ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY

Custom VLSI Circuits for High Energy Physics

Sherwood Parker
Physics Division
and University of Hawaii

June 1998

RECEIVED
AUG 27 1998
OSTI

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

This report has been reproduced directly from the best available copy.

Available to DOE and DOE Contractors
from the Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (615) 576-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, VA 22161

Ernest Orlando Lawrence Berkeley National Laboratory
is an equal opportunity employer.

LBNL-41444
University of Hawaii #UH-511-895-98

RECEIVED
AUG 27 1998
OSTI

Custom VLSI Circuits for High Energy Physics

Sherwood Parker
University of Hawaii

June 1998

MASTER

This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

**Portions of this document may be illegible
electronic image products. Images are
produced from the best available original
document.**

Contents

I. Introduction	1	Table 3. Failure Acceleration Factors for Silicon ICs	47
II. Advantages & Disadvantages of ASICs.	1	L. Assembly & Packaging	47
III. First-time User's Guide	4	L1. Dicing & Die Bonding	48
A. Basic References	4	L2. Attaching Leads	49
B. Components	7	L3. Encapsulation & Packaging	51
B1. Conductors & Resistors	7	M. Yield	53
B2. Capacitors	8	Table 4. Sources of Yield Loss	54
B3. Diodes	9	IV. Survey of ASICs in High Energy Physics	55
B4. Contacts	9	Table 5. Tabulated Survey	56
Table 1. Useful Constants	10	V. Technology Trends	62
Table 2. Some Formulae	11	A. Scaling	62
B5. Bipolar Transistors	14	Table 6. Scaling Changes	63
B6. Junction Field- Effect Transistors	15	B. Progress & Limits	64
B7. MOS Transistors	16	VI. New Technologies	65
B8. Scratch Masks	19	A. Gallium Arsenide	66
C. Noise	19	B. Band-Gap Engineering	67
D. Types of ICs	21	C. Fiber Optics	68
E. Fabrication	24	VII. Suggestions for Future VLSI Development	68
F. Input Protection	27	VIII. Non-conclusion	75
G. Designing ICs - Foundries	28	Glossary	76
H. Designing ICs - Layout	29	Additional References	81
I. Designing ICs - Simulation	32	Acknowledgments	98
J. Effects of Radiation	34	Figures	99
K. Testing	40		
K1. Designing for Testability	41		
K2. Apparatus for Testing	41		
K3. Test Structures	42		
K4. Stress Testing	44		

Custom VLSI Circuits for High Energy Physics

Sherwood Parker

University of Hawaii

Abstract

This article provides a brief guide to integrated circuits, including their design, fabrication, testing, radiation hardness, and packaging. It was requested by the Panel on Instrumentation, Innovation, and Development of the International Committee for Future Accelerators, as one of a series of articles on instrumentation for future experiments. Their original request emphasized a description of available custom circuits and a set of recommendations for future developments. That has been done, but while traps that stop charge in solid-state devices are well known, those that stop physicists trying to develop the devices are not. Several years spent dodging the former and developing the latter made clear the need for a beginner's guide through the maze, and that is the main purpose of this text.

I. Introduction

Understanding the physics of high energy particle interactions needs an ever increasing number of detector elements and channels of electronics as the center of mass energy increases. In many cases, the spatial density and device output impedance also increases, necessitating the development of compact electronics that can be placed directly on the detectors. At the same time, both the difficulty of designing a typical custom or application specific integrated circuit (ASIC), and the cost and time required for its production have decreased. This article is intended to be both a guide to the field for a first-time user and a brief review of some of the developments in the field. Introductory texts are listed in case the reader is not familiar with the most common types of *discrete* semiconducting components: diodes, bipolar transistors, junction field effect transistors (JFETs) where a silicon gate is insulated from the channel by a reverse biased pn junction, MOSFETs where the gate is insulated by an oxide layer, and metal - semiconductor (MESFETs) transistors, where the gate insulation is a reverse biased Schottky diode formed between a metal gate and the underlying semiconductor. However, brief descriptions of each and several useful formulas will be given. Unless explicitly stated otherwise, all circuits are assumed to be of silicon.

II. Advantages and Disadvantages of Application Specific Integrated Circuits.

Some of the advantages are immediately obvious. They include:

1. Less space and usually less power is required.

2. Interconnections, often the least reliable part of a device, are minimized. At the same time, speed is improved, due to decreased interconnection capacitance.
3. Incremental cost is greatly reduced. For example, discrete amplification and readout circuits for silicon strip detectors cost about \$20 per channel. The first custom chip to be used to read out silicon strip detectors, the 128 channel Microplex [1], [2] (see "Additional References") had an incremental cost of \$0.03 per channel.
4. You are riding an express train. Tens of thousands of people are working overtime just so you (and a few hundred million others) can get better chips.
5. It is easier to get precisely matched components.
6. Better computer-aided design (CAD) tools are available for integrated circuits than for discrete component circuits.

Some are less so. For instance:

7. With discrete circuits, you sketch the circuit, search parts catalogs, choose the most appropriate transistors available (often not quite what you really wanted), and design the circuit. With ASICs, you sketch the circuit, select the best transistor sizes for each specific application, and design and lay out the circuit.
8. Often cell libraries can be used, greatly speeding circuit design. (The price, however, will often be increased chip area and somewhat decreased performance, compared to a full-custom design.)
9. Reduced dimensions lower the capacitive loads, permitting increased circuit speed and decreased power.
10. With discrete circuits, stray capacitances and inductances usually cannot be accurately calculated. This may lead, particularly with circuits having feedback (intentional or otherwise), to unexpected oscillations. Integrated circuits are so planar, so thin, even compared with their small lateral dimensions, that capacitive coupling can be well calculated; inductive coupling is usually negligible. This permits a more accurate calculation of the circuit parameters. Well tested computer programs are available that can predict circuit performance from these parameters.
11. Good placement, routing, and shaping of circuits and their elements can make dramatic improvements. If you enjoy visual puzzles, you will probably enjoy IC design. And if you do -- a suggestion -- work *with* the chip designer(s), perhaps doing the simulations or some part of the circuit yourself. The double-checking the circuit will receive (and probably need) will be far better than anything coming from routine meetings -- however many.

Some disadvantages are obvious, or will become so with great rapidity:

1. The entrance fees are high. To do a reasonably good job, significant knowledge of the fabrication and properties of integrated circuits is necessary, particularly for analog circuits. Much of the material covered in typical detector articles was developed by or for particle physicists, and is

covered, in part, in physics courses. Often several key books or review papers will provide enough information to get started. On the other hand perhaps 15 or 20 books would be needed to cover integrated circuit physics, design, fabrication technology and testing, none of which are likely to be covered in a typical physics program. (See the next section for a list.)

2. It takes much longer than you think to do anything. The time for a single cycle of circuit definition, design, layout, simulation, mask making, fabrication, and testing will probably take several months at least. Circuits now in use or currently under development for high energy physics have required two to five or more of these cycles. There are however, developments, both in design software and fabrication procedures, that may dramatically shorten this time. The first sentence will remain true.
3. Quick fixes are rare. You usually have to start the fabrication over if an error is found. (But see section III.K 2 for some exceptions.)
4. There are technology-based limitations on availability of some components. Compromises are necessary since all components must be made in the same process sequence. The final steps must not destroy the work of earlier ones, either directly, or indirectly through a gradual loss of yield due to process complexity. Due to space limitations, large value capacitors and resistors are often difficult to fabricate. (When was the last time you had to consider the dimensions of a resistor in a discrete circuit?)

And others may only become obvious afterwards:

5. There are more ways to make mistakes. For instance, missing a sneak path (a conductive path - under certain circuit conditions - from one structure to another, that you didn't think of, but the circuit did). Or neglecting electromigration (material in metal traces flows in the direction of the current when it is more than a few mA per square micron).
6. The economic health of the foundry making your chips may be fragile. It is unlikely to get rich from your business (see advantage 3 above), and may not be there (at least as a foundry for outside users) the next time you come around. For example, a typical plant may start 1000 to 5000 wafers per week. Each of the wafers for the Microplex chip held the electronics for over 10,000 channels (and wafers have since doubled in diameter, and transistor gate lengths are 10 times smaller). A large silicon vertex detector might use 1% of one week's output. Making *all* the front end electronics for a large collider detector might keep the plant busy for about a week. Keeping fabrication sources available, however, may not be too serious a problem. An organization called MOSIS, has a partial solution for it (more later).

