72.61%**	**72.95%**	**73.25%**	**73.52%**	**73.76%**	**73.97%**
Generation at introduction §	**4G**	**4G**	**8G**	**8G**	**16G**	**16G**	**16G**
Functions per chip (Gbits)	**4.29**	**4.29**	**8.59**	**8.59**	**17.18**	**17.18**	**17.18**
Chip size at introduction (mm^2) §	**485**	**383**	**568**	**419**	**662**	**449**	**356**
Gbits/cm^2 at introduction §	**0.88**	**1.12**	**1.51**	**2.05**	**2.59**	**3.82**	**4.83**

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

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9.0**	**7.0**
Cell area factor [a]	**6**	**6**	**6**	**6**	**5**	**5**
Cell area [Ca = af2] (mm^2)	**0.012**	**0.0077**	**0.0061**	**0.0038**	**0.0025**	**0.0016**
Cell array area at introduction (% of chip size) §	**74.16%**	**74.47%**	**74.61%**	**74.83%**	**74.93%**	**75.09%**
Generation at introduction §	**32G**	**32G**	**64G**	**64G**	**128G**	**128G**
Functions per chip (Gbits)	**34.36**	**34.36**	**68.72**	**68.7**	**137.4**	**137.4**
Chip size at introduction (mm^2) §	**563**	**353**	**560**	**351**	**464**	**292**
Gbits/cm2 at introduction §	**5.2**	**8.2**	**10.4**	**16.5**	**24.9**	**39.5**

Notes for Tables 1e and 1f:

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

1999-2004/8×: 2005/7.5×; 2006-2007/7×; 2008-2015/6×; 2016–2018/5×. The delay of the "6" DRAM Cell design improvement Factor [a] by five years, from 2003 to 2008, requires the slowing of the addition of "Moore's Law" bits/chip from 2× every 1.5–2 years to 2× every 2.5–3 years in the 2003 ITRS DRAM Chip Size Model, which remains on a 3-year DRAM half-pitch node cycle after 2004.

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 16 Gbit generation/2007, the introduction rate is 4×/six years (2×/three years); and

2. at the Production phase, after the 1 Gbit generation/2003, the introduction rate is 4×/five years (2×/two–three years).

The original 2001 ITRS InTER-generation chip size growth rate was targeted to fit one chip per 572 mm^2 field at Introduction and two chips per 572 mm^2 field at Production. Due to the delay of the cell area factor reductions, Introduction chip sizes increased, but the new 704 mm^2 maximum affordable Litho field allows the Introduction chip to double bits per chip every two years through the 16 Gbit generation (660 mm^2/2007). Slowing the "Moore's Law" bits per chip of the Introduction-phase DRAM model to an average of 2× per 2.5 years enables the Introduction DRAMs to remain under the original 572 mm^2 affordable target after 2007. The InTRA-generation chip size shrink model remains at 0.5× every technology node in-between cell factor reductions, and eventually (ranging from five to six years), the Introduction-phase DRAMs shrink below the 140 mm^2 Production-phase chip size target.

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	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
DRAM ½ Pitch (nm)	**100**	**90**	**80**	**70**	**65**	**57**	**50**
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
SRAM Cell (6-transistor) Area factor ++	**120.3**	**117.8**	**115.6**	**113.7**	**111.9**	**110.4**	**109.0**
Logic Gate (4-transistor) Area factor ++	**320**	**320**	**320**	**320**	**320**	**320**	**320**
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 ++	**2.0**	**1.5**	**1.2**	**0.93**	**0.73**	**0.57**	**0.45**
Logic Gate (4-transistor) Area w/overhead ++	**6.5**	**5.2**	**4.1**	**3.3**	**2.6**	**2.1**	**1.6**
Transistor density SRAM (Mtransistors/cm^2)	**305**	**393**	**504**	**646**	**827**	**1,057**	**1,348**
Transistor density logic (Mtransistors/cm^2)	**61**	**77**	**97**	**122**	**154**	**194**	**245**
*Generation at introduction **	**--**	**p07c**	**--**	**--**	**p10c**	**--**	**--**
Functions per chip at introduction (million transistors [Mtransistors])	**180**	**226**	**285**	**360**	**453**	**571**	**719**
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) ‡	**110**	**138**	**174**	**219**	**276**	**348**	**438**
*Generation at production **	**--**	**p04c**	**--**	**--**	**p07c**	**--**	**--**
Functions per chip at production (million transistors [Mtransistors])	**153**	**193**	**243**	**307**	**386**	**487**	**614**
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) ‡	**110**	**138**	**174**	**219**	**276**	**348**	**438**