III. A First-time User's Guide To The Field

A. Some Basic References

Except for some of the classics, there is inevitably a fair amount of randomness that enters in selecting this short list from the many books available. (For example, not even counting course text books, there are over 360 separate titles on integrated circuits on the shelves at the Stanford University bookstore.)

Some good introductory books are:

- [3] Microelectronics, a Scientific American book, 1977. It originated from a single topic issue. As is the case with most Scientific American books, it is well organized and written, although, of course, it does not have the most recent developments.

- [4] Modular Series on Solid State Devices, Robert Pierret and Gerold Neudeck, editors, Addison-Wesley, 1983-8. Seven (thin) volumes including one on the solid state course you probably wish you had, set at just about the right level for high energy physicists. Others cover devices such as diodes, bipolar transistors, field effect transistors and fabrication. Much of the material has been combined into Semiconductor Device Fundamentals by Robert Pierret, also published by Addison-Wesley.

- [5] Operation and Modeling of the MOS Transistor, Yannis Tsividis, McGraw-Hill, 1987.

Rather detailed for an introduction, except it goes to considerable effort to explain things thoroughly and carefully.

The next one isn't one to start with, but you probably won't get very far before you want to look up something in:

- [6] Physics of Semiconductor Devices, S. Sze, Wiley, 1981.

Some books on devices and fabrication technology are:

- [7] Microelectronics Processing and Device Design, R. Colclaser, Wiley, 1980.

- [8] Physics and Technology of Semiconductor Devices, A. Grove, Wiley, 1967.

- [9] Device Electronics for Integrated Circuits, R. Muller and T. Kamins, Wiley, 1986. This provides the most complete description of the four above, of the actual devices as adapted for use in integrated circuits.

- [10] Semiconductor Devices, Physics and Technology, S. Sze, Wiley, 1985.

Four books that concentrate on fabrication technology are:

- [11] VLSI Fabrication Principles, Silicon and Gallium Arsenide, Sorab Ghandhi, Wiley, 1983.

- [12] VLSI Technology, S. Sze, editor, McGraw-Hill, 1983. This has much introductory material on testing, assembly, packaging, and yield.

- [13] Silicon Processing for the VLSI Era, S. Wolf and R. Tauber, Lattice Press, Vol. 1, 1986; S Wolf is the sole author of Vol. 2 (1990) and Vol. 3 (1995).

- [14] Integrated Circuit Fabrication Technology, D. Elliott, McGraw-Hill, 2nd Ed., 1989.

The next eight books are on circuit design. The first two are classics, but were largely written before CMOS circuits became common. The others continue with digital and analog CMOS circuit design, including the detailed practical information needed for designing circuits rather than just understanding designed circuits.

- [15] Analysis and Design of Analog Integrated Circuits, Paul Gray and Robert Meyer, Wiley, 1984. The second edition has material on MOS circuit design, but at heart, it's still a bipolar book--the best perhaps in its field.
- [16] Introduction to VLSI Systems, C. Mead and L. Conway, Addison-Wesley, 1980. (Out of print). Digital MOS circuits from a single transistor to a computer. This book changed a field as well as described it.
- [17] The Design and Analysis of VLSI Circuits, L. Glasser and D. Dobberpuhl, Addison-Wesley, 1985. An advanced book on NMOS and CMOS.
- [18] Analogue IC Design, C. Toumazou, F. Lidgley, and D. Haigh, ed., Peter Peregrinus Ltd., 1990.
- [19] Principles of CMOS VLSI Design, A Systems Perspective, N. Weste and K. Eshraghian, Addison-Wesley, 1985. This and the following reference have material on the layout of CMOS circuits.
- [20] Introduction to NMOS and CMOS VLSI Systems Design, Amar Mukherjee, Prentice-Hall, 1986.
- [21] CMOS Analog Circuit Design, P. Allen and D. Holberg, Holt, Rinehart and Winston, 1987.
- [22] VLSI Design Techniques for Analog and Digital Circuits, R. Geiger, P. Allen, and N. Strader, McGraw Hill, 1990.

And after your chip is designed, fabricated, and tested, you will need to know about:

- [23] Circuits, Interconnections, and Packaging for VLSI, H. Bakoglu, Addison-Wesley, 1990. Suggestion: read this before the chip is designed.

On the vital oxide layer:

- [24] MOS (Metal Oxide Semiconductor) Physics and Technology, E. Nicollian and J. Brews, Wiley, 1982.

This is a highly specialized book on the metal-oxide-silicon system and interfaces, and would not normally be included in a general list. However, the topics covered in this book are of great importance now, and will become of even greater interest as radiation hardening of MOS electronics becomes more important in high energy physics. It has only a small section on that, but much basic information on oxides and measurement techniques that are used in the field.

There are many review articles and books specifically on radiation effects:

- [25] Mechanisms of Radiation Effects in Electronic Materials, V. van Lint, T. Flanagan, R. Leadon, J. Naber, and V. Rogers, Wiley, 1980.

[26] Basic Radiation Effects in Nuclear Power Electronics Technology, J. Gover and J. Srour, Sandia Natl. Laboratories, Albuquerque, NM, Report SAND85-0776, 1985.

[27] Basic Mechanisms of Radiation Effects in Electronic Materials and Devices, F. Barry McLean and T. Oldham, Harry Diamond Laboratories, Adelphi, MD, Report HDL-TR-2129, 1987.

The most recent differ from the others in having chapters on radiation - hardening fabrication technology.

[28] Ionizing Radiation Effects In MOS Devices And Circuits, T. P. Ma and Paul Dressendorfer, Wiley, 1989; Hardening Semiconductor Components Against Radiation and Temperature, W. R. Dawes, et al., Noyes Data Corp., Park Ridge, N.J., 1989, also has a chapter on increasing hardness with circuit design techniques.

Finally, a book with the same overall goal as this article!

[29] John Schroeter, Surviving the ASIC Experience, Prentice Hall, (1992). In a similar vein, Ed Platner of Brookhaven National Lab. has written a note, "Report on Application Specific Integrated Circuits for Relativistic Heavy Ion Detectors" [BNL 41913 (Aug. 17, 1988)]. It describes some of the problems he had in developing one of the earliest ASIC chips made for high energy physics. Both are recommended for anyone who'd rather read about problems than have them.

In addition to the IEEE professional journals, there are many useful trade journals. If you are a likely customer of any of their advertisers, many will give you a free subscription. In addition to up-to-date articles, they often provide complete lists of manufacturers and suppliers within their specific areas. A few examples are:

[30] Advanced Packaging, IHS Publishing Group, 17730 W. Peterson Rd., PO Box 159, Libertyville, IL 60048-0159.

[31] CleanRooms, PennWell Publishing Co., 1421 S. Sheridan Rd., Tulsa OK 74112.

[32] Electronic Design, Penton Publishing Inc., 1100 Superior Ave., Cleveland OH 44114-2543 Mostly about components and putting them together to make systems - - of use here once you fab your chips (and they work).

[33] Electronic Packaging and Production, Cahners Publishing Co., 1350 E. Touhy Ave., Des Plaines, IL 60018; circulation department: Highlands Ranch, CO 80126.

[34] Integrated System Design (formerly ASIC & EDIA), The Verecom Group, Mountain View, CA 94043. About custom chip design and design tools. Foundry list issues of this and similar publications are given in references [63], [64], and [79].

[35] Semiconductor International, Cahners Publishing Co. Directed at the fabrication industry. Most recent buyer's guide: "SemiSource", 20 (Dec. 1997).

[36] Solid State Technology, PennWell Publishing Co., covers topics similar to those in Semiconductor International.

Finally, while not a journal, the following publication lists manufacturers, importers, distributors, representatives, and has a yellow pages section as well. It is issued annually:

[37] Electronic Industry Telephone Directory, Harris Publishing Co., 2057 Aurora Rd., Twinsburg, OH 44087-9000.

B. Components

The following sections will assume some familiarity with the discrete versions, and primarily cover changes needed for integrated circuits.

B 1. Conductors and Resistors

Aluminum is normally used as the basic low resistance conducting layer in integrated circuits. It will be deposited, typically to a depth of about half a micron, over the entire area of the wafer by evaporation or sputtering, and then etched away to form the desired pattern. When the trace is over silicon dioxide, the first aluminum down will react with the oxygen to form aluminum oxide, bonding strongly to the glass. Even when the trace is to make a direct contact to silicon, there will be a thin layer of native oxide covering the silicon, and the reaction of that layer with the aluminum allows good penetration and contact to the bulk of the trace.

One problem not found with normal wires is electromigration: the movement, in the direction of the current, of conductor atoms. It is inherently unstable, since the narrowing of the conductor due to electromigration increases the current density, and the rate of further narrowing. Under normal conditions, keeping the density below about 1 milamp per square micron will prevent it. Adding 2 to 4% by weight of copper can produce an order of magnitude or better improvement in that limit. Titanium is also used. References [10], [11], [12], [17] and [38], among others, have further information on electromigration. For long term reliability, aluminum should be protected from moisture. The top layer on the chip, the scratch mask, has this job. More complex metallization schemes involving such metals as tungsten in contacts, and titanium, platinum, and gold traces are sometimes used.

Recently (22 Sept. 1997) IBM announced plans to replace aluminum with copper to decrease RC times and produce microprocessors with 1 Ghz clock speeds. Techniques had to be developed for three difficult tasks: deposition, etching, and protection of the underlying silicon from the lifetime shortening effects of copper. Information can be found on the web at www.chips.ibm.com and in [P. Singer, "Tantalum, Copper and Damascene: The Future of Interconnects", *Semiconductor International* 21 (June 1980) 90.]

High-value resistors can be somewhat difficult to fabricate. Aluminum sheet resistivity is about 0.04 ohms per square, for typical layer thicknesses of about 0.6-0.7 microns. Heavily implanted poly or

substrate layers (diffusion resistors) have about 10 - 20 ohms per square. ("Poly" is polycrystalline silicon. It is the kind that will be formed on amorphous substrates such as SiO₂. Grain sizes of 0.1 micron are typical.) This isn't much when you need a kilohm or megohm resistor. Sometimes the poly is coated with a layer of tungsten, and then it will only have about 1 ohm per square, making it useful as a conductor. P wells, used as conductors, can reach values of about 1K to 10K ohms per square. However, such resistors have large temperature and voltage coefficients, the latter due to voltage dependence of the size of the surrounding depletion zone. Often transistors are used to reach the higher ranges.

If large R values are particularly important a special layer of lightly doped poly can be used. Poly can have bulk resistivities of over 10⁶ ohm-cm. Doping lightly with an impurity to provide control of the exact resistance value, produces practical values in the gigohm per square range. Poly has such a large resistivity because many charge carriers are trapped or reflected at the grain boundaries, and many dopant atoms are also segregated in that chaotic region where all their bonding sites can be occupied.

Poly has limitations. The first dopant atoms find their way into fairly deep traps, and so do not ionize easily. Resistivity is then strongly dependent on temperature, varying approximately as $\exp(U/kT)$ where U can range up to .555 eV, producing a change of 5% or more per °C. As the doping is increased, higher levels are filled and the resistivity decreases rapidly, falling as much as 5 decades with a factor of 4 increase in doping, making highly precise doping control necessary. At these lower resistivity levels, ionization is easier, U drops to about .01 eV, and the temperature variation can be 0.1% per °C. Reference [39] gives more detail.

B 2. Capacitors

Capacitors of values needed for integrated circuits are relatively easy to make. Any combination of the conducting layers such as metal, poly, the silicon substrate, or source/drain layers (in MOS circuits) will give, to the nearest order of magnitude, about 0.1 fF per square micron, with the exception of poly to substrate which forms transistors in MOS technology and has a voltage dependent capacitance about 10 times larger. If large capacitors are an important part of the circuit, two layers of poly with a thin (~ 20 nm) oxide layer between can be used to give about 1.7 fF per square micron. With fast circuits, a low dielectric constant is important for the layers separating metal interconnects. Work is underway on using fluorinated oxide films for this purpose [40]. These and other possible films must be able to stand high processing temperatures, have low moisture absorption, high thermal conductivity, and good adhesion.

Heavily doped n type silicon in contact with heavily doped p type, when reverse biased, will have a thin depletion depth and a large capacity per unit area. (Doping levels are designated by superscripts, so, for example, heavily doped n type is n⁺.)

B 3. Diodes

The most commonly used diodes have adjacent structures of semiconductor and metal (Schottky diodes) or of n and p type semiconductors. In either case, a difference in the energy distribution of charge carriers in the bulk of the two regions results in a transfer of charge where they contact, and, with zero applied bias, a resulting region of space charge in the semiconductor. The energy offset in the charge distribution causes current to flow more readily for one sign of voltage bias than the other. (See Table 1 for useful constants and Table 2 for several useful formulas.) These diodes are usually used, back biased, to isolate circuit elements from the bulk. (If a diode is needed directly in the circuit, either an insulating layer will be needed to isolate both sides from the bulk, or a more complicated structure such as a bipolar junction transistor will be needed.)

B 4. Contacts

The most common form of contact in integrated circuits is between layers of metal or between metal and poly or underlying layers of single crystal silicon (called diffusion or implant layers). Some fabrication schemes also allow buried contacts between poly and diffusion. In general, contacts to silicon would behave as diodes, as discussed above. If the contact is to be ohmic, as is usually desired, the silicon under it is usually heavily doped, even if the rest of the layer is not to be. This will produce a high density of charge in the depletion layer under reverse bias. The layer will thus be very thin, and current will be able to tunnel through the barrier. Due to surface effects p type contacts are usually ohmic, even when not heavily doped.

Contact design must also avoid the problem of spiking, in which silicon along a defect dissolves in the overlying aluminum during deposition. The aluminum then replaces the silicon along the defect path, and forms a conducting channel that shorts through the underlying diode. This is particularly a problem for small MOS transistors, which must have correspondingly thin source and drain structures (see section on scaling for more detail on this point). Adding 1% silicon to the aluminum helps prevent such dissolving. Deposition of an initial layer of another metal such as tungsten also helps.

Table 1
Some Useful Constants and Properties of Materials @ 300° K.^a

Property	units	Si	Ge	GaAs	SiO ₂	Si ₃ N ₄	Si ₃ N ₄ LPCVD	Si ₃ N ₄ Plasma Assisted CVD	Diamond ^f
Atomic number		14	32	31/33	14/8	14/7			6
Atomic or molecular weight		28.09	72.60	144.64	60.06	140.28			12.01
Density	g/cc	2.328	5.3267	5.32	2.19	3.44	2.9-3.1	2.4-2.8	3.5
Radiation length	g/cm ²	21.82	12.25	12.19	27.04	26.29			42.7
Interaction length	g/cm ²	106.0	140.5	140	97.45	94.26			
Lattice constant	nm	0.543	0.566	0.565					
Atomic or molecular density	cm ⁻³	5.00x10 ²²	4.42x10 ²²	2.21x10 ²²	2.20x10 ²²	1.48x10 ²²			
Effective density of states									
conduction band	cm ⁻³	2.8x10 ¹⁹	1.04x10 ¹⁹	4.7x10 ¹⁷					
valence band	cm ⁻³	1.04x10 ¹⁹	6.0x10 ¹⁸	7.0x10 ¹⁸					
Intrinsic carrier concentration	cm ⁻³	1.45x10 ¹⁰	2.4x10 ¹³	1.8 - 9x10 ⁶				<10 ³	
Young's modulus	dynes/cm ²	1.9x10 ¹²			0.73x10 ¹²	3.85x10 ¹²			
Linear thermal expansion coefficient	°C ⁻¹	2.5x10 ⁻⁶	5.7x10 ⁻⁶	5.94x10 ⁻⁶	5x10 ⁻⁷	2.8x10 ⁻⁶			0.8x10 ⁻⁶
Melting point	°C	1412	937	1237	~1700	~1900			
Specific heat	J/(g °C)	0.76	0.31	0.35	1.4	0.17			
Thermal conductivity	W/(cm °C)	1.5	0.61	0.46	0.014	0.18			10 - 20
Thermal diffusivity	cm ² /s	0.9	0.36	0.24-0.44	0.004	0.32			
Temperature dependence	eV/°C	-2.7x10 ⁻⁴	-3.7x10 ⁻⁴	-5.0x10 ⁻⁴					
Dielectric constant		11.9	16.0	13.1	3.9 ^b	9.4	5.8-6.1	6-9	5.7
Index of refraction		3.4	3.97	3.3	1.46	2.1	2	1.8-2.5	
Mobility									
Electron	cm ² /(V s)	1500	3900	8500	20				1800
Hole	cm ² /(V s)	450	1900	400	10 ⁻⁸				1200
Diffusion Constant									
electrons	cm ² /s	35							
holes	cm ² /s	12							
Intrinsic resistivity	ohm cm	2.3x10 ⁵	47	10 ⁸	10 ¹⁴ -10 ¹⁶	10 ¹⁵	10 ¹⁶	10 ⁶ -10 ¹⁵	> 10 ¹¹
Minority carrier lifetime	s	2.5x10 ⁻³	10 ⁻³	~10 ⁻⁸					
Breakdown electric field	V/cm	~3x10 ⁵	8x10 ⁴	~3.5x10 ⁵	~10 ⁷		10 ⁷	(1-6)x10 ⁶	10 ⁷
Saturation Velocity	km/s	100		60 ^c					220
Cohesive Energy_d									
eV/atom		4.63							7.37
<e-h pairs> / μ m		107		174					36
<e-h pairs> / 0.1% X ₀ ^d		10 ⁴		3980					4500
Energy gap	eV	1.124	0.67	1.42	9	~ 4	5	4-5	5.47
Electron affinity (E _{vac} - E _{bottom of conduction band}) ^d	eV	4.05	4.00	4.07	0.9				
<Ionization energy>/electron-hole pair ^d	eV	3.62	2.96 ^e	4.2	17				13
<dE/dx>, minimum	MeV/(g/cm ²)	1.66	1.37	1.37	1.75	1.75			1.745