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

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
DRAM ½ Pitch (nm)	**45**	**35**	**32**	**25**	**22**	**18**
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
SRAM Cell (6-transistor) Area factor ++	**107.8**	**105.7**	**104.8**	**103.4**	**102.8**	**101.7**
Logic Gate (4-transistor) Area factor ++	**320**	**320**	**320**	**320**	**320**	**320**
SRAM Cell (6-transistor) Area efficiency ++	**0.625**	**0.625**	**0.625**	**0.625**	**0.625**	**0.625**
Logic Gate (4-transistor) Area efficiency ++	**0.500**	**0.500**	**0.500**	**0.500**	**0.500**	**0.500**
SRAM Cell (6-transistor) Area w/overhead ++	**0.22**	**0.13**	**0.11**	**0.066**	**0.052**	**0.032**
Logic Gate (4-transistor) Area w/overhead ++	**1.30**	**0.82**	**0.65**	**0.41**	**0.32**	**0.20**
Transistor density SRAM (Mtransistors/cm^2)	**1,718**	**2,781**	**3,532**	**5,687**	**7,208**	**11,558**
Transistor density logic (Mtransistors/cm^2)	**309**	**490**	**617**	**980**	**1,235**	**1,960**
*Generation at introduction **	**p13c**	**--**	**p16c**	**--**	**p19c**	**--**
Functions per chip at introduction (million transistors [Mtransistors])	**1,546**	**2,454**	**3,092**	**4,908**	**6,184**	**9,816**
Chip size at introduction (mm^2) ‡	**280**	**280**	**280**	**280**	**280**	**280**
Cost performance MPU (Mtransistors/cm^2 at introduction) (including on-chip SRAM) ‡	**552**	**876**	**1,104**	**1,753**	**2,209**	**3,506**
*Generation at production **	**p10c**	**--**	**p13c**	**--**	**p16c**	**--**
Functions per chip at production (million transistors [Mtransistors])	**773**	**1,227**	**1,546**	**2,454**	**3,092**	**4,908**
Chip size at production (mm^2) §§	**140**	**140**	**140**	**140**	**140**	**140**
Cost performance MPU (Mtransistors/cm^2 at production, including on-chip SRAM) ‡	**552**	**876**	**1,104**	**1,753**	**2,209**	**3,506**

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 2018 (280 mm^2/cost-performance at introduction; 140 mm^2/cost-performance at production; 310 mm^2/high-performance at production). The MPU flat chip-size model is made possible by doubling the on-chip functionality every technology node cycle. The InTRA-generation chip size shrink model was 0.5× every two-year technology node through 2001, and is now 0.5× every three-year technology node cycle after 2003.

Refer to the Glossary for definitions

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

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Logic (Low-volume Microprocessor) High-performance ‡							
*Generation at production ***	**p03h**	**--**	**p05h**	**--**	**p07h**	**--**	**p09h**
Functions per chip (million transistors)	**439**	**553**	**697**	**878**	**1,106**	**1,393**	**1,756**
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) ‡	**142**	**178**	**225**	**283**	**357**	**449**	**566**
ASIC							
ASIC usable Mtransistors/cm^2 (auto layout)	**142**	**178**	**225**	**283**	**357**	**449**	**566**
ASIC max chip size at production (mm^2) (maximum lithographic field size)	**572**	**572**	**572**	**572**	**572**	**572**	**572**
ASIC maximum functions per chip at production (Mtransistors/chip) (fit in maximum lithographic field size)	**810**	**1,020**	**1,286**	**1,620**	**2,041**	**2,571**	**3,239**

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

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
Logic (Low-volume Microprocessor) High-performance ‡						
*Generation at production ***	**--**	**--**	**p13h**	**p15h**	**--**	**--**
Functions per chip (million transistors)	**2,212**	**3,511**	**4,424**	**7,022**	**8,848**	**14,045**
Chip size at production (mm^2) §§	**310**	**310**	**310**	**310**	**310**	**310**
High-performance MPU Mtransistors/cm^2 at production (including on-chip SRAM) ‡	**714**	**1,133**	**1,427**	**2,265**	**2,854**	**4,531**
ASIC						
ASIC usable Mtransistors/cm^2 (auto layout)	**714**	**1,133**	**1,427**	**2,265**	**2,854**	**4,531**
ASIC maximum chip size at production (mm^2) (maximum lithographic field size)	**572**	**572**	**572**	**572**	**572**	**572**
ASIC maximum functions per chip at ramp (Mtransistors/chip) (fit in maximum lithographic field size)	**4,081**	**6,479**	**8,163**	**12,958**	**16,326**	**25,915**

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, the p03h in2003, and so forth.*

‡ 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 2018 (280 mm^2/cost-performance at introduction; 140 mm^2/cost-performance at production; 310 mm^2/high-performance at production). The MPU flat chip-size model is made possible by doubling the on-chip functionality every technology node cycle. The InTRA-generation chip size shrink model was 0.5× every two-year technology node through 2001, and is now 0.5× every three-year technology node cycle after 2003.

Refer to the Glossary for definitions

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 as flat as possible. 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.