(SEE NEXT PAGE FOR FOOTNOTES)

^a Some entries give a range instead of a unique value. These uncertainties follow from differences among references from which these data are drawn. These references are: [6], 848-852; [8], 102-103; [9], 31-37, 54-56; [10], 513; [11], 45; [12], 121, 639.

The properties are for 300° K. and, for the three semiconductors, for very lightly doped or intrinsic ones. The temperature and doping dependence of the mobility, resistivity, and breakdown electric fields can be found in Chapter 1 of Ref. [6], Chapter 3 of Ref. [10], and Table 1.1 of Ref. [9]. An approximate formula for the avalanche field in one-sided step junctions in silicon is given in Table 2. The strong dependence of mobility in silicon on transverse electrical fields in MOS transistors is given on page 449 of Ref. [6] and discussed in more detail in Ref. [44]. The last two columns for silicon nitride are for (1) films made by low pressure chemical vapor deposition, normally used to mask active (non-field) regions during field oxidations, and (2) plasma assisted chemical vapor deposition, a low temperature process used for scratch masks, and not having a fixed 3:4 element ratio. The former films can have 4-8%, and the latter 20-25% concentrations of hydrogen. Some selected properties that can have significant variations between these films are shown. In addition, some numbers, for example, the thermal expansion coefficient for SiO₂, from tables in the seven books used to compile this table, can differ by about 5%.

^b For SiO₂ formed at high temperature in an oxygen atmosphere by the oxidation of the underlying silicon. For deposited oxide, formed at low temperature (400° K.) with an atmosphere of silane and oxygen, the dielectric constant will be about 4.5

^c Maximum velocity = 200 @ 3kV / cm for electrons.

^d Mean values are tabulated. Most probable values are ~ 3/4 of mean values.

^e Measured at 90° K. Because of its small band gap, Ge is not normally used as a particle detector at 300° K. The ionization energy per electron-hole pair decreases by about 0.8 mV / deg. K. (See Nucl. Instr. and Meth. 59 (1968) 45.)

^f C. Bauer, et al., "Recent Results from the RD42 Diamond Detector Collaboration", NIM A 383 (1996) 65.

Table 2

Some Formulae

Useful collections of formulae can be found in Grove [8] (Table 6.1, p. 207, One-sided Step Junctions; Table 7.1, p. 242 Bipolar (junction) Transistors; Table 11.1, p. 333 MOSFETs) and Colclaser [7] (Table 10.1, p. 261 MOSFETs).

Bulk Silicon

$$k = 8.617 \times 10^{-5} \text{ eV/}^\circ\text{K}$$

$$\epsilon_0 = 8.85 \times 10^{-12} \text{ Farads/Meter}$$

$$q = \text{electronic charge} = 1.602 \times 10^{-19} \text{ Coulombs}$$

$$\phi = \text{bulk potential}$$

$$E_g = \text{energy gap}$$

$$N_A = \text{acceptor concentration}$$

$$N_D = \text{donor concentration}$$

$$N = \text{larger of } N_A, N_D$$

$$n = \text{electron concentration}$$

$$p = \text{hole concentration}$$

$$n_i = \text{intrinsic hole/electron concentration}$$

$$n_i \approx 1.45 \cdot 10^{10} \left(\frac{T}{300} \right)^{1.5} \exp \left(-\frac{E_g}{2k} \left(\frac{1}{T} - \frac{1}{300} \right) \right)$$

$$n_i \approx 1.45 \cdot 10^{10} \left(\frac{T}{300} \right)^{1.5} \exp \left(-21.7 \left(\frac{1}{T} - \frac{1}{300} \right) \right)$$

T in degrees Kelvin.

for p type silicon,

$$p = N_A - N_D$$

$$n = \frac{n_i^2}{p}$$

$$\phi = \frac{kT}{q} \ln \frac{n_i}{N_A}$$

and for n type silicon,

$$n = N_D - N_A$$

$$p = \frac{n_i^2}{n}$$

$$\phi = \frac{kT}{q} \ln \frac{N_D}{n_i}$$

Drift velocity v_d and mobility μ from [44] (first reference),

where for low N ,

$$v_d = v_m \frac{E/E_c}{\left[1 + (E/E_c)^\beta \right]^{1/\beta}}$$

	electrons	holes	units
v_m	$1.53 \cdot 10^9 T^{-0.87}$	$1.62 \cdot 10^8 T^{-0.52}$	cm / sec
E_c	$1.01 T^{1.55}$	$1.24 T^{1.68}$	V / cm
β	$2.57 \cdot 10^{-2} T^{0.66}$	$0.46 T^{0.17}$	---

T is measured in °K

and for low E ,
$$\mu = \mu_{\min} + \frac{\mu_{\max} - \mu_{\min}}{1 + (N/N_{\text{ref}})^\alpha}.$$

	electrons	holes	units
$\mu_{\min}$	92.	47.7	$\text{cm}^2 / \text{V} \cdot \text{sec}$
$\mu_{\max}$	1360.	495.	$\text{cm}^2 / \text{V} \cdot \text{sec}$
N_{ref}	$1.3 \cdot 10^{17}$	$6.3 \cdot 10^{16}$	$\text{cm}^2 / \text{V} \cdot \text{sec}$
α	0.91	0.76	---

One-sided Planar Step Junctions

$$\phi_{\text{BI}} = \text{built-in voltage} = \frac{2kT}{q} \ln \frac{N}{n_i}$$

$$\text{depletion region width} = \left(\frac{2\epsilon_{\text{Si}}\epsilon_0(\phi_{\text{BI}} \pm |V|)}{qN} \right)^{\frac{1}{2}}$$

where $+(-)$ is for reverse (forward) applied voltage, V

$$\text{maximum electric field} = E_{\max} = 2 \frac{\phi_{\text{BI}} \pm |V|}{W}$$

E_{BD} = break-down electric field (avalanche)

$$E_{\text{BD}} \approx 678 [N(\text{cm}^{-3})]^{0.175} \left(\frac{\text{Volts}}{\text{cm}} \right)$$

for $N \leq 10^{17}(\text{cm})^{-3}$

Above $\approx 10^{18}(\text{cm})^{-3}$ and $E_{\text{BD}} \sim 1.4 \cdot 10^6 \text{ V/cm}$ tunneling (Zener) breakdown becomes the dominant mechanism.

Temperature dependence of reverse current ($T \approx 300^\circ \text{K}$): I doubles for $\Delta T \approx +8^\circ \text{C}$.

MOSFETs

I_{DS} = drain-source current

W = transistor width

L = transistor length

L_{ch} = channel length

V_{GS} = gate-source voltage

V_{DS} = drain-source voltage

V_t = threshold voltage

ϵ_{ox} = dielectric constant of gate oxide

ϵ_{Si} = dielectric constant of silicon

t_{ox} = thickness of gate oxide

V_{SB} = source-bulk voltage

V_{DB} = drain-bulk voltage

N_B = bulk (substrate) doping concentration

N_{SD} = source-drain doping concentration

Neglecting the effect of the substrate voltage, we have

$$I_{DS} \cong \frac{W}{L} \mu \frac{\epsilon_0 \epsilon_{ox}}{t_{ox}} \left[(V_{GS} - V_t) V_{DS} - \frac{1}{2} V_{DS}^2 \right]$$

for $V_{GS} - V_t > V$ of the entire length of the conducting channel.

If not, and where

L_{ch} = distance from source to the point

and where $V_{GS} - V_t = V$ of the conducting channel

$$I_{DS} \cong \frac{W}{L_{ch}} \mu \frac{\epsilon_0 \epsilon_{ox}}{t_{ox}} \left[\frac{1}{2} (V_{GS} - V_t)^2 \right]$$

Taking substrate charge into account,

$$I'_{DS} = I_{DS} + \frac{W}{L} \mu (2q \epsilon_0 \epsilon_{Si} N_B)^{\frac{1}{2}} \left[V_{DS} (V_{SB} + \phi_B)^{\frac{1}{2}} + \frac{2}{3} (V_{SB} + \phi_B)^{\frac{3}{2}} - \frac{2}{3} (V_{DB} + \phi_B)^{\frac{3}{2}} \right]$$

$$\phi_B = \frac{2kT}{q} \ln \frac{N_B}{N_{SD}}$$

B 5. Bipolar Transistors

Fig. 1 shows two common bipolar transistors, adapted for life on an integrated circuit. Many more types, specialized for different purposes, are made. Some are described in chapters 6 and 7 of Muller

and Kamins [9]. The transistor of Fig. 1a, a vertical npn, is the preferred variety. Electrons have a higher mobility than holes, increasing device speed compared to pnp transistors. The main parts of the emitter, base, and collector taking part in transistor action are arranged vertically, providing a thin base located away from the surface. The base contact to its aluminum trace is on the left, and that of the collector is on the right. The base doping must be high enough to form a low resistance path to the aluminum (as well as an ohmic contact), but low enough so the main current across the emitter-base junction is from electrons diffusing through to the collector.

The heavily doped, buried layer, unique to integrated circuit technology, forms a low resistance structure to bring the collector current in a vertical transistor laterally out towards the n^+ collector contact. For extra low resistance, the contact may extend all the way down to the buried layer. The epitaxial layer, deposited on the substrate from the decomposition of a gas such as silane, must be lightly doped, and therefore of low conductivity, so that the depletion region of the reverse biased base - collector junction is thick enough to withstand voltage breakdown. Also, the base doping must not be so low that the entire base thickness is depleted, since an electric field would then directly connect two n regions, resulting in a large current that could not be controlled by the base. This condition is known as "punch-through". The isolation structures, here formed from the reverse biased p regions on each side and below the transistor, are also found only on integrated circuits.

The substrate thickness depends on the diameter of the wafer from which the chip was cut, and is a compromise between the desire to save silicon and the need to have a robust wafer that will not distort while placed on edge during high-temperature processes or break during normal handling. A typical thickness for 10 cm diameter wafers is 550 microns and for 20 cm wafers is 725 microns [41]. Since the active elements are almost always in the upper few microns, the thickness is normally not important. One exception is a new class of devices now being developed as particle detectors (see the material on pixel devices in Section VII).

If pnp transistors are also needed, the lateral one shown in Fig. 1b can be fabricated without adding additional process steps. The emitter and collector are made from the same p layer that is used to make npn bases, and the base is made from the n layer used for npn collectors. The freedom from extra steps increases yield and reduces costs. As usual in a world with no free lunches, there are some disadvantages. Besides losing the advantages of the vertical npn mentioned above, the light doping of the n layer, optimized for an npn collector, makes it susceptible to punch-through. To prevent that, a large spacing must be used between the emitter and collector, which lowers the transistor speed. Also, they have current gains of 10 to 100, rather than the values of 100 to 1000 common with vertical npn transistors.

B 6. Junction Field-Effect Transistors

Fig. 2 shows an n -channel junction field-effect transistor (JFET). Current flows between the n^+ source and drain, passing through the channel whose effective conducting thickness can be controlled by the

voltage on the p^+ gate electrode. Changing the reverse bias across the gate-channel diode alters the size of the depletion region, and thus the thickness of the remaining undepleted silicon which conducts the current. All but the top few microns of the chip is a p^+ substrate which lies just under the channel and forms a reverse biased diode that makes the bottom of channel. A metal layer in contact with the substrate covers the bottom of the chip.

The n type layer is formed from epitaxial (or epi) material which is deposited on the substrate, usually by the thermal decomposition of a gas such as SiH_4 , $SiHCl_3$, or $SiCl_4$. By adding a dopant such as arsene to the gas, a uniform and carefully controlled material can be made. The p^+ gate and n^+ source and drain regions are formed from two additional implants.

An alternate design with control gates below as well as above the channel is shown in Fig. 1.26 of Ref. [9]. The double sided device provides a larger current swing for a given voltage change but requires more fabrication steps and chip area. Figs. 2b and 2c show the transistor as the drain voltage is increased. When the $I \cdot R$ voltage drop from channel current increases, the increasing gate and substrate to channel voltages cause an increase in depletion widths, a decrease in channel width, and a consequent decrease in the rate of current increase. Channel pinch-off is shown in Fig. 2b. Current still flows to the drain, since the longitudinal component of the electric field pulls electrons out of the end of the channel. Further increases in drain voltage as in Fig. 2c, appear mostly in the high-field depletion region between the channel and the drain, so the further increase in current, due now in part to the shortening of the channel, is slow.

I-V characteristics are qualitatively similar to those of the more common MOS transistors. Since the JFET current does not have to flow along an oxide - silicon boundary, where capture and release of carriers at interface states is common, JFETs tend to have less noise. They also tend to have less damage from ionizing radiation (see section on radiation hardness for details). However, they take more room, and are more difficult to make. The gate diffusion, in particular, must have a carefully controlled depth. The gate input resistance is higher than that of a bipolar transistor base, since it is only connected to a back biased diode, while the base also contacts a forward biased one. It is lower, however, than that of a MOS transistor, due to current leakage across the diode.

B 7. MOS Transistors

Fig. 3 shows the cross section of a p channel metal-oxide-silicon (PMOS) transistor. The p^+ source and drain regions are heavily implanted ($\sim 10^{19}$ boron atoms/cc), both to provide low resistance and to make ohmic contacts with the overlying metal layer. The substrate, often formed directly from the bulk material of the wafer itself, is less heavily doped ($\sim 10^{14}$ to 10^{15} atoms / cc -- around 10 ohm-cm) n type material. When the source and drain are biased negatively with respect to the substrate, reverse biased n-p junctions will form with depletion regions indicated by the letters "d", as in Fig. 3b, isolating them from other structures. Even when they are at different voltages, no significant current will flow between them so long as the depletion regions do not touch. The heavier n-type field or

channel-stop implant insures that the region under the field oxide does not become inverted and short adjacent transistors together.

A thin layer of oxide (typically a few hundred Angstroms thick) is grown on top of the channel, and a gate electrode, now usually made of polysilicon that will be heavily implanted (for good conductivity), is grown on top of the oxide. The use of poly, rather than metal, permits high temperature steps to follow gate fabrication. With a potential that is sufficiently negative, holes can be attracted to the oxide - silicon interface just below the gate, and form a continuous p conducting channel between the p source and drain as in Fig. 3c.

The oxide coating has played a key role in silicon technology, and the inability to grow an oxide coat of comparable quality on GaAs or Ge is one of the things limiting their wider use. GaAs transistor gates are isolated from the substrate by use of a reverse biased diode (hence the name MESFET -- Metal Silicon Field Effect Transistor -- rather than MOSFET -- Metal Oxide Silicon FET). When a silicon crystal is cut to make a wafer, the dangling bonds at the surface produce states that fill the normal 1.12V energy gap between the conduction and valence bands, and make the surface conducting. Properly made oxides (for instance, ones formed by exposing the silicon surface to oxygen at about 1000° C) tie up these bonds. They also serve as a barrier to impurities and are used to insulate gates from the underlying substrate, as well as metal paths from both polysilicon and substrate layers. Oxygen with steam is usually used to speed up the fabrication of thick oxide coatings, as the concentration of oxygen on the surface, and so its diffusion rate into the silicon dioxide to the silicon interface, is increased.