In order for affordable DRAM products to achieve virtually flat intra-generation chip-sizes, they must also maintain a cell area array efficiency ratio of 63% of total chip area. Therefore, DRAM products require reduction of cell area design factors (cell area in units of minimum-feature-size-squared). The PIDS and FEP ITWGs have provided member survey data for the array efficiency targets, the cell area factors, and bits per chip. In addition, detailed challenges and needs for solutions to meet the aggressive cell area goals are documented 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. (also refer to the Glossary for additional details). Notably, the reduction rate of DRAM cell area factors for the *2003 ITRS* models has been slowed significantly (the 6 factor moved from 2003 to 2008, and the 4 factor has been increased to 5 and moved from 2011 to 2016). In order to maintain the goal of flat chip sizes, the *2003 ITRS* DRAM chip size model now includes more aggressive array efficiency targets, and the rate of increase of "Moore's Law" bits per chip targets has been slowed from 2× every three years to 2× every three years.

In the *2001 ITRS* 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 *2003 ITRS* MPU chip size model is unchanged from the *2001 ITRS*, and continues to reflect the additional competitive requirements for affordability and power management by targeting flat chip size trends for both high-performance MPUs (310 mm^2) and cost-performance MPUs (140 mm^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 inter-generation MPU chip size model, which is indexed to the ITRS technology node, 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× 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 (25×32 = 800 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 of a requirement for the economically affordable lithography field was reduced to 572 mm^2 (22×26) by the 90 nm node. After addition review, the Lithography TWG increased the "Affordable" field size to 704 mm^2 (22×32) for the *2003 ITRS*. 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 *2003 ITRS* chip-size model for DRAMs, the introduction-level chip size is targeted to be smaller than the new affordable 704 mm^2 lithography field size, fitting at least one introduction-level chip size within the field. The new production-level DRAM chip size model (less than 140 mm^2 flat target) fits four 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, as mentioned above, the slowing of DRAM design improvements causes a requirement to add fewer on-chip bits to stay under the affordable lithography field limit. This accomplished in the present DRAM model by slowing the Moore's Law bits/chip rate to 2×/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 has created financial challenges and limited capital investment. The maximum substrate diameter in Tables 2a and b (and in additional detail in the FEP chapter) is consistent with the ramp of 300 mm capacity beginning 2001. Also, the first manufacturing capability for the next 1.5× wafer size conversion to 450 mm diameter is not anticipated to be required until 2011–2012 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, or an equivalent processing platform, as a productivity improvement.

The effects 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 Semiconductor Equipment and Materials International (SEMI) and International SEMATECH.

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

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Lithography Field Size							
Lithography Field Size—area (mm^2)	**704**	**704**	**704**	**704**	**704**	**704**	**704**
Lithographic field size—length (mm)	**32**	**32**	**32**	**32**	**32**	**32**	**32**
Lithographic field size—width (mm)	**22**	**22**	**22**	**22**	**22**	**22**	**22**
Maximum Substrate Diameter (mm)—High-volume Production (>20K wafer starts per month)							
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*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
Lithography Field Size						
Lithography Field Size—area (mm^2)	**704**	**704**	**704**	**704**	**704**	**704**
Lithographic field size—length (mm)	**32**	**32**	**32**	**32**	**32**	**32**
Lithographic field size—width (mm)	**22**	**22**	**22**	**22**	**22**	**22**
Maximum Substrate Diameter (mm)—High-volume Production (>20K wafer starts per month)						
Bulk or epitaxial or SOI wafer	**300**	**450**	**450**	**450**	**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

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Number of Chip I/Os (Number of Total Chip Pads)—Maximum							
Total pads—MPU	**3,072**	**3,072**	**3,072**	**3,072**	**3,072**	**3,328**	**3,584**
Signal I/O—MPU (1/3 of total pads)	**1,024**	**1,024**	**1,024**	**1,024**	**1,024**	**1,109**	**1,195**
Power and ground pads—MPU (2/3 of total pads)	**2,048**	**2,048**	**2,048**	**2,048**	**2,048**	**2,219**	**2,389**
Total pads—ASIC high-performance	**3,400**	**3,600**	**4,000**	**4,200**	**4,400**	**4,600**	**4,800**
Signal I/O pads—ASIC high-performance	**1,700**	**1,800**	**2,000**	**2,100**	**2,200**	**2,300**	**2,400**
Power and ground pads—ASIC high-performance (½ of total pads)	**1,700**	**1,800**	**2,000**	**2,100**	**2,200**	**2,300**	**2,400**
Number of Total Package Pins—Maximum [1]							
Microprocessor/controller, cost-performance	**500–1,452**	**500–1,600**	**550–1,760**	**550–1,936**	**600–2,140**	**660–2,354**	**720–2,568**
Microprocessor/controller, high-performance	**1,452**	**1,600**	**1,760**	**1,936**	**2,140**	**2,354**	**2,568**
ASIC (high-performance)	**2,400**	**3,000**	**3,400**	**3,800**	**4,000**	**4,400**	**4,600**