The thick oxide (usually 0.6 microns or more) between transistors (see Fig. 3) is called the field oxide. The increased thickness (over that of gate oxides) together with the channel stop implant. prevents signals on any overlying conductors from turning on the field transistor formed by that conductor, the oxide, and the substrate, and shorting together adjacent legitimate transistors. It is also the primary impurity barrier. Over the channel, where there is only a thin gate oxide, the polysilicon gate serves as the barrier. (It was the development of effective barriers against, and control of impurities, particularly sodium, that made MOS transistors practical [42]).

If the drain-source voltage difference is made large enough to deplete the entire gap between them as in Fig. 3d, the depletion region field connects electrodes of like type (rather than of opposite type as in the case of a pn diode). For a large range of gate and substrate voltages, a punch-through current may flow between them, even though the gate is biased to a normally-off voltage. Punch-through is usually prevented by using low resistivity silicon with enough dopant ions to prevent full depletion at normal operating voltages. Making the channel doping level too high, however, will shorten the depletion distance between sources and drains and the substrate, increase their capacitances to ground, and thus produce slower circuits. The channel doping is also adjusted to set the gate voltage, called the threshold voltage, at which the induced hole density equals the electron density in the bulk material, and at which effective conduction starts.

In NMOS transistors, all doping types are reversed, and electrons conduct the current in the channel. Their characteristics are similar to equivalent PMOS transistors, except, since electron mobility is between 2 and 3 times greater than hole mobility, currents at equal voltages are proportionately greater. Typical voltage-current characteristics of an enhancement-mode (one which requires a non-zero gate to source voltage to turn on strongly) NMOS transistor are shown in Fig. 4 (top). The increase in current with gate voltage is due to the induction of more charge carriers in the channel which terminate the lines of force from the gate. Not all terminating charges, however, are mobile. When the substrate voltage is varied, the size of the depletion region under the gate changes, changing the net substrate charge as mobile carriers are removed. Field lines from the gate now terminate on fixed charges, rather than on current carriers. This is the cause of the weak (relative to gate voltage changes) variation shown within each group of three lines.

Following any one line, when the drain to source voltage is increased, the drain current first increases rapidly. Once the $I \cdot R$ voltage drop causes the channel voltage to fall below threshold, a depletion zone starts. As in JFETs, further voltage increases appear mostly across that depletion zone as in Fig. 3e, producing the relatively slow increases shown on the right side of Fig. 4 (top). Note that a MOS transistor has *four* terminals: source, drain, gate, and substrate. The source and drain can be identical; which is which depends on their relative voltages--drains are relatively positive in NMOS and negative in PMOS. (Special doping profiles, however, can be used for drains to keep electric fields low and reduce gate oxide damage from electron impact.) The current also depends on the cross section of the channel under the gate. Its thickness is of the order of kT/qE where q is the electron charge and E is the perpendicular component of the electric field in the channel. (See Ref. [43] for a calculation of the thickness of the conducting layer in a MOS transistor.)

Depletion mode transistors are fabricated with a conducting channel: *n* type for NMOS transistors and *p* type for PMOS. They can be turned off by depleting the channel. In the case of NMOS transistors, the characteristics will be similar to the ones in the figure, but the corresponding gate voltages will be perhaps 4 or 5 volts lower. As mentioned earlier, such transistors, with the gate permanently tied to the source, can be used as resistors.

Even below threshold, some current flows from the source to the drain, if they are at different voltages, due to diffusion through the channel between them. The diffusion, driven by source-drain concentration differences, is similar to the emitter-to-collector current in bipolar transistors, with the channel playing the role of the depletion regions between the base and collector electrodes. The concentrations are exponentially dependent on the channel potentials at the source and drain. Usually, the drain potential is so large that the carrier density there is nearly zero. Then only the source and gate voltage determine the current as shown in Fig. 4 (bottom).

The formulas for the current given in Table 2 are the standard ones derived in many texts. More accurate treatments are given in references [4], [6] and [17]. One particularly important correction is the

variation of mobility with both longitudinal and transverse electric fields. The former is given in the first reference of [44]. The latter is shown in Fig. 5, (see remaining references [44]). The large decrease with increasing transverse field is due to the increased scattering of the carriers from the oxide - silicon interface as they are pulled closer to the oxide. This variation is currently determined only by a fit to experimental data.

B 8. Scratch Masks

The scratch mask is a normally transparent layer about one micron thick that protects the chip from moisture and other contaminants as well as from mechanical damage. It can be made of phosphorus glass (by reacting silane, PH_3 , and oxygen), silicon nitride (from dichlorosilane and ammonia), or polyimide. Phosphorus glass has a low melting point, so it flows easily over the lower circuit levels which are generally not completely planar. Sodium, an ever-present enemy, is especially soluble in glass containing P_2O_5 , and so a layer of phosphorus glass over a non-phosphorus one serves as an effective barrier. The nitride is mechanically strong. Both it and polyimide have good conformal coverage, freedom from pinholes, and form good barriers to sodium and moisture. However hydrogen, released in their fabrication may affect the behavior of the chip under radiation. (See the section on radiation hardness). Any of them, even though they are insulators on the top of the chip, may affect the circuit performance by charging up. Since probing internal nodes requires the scratch mask be left off or removed, at least in part, some care may be needed in the design of even this almost trivial layer.

C. Noise

Noise is discussed in [45] and [46] as well as in several of the basic references already listed, including references [15], [17], and [20]. Three important sources of noise in MOS circuits are $1/f$ noise, shot noise, and white or thermal noise.

The first is due to the capture and release of charge carriers at the gate oxide-silicon interface. This produces steps in the current flow which will have a Fourier transform proportional to $1/f$ where f is the frequency. (The Fourier transform of a square wave of length $2t$ is proportional to $(\sin \omega t)/\omega$; multiple t 's will tend to average out the sine factors and leave the $1/\omega$.) The mean square value of the noise current per unit frequency interval, $\langle i^2 \rangle$, is given by:

$$\langle i^2 \rangle df = K \frac{I_{dc}^a}{f^b} df \quad (1)$$

where K is a constant, and a and b are empirical constants defined by this equation and are approximately equal to 1. Since this type of noise depends on the presence of traps at the interface, it

can vary significantly with different fabrication procedures. Increasing the gate area reduces its magnitude. JFETs, having no oxide-silicon interface at their channel boundaries, have less $1/f$ noise than comparable MOS transistors. PMOS transistors generally have less than comparable NMOS ones.

Shot noise is due to the quantized nature of the charge carriers. It can be important when small amounts of charge are involved, for instance in the leakage current in silicon diode ionization detectors. The mean square value within the frequency band Δf is given by:

$$\langle i^2 \rangle \Delta f = 2q \langle I \rangle \Delta f \quad (2)$$

where q is the charge of an electron, and $\langle I \rangle$ is the average current. The flat frequency spectrum comes from the underlying randomly spaced delta functions of current flow. The actual spectrum will tend to cut off at times characteristic of electron traversal of the relevant structures, usually depletion regions, in the case of integrated circuits.

Thermal noise is developed across any resistive element (R ohms) in the circuit. The mean square noise voltage within the frequency band Δf is equal to:

$$\langle V^2 \rangle \Delta f = 4kTR\Delta f \quad (3)$$

The largest resistances in most circuits are the transistor channels (ignoring reverse-biased diodes and gate leakage resistances). Their noise, referred to an equivalent voltage at the input gate, can be approximated by substituting $0.7/g_m$ for R , where g_m is the transconductance, dI (drain) / dV (gate). The limits on this approximation are discussed in a paper by Klaassen and Prins [45]. At high frequencies this noise source dominates $1/f$ noise. The transition or corner frequency where the descending $1/f$ curve meets the white noise plateau depends on circuit, layout, and fabrication details, but is often in the range of 1 KHz to 1 MHz.

One common configuration used for particle physics is a detector of capacitance, C , in parallel with a reset transistor of on-resistance, R . The thermal noise generates a current which leaves a charge on the capacitor when the reset is turned off, and thus contributes to the final noise. Smaller values of R will make lower voltages, but the faster RC response time increases the high frequency currents resulting in a mean square charge integrated over all frequencies, kTC , that is independent of R . The finite value of the upper frequency limit of the circuit will tend to reduce the noise somewhat below kTC . More complex circuits will have additional noise sources (for example, see reference [47] p512). Others may use special noise-reduction techniques such as double-correlated sample and hold (some examples are the circuits described in references [48], [2], and [49]) or special devices (see reference [50]). The decrease of signal voltage at the amplifier input with increasing detector capacitance is common to all, so even with these techniques, some degradation should be expected.