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*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
Number of Chip I/Os (Number of Total Chip Pads)—Maximum						
Total pads—MPU	**3,840**	**4,096**	**4,224**	**4,352**	**4,416**	**4,544**
Signal I/O—MPU (1/3 of total pads)	**1,280**	**1,365**	**1,408**	**1,451**	**1,472**	**1,515**
Power and ground pads—MPU (2/3 of total pads)	**2,560**	**2,731**	**2,816**	**2,901**	**2,944**	**3,029**
Total pads—ASIC high-performance	**4,800**	**5,200**	**5,400**	**5,800**	**6,000**	**6,400**
Signal I/O pads—ASIC high-performance	**2,400**	**2,600**	**2,700**	**2,900**	**3,000**	**3,200**
Power and ground pads—ASIC high-performance (½ of total pads)	**2,400**	**2,600**	**2,700**	**2,900**	**3,000**	**3,200**
Number of Total Package Pins—Maximum [1]						
Microprocessor/controller, cost-performance	**780–2,782**	**936–3,338**	**1014–3,616**	**1217–4,340**	**1318–4,702**	**1521–5,426**
Microprocessor/controller, high-performance	**2,782**	**3,338**	**3,616**	**4,340**	**4,702**	**5,426**
ASIC (high-performance)	**4,009**	**4,810**	**5,335**	**6,402**	**7,042**	**8,450**

Package pin count (Tables 3a and 3b) and cost-per-pin (Tables 4a and 4b), 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 cost/pin × 1.10 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 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 will continue to demand the development of novel process, design, and packaging techniques.

These considerations are reflected in Tables 4c and 4d, which includes line items contributed by the Design ITWG and the Assembly and Packaging ITWG 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 (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—Near-term Years

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Chip Pad Pitch (micron)							
Pad pitch—ball bond	**40**	**35**	**30**	**25**	**25**	**20**	**20**
Pad pitch—wedge bond	**30**	**25**	**20**	**20**	**20**	**20**	**20**
Pad pitch—wedge bond	**30**	**25**	**20**	**20**	**20**	**20**	**20**
Pad Pitch—area array flip-chip (cost-performance, high-performance)	**150**	**150**	**130**	**130**	**120**	**110**	**100**
Pad Pitch—peripheral flip-chip (handheld, low-cost, harsh)	**60**	**60**	**40**	**40**	**30**	**30**	**20**
Cost-Per-Pin							
Package cost (cents/pin) (cost-performance)— minimum-maximum	**.75–1.30**	**.71–1.24**	**.67–1.17**	**.64–1.11**	**.61–1.05**	**.58–1.00**	**.55–.96**
Package cost (cents/pin) (Memory)— minimum–maximum	**0.30–0.56**	**0.29–.53**	**.27–.50**	**,26–.48**	**,25–.45**	**.23–.43**	**.22–.41**

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

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
Chip Pad Pitch (micron)						
Pad pitch—ball bond	**20**	**20**	**20**	**20**	**20**	**20**
Pad Pitch—wedge bond	**20**	**20**	**20**	**20**	**20**	**20**
Pad Pitch—area array flip-chip (cost-performance, high-performance)	**100**	**90**	**90**	**80**	**80**	**70**
Pad Pitch—peripheral flip-chip (handheld, low-cost, harsh)	**20**	**20**	**20**	**15**	**15**	**15**
Cost-Per-Pin						
Package cost (cents/pin) (cost-performance)— minimum–maximum	**0.52–0.94**	**0.5–.86**	**0.5–.77**	**0.5–0.69**	**0.5–0.65**	**0.5–0.59**
Package cost (cents/pin) (Memory)— minimum–maximum	**.22–.41**	**0.22–0.36**	**0.22–0.35**	**0.22–0.31**	**0.22–0.29**	**0.22–0.27**

Table 4c Performance and Package Chips: Frequency On-chip Wiring Levels—Near-term Years

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Chip Frequency (MHz)							
On-chip local clock	**2,976**	**4,171**	**5,204**	**6,783**	**9,285**	**10,972**	**12,369**
Chip-to-board (off-chip) speed (high-performance, for peripheral buses)[1]	**2,000**	**2,500**	**3,125**	**3,906**	**4,883**	**6,103**	**7,629**
Maximum number wiring levels—maximum	**13**	**14**	**15**	**15**	**15**	**16**	**16**
Maximum number wiring levels—minimum	**9**	**10**	**11**	**11**	**11**	**12**	**12**

Table 4d Performance and Package Chips: Frequency, On-chip Wiring Levels—Long-term Years

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
Chip Frequency (MHz)						
On-chip local clock	**15,079**	**20,065**	**22,980**	**33,403**	**39,683**	**53,207**
Chip-to-board (off-chip) speed (high-performance, for peripheral buses)[1]	**9,536**	**14,901**	**18,626**	**29,103**	**36,379**	**56,843**
Maximum number wiring levels—maximum	**16**	**16**	**16**	**17**	**18**	**18**
Maximum number wiring levels—minimum	**12**	**12**	**12**	**13**	**14**	**14**

Note for Tables 4c and 4d:

[1] The off-chip frequency is expected to increase for a small number of high-speed pins that 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.