D. Types of Integrated Circuits

Among the characteristics that distinguish various types of integrated circuits from each other are the types of transistors used, the method employed to isolate circuit elements from each other, the minimum feature size, and over-all circuit size. The smallest element size in most MOS circuits currently ranges from approximately one third to two microns. The definitions of the terms MSI, LSI, and VLSI, referring to medium, large and very large scale integration have tended to change somewhat over time as the minimum feature size has decreased, and chip size, more slowly, has increased. Large circuits now can have millions of transistors, and overall chip sizes range from about 1 mm to over 1 cm.

Most integrated circuits use either bipolar or MOS transistors. Recently, the two have been combined on the same chip, at the price of increased fabrication complexity and some compromises on device performance [51]. Two other types, the MESFET and JFET, are more difficult to fabricate and use, but have specialized properties which may enable them to play an important role in future high energy experiments. The control element in vertical bipolar transistors, the base, is made by diffusing or implanting an n or p type region between emitter and collector regions of the opposite type. An extremely thin base layer can be made by these processes, and the resultant transistor can be fast. The control region in MOS electronics, the channel under the gate that connects the source and drain, on the other hand, runs in a horizontal direction, and can be no shorter than the minimum feature size [52]. MOS circuitry tends to be somewhat slower than bipolar, has higher $1/f$ noise, and lower transconductance. Gallium arsenide MESFETs, with their higher carrier mobility, however, can operate at high speeds. The same differences in geometry between bipolar and field effect transistors allow bipolar circuits, in the emitter follower mode, more readily to have a low output impedance, and so to be better suited for driving heavy loads. On the other hand, MOS transistors use less space, can store charge efficiently, and digital MOS can be designed to use far less power. At low power, MOS can have a larger gain-bandwidth product than bipolar since the capacity of a minimum geometry drain is usually less than that of a minimum size collector.

CMOS digital circuits use matched sets of NMOS and PMOS transistors to achieve low power consumption, increased speed (often), and decreased sensitivity to noise, to supply voltage variations, and to temperature changes. Figs. 6a and 6b show NMOS and CMOS inverters. When V_{in} is high, the NMOS enhancement transistor is on and V_{out} is low. The resistance of the always turned-on depletion transistor, which serves as the load, cannot be too small or V_{out} will not be low enough. But it cannot be too large, or the rise of V_{out} when V_{in} goes low will be too slow. In the CMOS circuit, the PMOS transistor is off when V_{in} is high, so it can be any size. When V_{in} goes low, it turns on and can bring V_{out} high rapidly. Thus, even though the hole mobility is lower than the electron mobility, we can have a fast circuit with a large output swing. Once the CMOS transition is complete, no current flows: one transistor or the other is off. The NMOS inverter, on the other hand, conducts whenever V_{in} is high.

The abrupt transition improves the noise immunity of CMOS logic circuits. The input voltage can range from zero to close to the transition voltage V_t , while the output stays near the V_{DD} high level, and from slightly above V_t to the full V_{DD} while the output remains near zero. Thus noise lower than $V_{DD}/4$, in some designs, added to an output, will not change the result of the next gate for which it is an input.

Fig. 6c shows a cross section through a CMOS inverter. The p transistors are set directly into the substrate, while the n transistors are in a p well, which is then set into the substrate. The p wells must be charged negatively with respect to the substrate to back bias the junction and prevent leakage between the two. Thus the n transistors in the inverter will normally be at a lower voltage with their sources tied to ground, while the sources of the p transistors will be connected to the positive voltage, V_{DD} . Both sources and drains will be reverse biased with respect to their immediate substrate to prevent leakage.

Other arrangements can be used, for instance n wells in a p substrate, and "twin tubs" -- which uses both n and p wells. Transistors in wells usually have somewhat degraded mobility compared to ones set directly in the wafer material, so if comparable performance is important, it might be best to put the intrinsically faster n channel devices in a p well. Logic and memory chips that use one type predominantly might be best off making it n type, and using an n well for the smaller number of p channel transistors. Twin tub technology allows separate tuning of the channel doping profiles, so better control of the lateral spreading of fields is realized, an advantage for circuits with small feature sizes.

Another comparison of NMOS and CMOS circuits is shown in Figs. 6d and 6e. The CMOS transmission gate shown in Fig. 6e has similar strengths compared to the NMOS gate (Fig. 6d). If V_C , the control voltage, is high, the gate conducts. However, if V_{in} is also high, the drive voltage ($V_{gate} - V_{channel}$) is lowered, reducing the channel conductivity of the NMOS gate. In the CMOS gate, the parallel PMOS transistor has its drive voltage raised when V_{in} is high, and so carries the bulk of the current. The reverse holds for V_{in} low: the "on" CMOS gate is always ready to conduct strongly, in this case by using the NMOS transistor. Also, since the two gate signals have opposite polarity, their capacitive feed through to the output tends to cancel.

Two general types of circuits are commonly found in digital logic: dynamic and static. The former stores information in the form of charge stored on isolated structures such as transistor gates. The latter stores it in the form of circuits, such as flip-flops, that use feedback to latch themselves in some definite state. Dynamic circuits usually need to have the charge renewed periodically, are often more prone to radiation damage, and are usually slower. However, they normally need far fewer transistors to perform a given function.

CMOS also provides advantages for analog circuits. For instance, far higher gain is available from a single stage of amplification since the "load resistor" is now a transistor which can be biased in or

near saturation. Here it can have a large dynamic resistance and still carry enough current to drive the output as well as its own capacitive load. This in turn may eliminate the need for networks that reduce the gain for high frequencies. Such networks are used to prevent oscillation in multi - stage feedback amplifiers. At a sufficiently high frequency, the cumulative delay time, measured in cycles, becomes equal to one half or more. Then, a circuit with negative feedback at low frequencies may have positive feedback at high frequencies.

The price to be paid for these advantages is an increase of up to a factor of two in the number of transistors, a significant but smaller increase in circuit area, a more complex fabrication process (for instance, a 7 mask process may increase to 12), and some brand new problems. One example is latchup, in which the transistors, substrate and well form parasitic pnp and npn transistors. Currents flowing through them, as indicated in Fig. 7, may result in positive feedback and breakdown. Lowering the resistance of the substrate paths and increasing the distance from the drain - source to the well - substrate boundary (to increase minority carrier recombination) make this condition less likely. However, increasing substrate conductivity reduces the depletion depth and increases the drain - source capacitance. Inserting highly conductive guard bands (shown dotted) would lower the resistance without greatly increasing the capacity, but would increase the circuit area. One way to lower the resistance without increasing the capacity to the substrate is to use wafers made of highly conductive silicon (less than 0.1 ohm cm) with a top epi layer of normal conductivity.

Since the components of an integrated circuit rest on a common semiconducting substrate, some way must be found to insulate them from each other. Two basic methods are in common use. Each element to be isolated is surrounded either with a back biased pn junction or with an insulating layer. The former is easiest to fabricate and will be described later. Insulating layers can be formed by a number of methods, for instance:

1. by depositing silicon on an insulating substrate (SOI - silicon on insulator) such as sapphire (SOS),
2. by implanting oxygen through the silicon to form an underlying layer, or,
3. by oxidizing the silicon surface, depositing a mechanically supporting layer of polysilicon, inverting the wafer, and etching islands of silicon on the now underlying oxide layer.

SOI methods completely eliminate, rather than just reduce, the probability of latchup, and also provide low capacity to the substrate. However, the parasitic pnp or npn substrate transistors remain, and may cause leakage to the bulk to be multiplied by the transistor beta.

There are several additional problems with SOI technology. If the substrate potential is left floating and the silicon is not fully depleted, the threshold voltage will change with changes in drain voltage, putting an undesirable kink in the current -- voltage characteristics. In SOS, where the insulating substrate is sapphire (Al_2O_3), the lattice mismatch with silicon introduces a stress and silicon defects which reduce mobilities. Silicon grown on amorphous silicon dioxide nucleates from

many centers to make a poorly conducting polycrystalline layer, and must be recrystallized by some means. Forming the oxide layer by implanting oxygen (as in 2 above) removes this problem but adds another: crystal damage from the oxygen tracks that is difficult to remove. Generally, 1 and 2 produce transistors with higher leakage currents.