[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	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
DRAM Random Defect D_0 at production chip size and 89.5% yield (faults/m^2) §	**2,216**	**2,791**	**3,751**	**2,532**	**3,190**	**2,345**	**2,954**
MPU Random Defect D_0 at production chip size and 83% yield (faults/ m^2) §§	**1,395**	**1,395**	**1,395**	**1,395**	**1,395**	**1,395**	**1,395**
# Mask Levels—MPU	**29**	**31**	**33**	**33**	**33**	**35**	**35**
# Mask Levels—DRAM	**24**	**24**	**24**	**24**	**24**	**24**	**24**

Table 5b Electrical Defects—Long-term Years

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
DRAM Random Defect D_0 at production chip size and 83% yield (faults/ m^2) §	**3,722**	**2,954**	**3,722**	**2,954**	**2,233**	**3,545**
MPU Random Defect D_0 at production chip size and 89.5% yield (faults/ m^2) §§	**1,395**	**1,395**	**1,395**	**1,395**	**1,395**	**1,395**
# Mask Levels—MPU	**35**	**35**	**35**	**37**	**39**	**39**
# Mask Levels—DRAM	**26**	**26**	**26**	**26**	**26**	**26**

Notes for Tables 5a and 5b:

D_0 — defect density

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

1999–2004/8×: 2005/7.5×; 2006–2007/7×; 2008–2015/6×; 2016–2018/5×. The delay of the "6" DRAM cell design improvement factor [a] by five years, from 2003 to 2008, requires the slowing of the addition of "Moore's Law" bits/chip from 2× every 1.5–2 years to 2×.

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 16 Gbit generation, the introduction rate is 4×/six years (2×/three years); and

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

The original 2001 ITRS InTER-generation chip size growth rate was targeted to fit one chip per 572 mm² field at Introduction and two chips per 572 mm² field at Production. Due to the delay of the cell area factor reductions, Introduction chip sizes increased, but the new 704 mm² maximum affordable lithography field allows the Introduction chip to double bits per chip every two years through the 16 Gbit generation (660 mm²/2007). Slowing the "Moore's Law" bits per chip of the Introduction-phase DRAM model to an average of 2× per 2.5 years enables the Introduction DRAMs to remain under the original 572 mm² affordable target after 2007. The InTRA-generation chip size shrink model remains at 0.5× every technology node in-between cell factor reductions, and eventually (ranging from five to six years), the Introduction-phase DRAMs shrink below the 140 mm² Production-phase chip size target.

§§ MPU Chip Size Model—Both the cost-performance and high-performance MPUs InTER-generation chip size growth rates are targeted to be flat through 2018 (280 mm²/cost-performance at introduction; 140 mm²/cost-performance at production; 310 mm²/high-performance at production). The MPU flat chip-size model is made possible by doubling the on-chip functionality every technology node cycle. The InTRA-generation chip size shrink model was 0.5× every two-year technology node through 2001, and is now 0.5× every three-year technology node cycle after 2003.

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

Power Supply and Power Dissipation

The reduction of power supply voltage 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 not expected to be achieved by high-performance processors until beyond 2018 (versus 2013 in the *2001 ITRS*). The lowest V_{dd} target is now 0.5V in 2016 for the low operating power applications, a lower target than the 0.6V goal in the *2001 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 (now designated as the "Harsh" application category by the Assembly and Packaging TWG). 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	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Power Supply Voltage (V)							
V_{dd} *(high-performance)*	**1.2**	**1.2**	**1.1**	**1.1**	**1.1**	**1.0**	**1.0**
V_{dd} *(Low Operating Power, high* V_{dd} *transistors)*	**1.0**	**0.9**	**0.9**	**0.9**	**0.8**	**0.8**	**0.8**
Allowable Maximum Power [1]							
High-performance with heatsink (W)	**149**	**158**	**167**	**180**	**189**	**200**	**210**
Cost-performance (W)	**80**	**84**	**91**	**98**	**104**	**109**	**114**
Battery (W)—(low-cost/hand-held)	**2.1**	**2.2**	**2.3**	**2.4**	**2.5**	**2.6**	**2.7**

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

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
Power Supply Voltage (V)						
Vdd (high-performance)	**1.0**	**0.9**	**0.9**	**0.8**	**0.8**	**0.7**
Vdd (Low Operating Power, high Vdd transistors)	**0.7**	**0.7**	**0.6**	**0.6**	**0.5**	**0.5**
Allowable Maximum Power [1]						
High-performance with heatsink (W)	**218**	**240**	**251**	**270**	**288**	**300**
Cost-performance (W)	**120**	**131**	**138**	**148**	**158**	**168**
Battery (W)—(hand-held)	**2.8**	**3.0**	**3.0**	**3.0**	**3.0**	**3.0**

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

COST

Tables 7a and 7b are dedicated to cost trends. The historical ability to reduce the cost per function by an average 25–30% each year has represented 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 market cost reduction mandate, 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 an 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 challenging economic and competitive market environment continue to be resistant to even "moderate" increases in cost, putting pressure upon the semiconductor industry to slow rate of doubling functions per chip (Moore's Law) in order to keep chip and unit costs under control. The semiconductor manufacturers had to 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 constant cost targets. The *2001 and 2003 ITRS models both continue to use that model*, which results in 29% cost reduction of a function (bit, transistor, etc.). That rate of function cost reduction was achieved historically (prior to 1999) by delivering four times the functionality per chip every three years at 1.4× increase in cost per unit.

The *2003 ITRS* 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 10.5 microcents for 4-Gbit DRAMs in 2003. 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] Therefore, 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.

The *2003 ITRS* retains the original 2001 MPU chip size model. The Design ITWG updated the MPU model in the *2001 ITRS*, based upon available data. At that time, the data indicated 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 continuing to slow, as reflected in the *2003 ITRS* DRAM Chip Size Model targets. The "6" design factor, a 25% improvement over the "8" factor, was expected to be implemented in 2003, but has now been delayed five years to 2008. Furthermore the "4" design factor, a 33% improvement over the "6" factor, was changed to "5" in the *2003 ITRS*, and delayed from 2011 to 2016. Consequently the target for the cell array efficiency percentage was increased to 63% and the rate of bits per chip was slowed in the future from 2×/two years to 2×/2.5-3 years. These adjustments to the *2003 ITRS* DRAM chip size model were required in order to preserve a constant chip size target of less than 140 mm^2.

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, package, board, 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, though at a slower rate. As the number of functions/chip continues to increase, it becomes

[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.

increasingly difficult and, therefore, costly to test the final products. This issue is reflected in the escalating cost of testers. The cost/pin of testers is forecast to increase (Tables 7 a and 7b), and the also the number of tested pins (Tables 4 a and 4b). Therefore, there will be an ongoing need for accelerated implementation of Built-In-Self-Test (BIST) and Design-For-Testability (DFT) techniques within the time frame of the 2003 ITRS. Further discussion is detailed in the *Test* chapter.

Table 7a Cost—Near-term Years

Year of Production	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Technology Node		*hp90*			*hp65*		
DRAM ½ Pitch (nm)	*100*	*90*	*80*	*70*	*65*	*57*	*50*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**120**	**107**	**95**	**85**	**76**	**67**	**60**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**107**	**90**	**80**	**70**	**65**	**57**	**50**
MPU Printed Gate Length (nm) ††	**65**	**53**	**45**	**40**	**35**	**32**	**28**
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**
Affordable Cost per Function ++							
DRAM cost/bit at (packaged microcents) at samples/introduction	**10.5**	**7.4**	**5.3**	**3.7**	**2.6**	**1.9**	**1.3**
DRAM cost/bit at (packaged microcents) at production §	**3.8**	**2.7**	**1.9**	**1.4**	**0.96**	**0.7**	**0.5**
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at introduction §§	**88**	**62**	**44**	**31**	**22**	**15.6**	**11.0**
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at production §§	**53**	**38**	**27**	**19**	**13.3**	**9.4**	**6.7**
High-performance MPU (microcents/transistor) (including on-chip SRAM) at production §§	**49**	**34**	**24**	**17**	**12**	**8.6**	**6.1**
Cost-Per-Pin							
Test Cost							
Volume tester cost per high-frequency signal pin ($K/pin) (high-performance ASIC)—maximum	**3**	**3**	**3**	**3**	**3**	**3**	**3**
Volume tester cost per high-frequency signal pin ($K/pin) (high-performance ASIC)—minimum	**1**	**1**	**1**	**1**	**1**	**1**	**2**

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 to four years after introduction).

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

1999–2004/8×: 2005/7.5×; 2006–2007/7×; 2008–2015/6×; 2016–2018/5×. The delay of the "6" DRAM cell design improvement factor [a] by five years, from 2003 to 2008, requires the slowing of the addition of "Moore's Law" bits/chip from 2× every 1.5–2 years to 2× every 2.5–3 years in the 2003 ITRS DRAM Chip Size Model, which remains on a three-year DRAM half-pitch node cycle after 2004.

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 16 Gbit generation, the introduction rate is 4×/six years (2×/three years); and

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

InTER-generation chip size growth rate model target for Production-phase DRAMs is now "flat" at less than 140 mm^2, similar to the MPU model. This new flat-chip-size model target requires the bits/chip "Moore's Law" model for DRAMs to increase the time for doubling bits per chip to an average of 2×/2.5 years by alternating between 2×/2 years and 2×/3 years (see ORTC Tables 1c and d). In addition, the Cell Array Efficiency (Array % of total chip area) was increased to 63%, which also assists in the achievement of the target flat-chip-size model for the Production-phase product chip size, even under the slower design improvement factor contribution (see note above). The InTRA-generation chip size shrink model is 0.5× every technology node in-between cell factor reductions.

§§ MPU Chip Size Model—Both the cost-performance and high-performance MPUs InTER-generation chip size growth rates are targeted to be flat through 2018 (280 mm^2/cost-performance at introduction; 140 mm^2/cost-performance at production; 310 mm^2/high-performance at production). The MPU flat chip-size model is made possible by doubling the on-chip functionality every technology node cycle. The InTRA-generation chip size shrink model was 0.5× every two-year technology node through 2001, and is now 0.5× every three-year technology node cycle after 2003.

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

Table 7b Cost—Long-term Years

Year of Production	*2010*	*2012*	*2013*	*2015*	*2016*	*2018*
Technology Node	*hp45*		*hp32*		*hp22*	
DRAM ½ Pitch (nm)	*45*	*35*	*32*	*25*	*22*	*18*
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	**54**	**42**	**38**	**30**	**27**	**21**
MPU/ASIC ½ Pitch (nm) (Un-contacted Poly)	**45**	**35**	**32**	**25**	**22**	**18**
MPU Printed Gate Length (nm) ††	**25**	**20**	**18**	**14**	**13**	**10**
MPU Physical Gate Length (nm)	**18**	**14**	**13**	**10**	**9**	**7**
Affordable Cost per Function ++						
DRAM cost/bit (packaged microcents) at samples/introduction	**0.93**	**0.46**	**0.33**	**0.16**	**0.12**	**0.06**
DRAM cost/bit (packaged microcents) at production §	**0.34**	**0.17**	**0.12**	**0.06**	**0.042**	**0.021**
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at introduction §§	**7.78**	**3.89**	**2.75**	**1.38**	**0.97**	**0.49**
Cost-performance MPU (microcents/transistor) (including on-chip SRAM) at production §§	**4.71**	**2.35**	**1.66**	**0.83**	**0.59**	**0.29**
High-performance MPU (microcents/transistor) (including on-chip SRAM) at production §§	**4.305**	**2.15**	**1.52**	**0.76**	**0.54**	**0.27**
Cost-Per-Pin						
Test Cost						
Volume tester cost per high-frequency signal pin ($K/pin) (high-performance ASIC)—maximum	**4**	**4**	**4**	**4**	**4**	**4**
Volume tester cost per high-frequency signal pin ($K/pin) (high-performance ASIC)—minimum	**2**	**2**	**3**	**3**	**4**	**4**

GLOSSARY

KEY ROADMAP TECHNOLOGY CHARACTERISTICS TERMINOLOGY (WITH OBSERVATIONS AND ANALYSIS)

CHARACTERISTICS OF MAJOR MARKETS

Technology Node—The minimum half-pitch of custom-layout (i.e., with staggered contacts/vias) metal interconnect is most representative of the process capability enabling high-density (low cost/function) integrated circuits and is selected to define an ITRS Technology Node. For each Node, this defining metal half-pitch is taken from whatever product has the minimum value. Historically, DRAMs have had leadership on metal pitch, but this could potentially shift to another product in the future.

Other parameters are also important for characterizing IC technology. For example, in the case of microprocessors (MPUs), physical bottom gate length is most representative of the leading-edge technology level required for maximum performance. Each technology node step represents the creation of significant technology progress in metal half-pitch — approximately 70% of the preceding node, 50% of two preceding nodes.

Example: DRAM half pitches of 180 nm, 130 nm, 90 nm, 65 nm, 45 nm, 32 nm, and 22 nm.

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.

An official *2003 ITRS* metal half-pitch node indicator, "hpXX," has been added to differentiate the ITRS definition from commercial technology generation numbers .

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 three years, as suggested by consensus DRAM and MPU models of the *2003 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) increased density and smaller chip sizes from technology and design improvements; 2) increasing wafer diameters; 3) decreasing equipment cost-of-ownership; 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-level, Introduction-level, Production-level, Ramp-level, 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-level, Production-level, Ramp-level, Peak).

Cost-Performance MPU—MPU product optimized for maximum performance and the lowest cost by limiting the amount of on-chip SRAM level-two (L2) cache (example 1Mbytes/2001). Logic functionality and L2 cache typically double every three-year generation.

High-performance MPU—MPU product optimized for maximum system performance by combining a single or multiple CPU cores (example two cores at 25Mt cores in 2001) with a large (example 4Mbyte/2001) level-two (L2) SRAM. Logic functionality and L2 cache typically double every three-year technology generation by doubling the number of on-chip CPU cores and associated memory.

Product inTER-generation—Product generation-to-generation targets for periodically doubling the on-chip functionality at an affordable chip size. The targets are set to maintain Moore's Law (2×/two years) while preserving economical manufacturability (flat chip size and constant manufacturing cost per unit). This doubling every two years at a constant cost assures that the cost/function reduction rate (inverse productivity improvement) is -29% per year (the target historical rate of reduction). In order to double the on-chip functionality every two years, when technology-node scaling (.7× linear, .5× area) is every three years, an additional device/process design improvement of .8× per two years must be achieved. This requirement represents a design-related (cell-area-factor) area-reduction improvement of at least -11% per year, and this design-related productivity improvement is in addition to the basic lithography-based area reduction of -21% per year (three-year node cycle).

The present *2003 ITRS* consensus target for the rate of increase of DRAM bits/chip has increased from 2× bits/chip every two years to 2×/chip every two and half years average. This slower bits/chip growth is required due to the new consensus *2003 ITRS* forecast of cell-area-factor improvement of only negative 4–6% per year on average rather than the *2001 ITRS* target of -7% per year average. This results in an average DRAM inTER-generation approximately flat chip-size growth. Presently, the MPU transistor area is shrinking only at lithography-based rate (virtually no design-related improvement). Therefore, the *2003 ITRS* MPU inTER-generation functionality model target is 2× transistors/chip every technology node, in order maintain a flat chip size growth throughout the roadmap period.

Product inTRA-generation—Chip size shrink trend within a given constant functions-per-chip product generation. The *2003 ITRS* consensus-based model 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* targets for both DRAM and MPU reduce chip size within a generation by minus 50% per technology node.

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 four years at the leading-edge process technology node, typically two–three years in advance of actual market introduction. DRAM demonstration chip sizes have doubled every eight 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, versus *ITRS* 1Gb 1999 Introduction-level, 2003 Production-level targets.

Year of INTRODUCTION—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, products will be introduced at 2× functionality per chip every two years (every technology node, in the case of MPUs). In addition, manufacturers will delay production until a chip-size shrink or "cut-down" level is achieved which limits the inTER-generation chip-size growth to be flat.

Year of PRODUCTION—Year in which at least one leading chip manufacturer begins shipping volume quantities (initially, at least 10K/month) of product manufactured with customer product qualified* production tooling and processes and is followed within three months by a second manufacturer. (**Note: Start of actual volume production ramp may vary between one to twelve months depending upon the length of the customer product qualification*). As demand increases for the leading-edge performance and shrink products, the tooling and processes are being quickly "copied" into multiple modules of manufacturing capacity.

For high-demand products, volume production typically continues to ramp to fab design capacity within twelve months. Alpha-level manufacturing tools and research technology papers are typically delivered 24–36 months prior to volume

production ramp. Beta-level tools are typically delivered 12-24 months prior to ramp, along with papers at industry conferences. The beta-level tools are made production-level in pilot-line fabs, which must be ready 12–24 months prior to Production Ramp "Time Zero" [see Figure 2 in the Executive Summary] to allow for full customer product qualification. The production-level pilot line fabs may also run low volumes of product that is often used for customer sampling and early qualification prior to volume production ramp. Medium-volume production-level DRAMs will be in production concurrently with low-volume introduction-level DRAMs, and also concurrently with very-high-volume, shrunken, previous-generation DRAMs (example: 2003: 1Gb/production, 4G/introduction, plus 512Mb/256Mb/128Mb/64Mb high-volume). Similarly, high-volume cost-performance MPUs are in production concurrently with their lower-volume, large-chip, high-performance MPU counterparts, and also with very-high volume shrinks of previous generations.

Functions/Chip—The number of bits (DRAMs) or logic transistors (MPUs/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 gate-function 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* consensus models).

Functions/cm^2—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. In the *2003 ITRS*, the typical high-performance ASIC design is assumed to have the same average density as the high-performance MPUs, which are mostly SRAM transistors.

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 63% at the production level, and less than 50–55% for smaller previous generation shrunk die at the high-volume ramp level).

DRAM Cell Area (μm^2)—The area (C) occupied by the DRAM memory bit cell, expressed as multiplication of a specified *ITRS*-consensus cell area factor target (A) times the square of the minimum half-pitch feature (f) size, that is: $C = Af^2$. To calculate the chip size, the cell area must be divided by the array efficiency, a factor (E) that is statistically derived from historical DRAM chip analysis data. Thus an average cell area (C_{AVE}) can be calculated, which is burdened by the overhead of the drivers, I/O, bus lines, and pad area. The formula is: $C_{AVE} = C/E$.

The total chip area can then be calculated by multiplying the total number of bits/chip times the C_{AVE}.

Example: 1999: A=8; square of the half-pitch, $f^2 = (180\ nm)^2 = .032\ \mu m^2$; cell area, $C = Af^2 = 0.26\ \mu m^2$; for 1 Gb introduction-level DRAM with a cell efficiency of E=70% of total chip area, the $C_{AVE} = C/E = 0.37\ \mu m^2$; therefore, the 1 Gb Chip Size Area=2^{30} bits * 0.37e-6 mm^2/bit = 397 mm^2.

DRAM Cell Area Factor—A number (A) that expresses the DRAM cell area (C) as a multiple of equivalent square half-pitch (f) units. Typically, the cell factor is 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.).

SRAM Cell Area Factor—Similar to the DRAM area factor, only applied to a 6-transistor (6t) logic-technology latch-type memory cell. The number expresses the SRAM 6t cell area as a multiple of equivalent square technology-node half-pitch (f) units. Typically, the cell factor of the SRAM 6t cell is 16–25 times greater than a DRAM memory cell area factor.

Logic Gate Cell Area Factor—Similar to the DRAM and SRAM cell area factors, only applied to a typical 4-transistor (4t) logic gate. The number expresses the logic 4t gate area as a multiple of equivalent square technology-node half-pitch (f) units. Typically, the cell factor of the logic 4t gate is 2.5–3 times greater than an SRAM 6t cell area factor, and 40–80 times greater than a DRAM memory cell area factor.

Usable Transistors/cm^2 *(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.

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. The specification represents the minimum specification that a semiconductor manufacturer might specify for a given technology node. The maximum field size may be specified higher than the ORTC target values, and the final exposure area may be achieved by various combinations of exposure width and scan length.

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 ITWG, are based on the first 20K wafer-starts-per-month manufacturing facility.

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.