A fourth method that might be free of many of these disadvantages involves oxidizing a wafer, bonding on a second one using pressure and heat, and then grinding it down so the silicon above the oxide is of the desired thickness. Controlling that thickness, particularly for thin layers may require special techniques. Reference [53], a review article on CMOS technology trends, has material on other isolation methods.

The types of integrated circuits listed above are the ones most likely to be used for application specific uses. Within each type, special features may be added for specific purposes, such as the addition of an extra layer of poly, separated by a thin layer of oxide, for the large value capacitors mentioned earlier. In addition, there are a number of types, not described in this article, that are used for specialized purposes. For example, EEPROMs, electrically erasable programmable read-only memories, use a special poly gate that is isolated from all other conducting structures with a very thin oxide layer. Its charge will stay indefinitely, but can be changed with a tunneling current induced by a sufficiently large applied voltage.

E. Fabrication

A typical integrated circuit fabrication process may require several thousand steps. However, only a few basic things are being done. The surface must be passivated with a thermal oxide in regions between transistors. Conducting layers, interspersed with insulating ones, will be made. These layers must be removed or have their properties altered in certain parts of the chip. For example, the substrate will have dopant atoms added between transistors to prevent those regions from becoming conducting. Insulating layers will have holes etched for interlayer contacts (vias). Metal and poly layers will be etched to define circuit paths. A final insulating layer, the scratch mask, will have holes etched for contacts to off-chip electronics.

The patterning for these steps requires the application of a layer of photo - sensitive liquid, the photoresist, which is usually applied to the center of a wafer. The wafer is then accelerated to a high rotation speed to produce a layer of uniform thickness. It is then baked to harden the resist, and exposed to ultraviolet light through a mask that is carefully aligned with the wafer. The resist is removed from exposed or unexposed regions (depending upon resist type) by development, and is baked again for hardening. The remaining resist then protects the underlying material from the following processes such as etching and ion implanting. Finally, the resist is removed, and the wafer is cleaned in preparation for the next set of steps. Fig. 8 shows the basic steps used in this photolithographic patterning. An alternate method under development is to deposit the photoresist directly from a gas

phase, using RF fields to make large molecules out of small ones [54]. These stick on the wafer and make a very uniform coating. Fig. 9 shows typical device cross sections after each of the major steps for a twin tub CMOS process.

Given the large number of steps and their associated variables, and since later steps, particularly high temperature ones such as oxidation, diffusion, and annealing, often affect the results of earlier ones, it is essential to first simulate them using a computer program such as SUPREM [55], [56]. If new types of structures are being constructed, as is often the case when new fabrication steps are being developed, it may be important to take the structure predicted by SUPREM, and use it as input to a program such as PISCES [57]. PISCES and other commercially available programs will calculate from the externally applied voltages and/or currents, the internal field, voltage, current and charge distributions. It allows for such things as the effects of work function differences, fixed charges, and diffusion currents, and can recognize a region in the semiconductor that will be depleted of carriers. PISCES devours computer time, and will happily swallow a supercomputer for breakfast (and the rest of the day). Commercial versions of these programs are now available [58]. They have corrected bugs, have added capabilities, are faster and easier to use. They are not inexpensive (except for academic institutions).

Major pieces of fabrication equipment typically cost about a million dollars. An entire plant, with its special air, water, and chemical supplies, and having 50 to 100 major pieces of production equipment, will cost from many millions to over a billion dollars. The rooms must be made of, and contain, only lint-free material. They need to be isolated from vibrations, with no motion from background vibration faster than 5 microns per second, and have highly filtered air with temperature and humidity control. The maximum number of particles per cubic foot larger than $1/2$ micron ranges from 10,000 in non-critical areas (class 10,000) to 1 in critical areas such as those devoted to photolithography (class 1--here the size limit is dropped to 0.2 microns). To reach the lowest levels, the air is filtered, and blown straight down from the ceiling to small holes in the floor. If a table is in the way, it too will have holes. Workers, who would otherwise emit thousands of particles per minute, are gowned in bunny suits (one piece white, lint free head to foot coveralls) with air filters. Only the cleanest parts of the machines are in the room; the rest are in another room just below or alongside.

Silicon wafers are usually purchased from specialized manufacturers who prepare the final boules of silicon from which the wafers are cut by pulling, from a melt, a starter crystal that is both rotated and pulled up as the boule grows (Czochralski growth). For special high-purity silicon (for detectors, for example), crystals are prepared by float zone growth, in which a molten zone made by a moving rf coil is passed through the silicon. The starting material for the float zone process is just polysilicon. No crucible walls touch the melt, and all significant impurities other than boron, preferentially remain in the melt, so with each passage of the coil, the purity tends to improve.

The most accurate masks, called reticles, are usually made of metal covered quartz plates which have a layer of resist that is exposed by an electron beam rather than by light. The reticles are much smaller than the wafer, and are stepped across it by an automatic mask aligner. (See Section VII, Item 8, below). This permits an adjustment to be made at each exposure point, both for the non-planarity of the wafer and for any changes in size (SiO_2 has twice the volume of silicon, so oxidation increases wafer size). Steppers with 4 or 5 to 1 reduction optics between the reticle and the wafer currently provide the highest resolution since it is easier to make defect free masks at the larger scale. The shortest wavelengths used are now at 365 nm and dropping.

Numerous optical techniques have pushed the resolution well below what most basic optics textbooks said could be done [59].

Furnaces find many uses including the growing of oxide layers, the reflow or densification of layers that have been formed on the wafer surface from the chemical reactions of vapors, the diffusion of dopant atoms from the surface into the bulk, and the annealing of the crystal lattice after ion implantation. Quartz tubes, through which appropriate gasses flow, are normally used to hold a "boat" which will contain perhaps 25 or 50 wafers. Normally, only one type of task will be done in each furnace.

Chemical vapor deposition (CVD) systems use furnaces in combination with gas control systems to form, directly on heated wafers, reaction products such as SiO_2 , Si_3N_4 , polysilicon, W, and WSi_2 . Both the glass and poly layers can be combined with dopant atoms such as phosphorus and boron. Low pressure is often used, since the longer diffusion length and thus, higher gas delivery rate, means the reaction rate is limited by the (easily controlled) wafer temperature. The addition of a plasma discharge system provides free radicals which improve bonding and thus film quality.

Dopant atoms can be inserted with precision using a technique called ion implantation. This is one of the vital advances in fabrication technology (another is plasma etching) that fueled much of the growth in the industry. It is also the one advance coming straight from high energy physics (circa 1930). It consists of an ion source followed by a mass spectrometer, an accelerating section, and a deflection system. Among the many advantages (in addition to better control of dose, location, and depth profile), the mass spectrometer provides a higher purity than is easily available from diffusion sources, and the low operating temperature allows photoresist to be used as the masking layer. One important use of ion implantation is the simultaneous doping of the gate and drain/source structures (see Fig. 9). The gate is changed from a poorly conducting material to one with about 20 ohms per square, and the source/drains are automatically aligned with the gates, since the poly is thick enough to stop all ions before reaching the channel silicon and the gate oxide. This process is often referred to as a "self-aligned" one. All implants must be followed by an anneal to jostle the dopant atoms from their stopping points to nearby active lattice sites, as well as to repair radiation damage caused by the incoming ions. Aluminum gates, used before poly was developed, had too low a melting point for this

anneal. Formed after the implant and anneal, they always had to have a large drain/source overlap (and therefore a large capacity) to ensure gate control of the entire channel length.

Sputter deposition equipment is used to deposit metal layers on the wafer. An electrical discharge, often made with a magnetron, sends argon ions into a target made of the metal to be evaporated. The impact knocks out metal atoms which cross a short gap to the substrate being coated. The process is easier to control than deposition of evaporated metal, particularly when two metals with different vaporization temperatures must be deposited together. Evaporation using electron beams, while the cleanest method of vaporizing metals, has the disadvantage of generating X-rays which can produce radiation damage in underlying oxide layers.

Plasmas are also used in modern etching equipment. Gasses are dissociated to make ions and free radicals such as F , F^+ , Cl , Cl^+ , O , and O^+ . These highly reactive species combine with unmasked parts of the wafer to make gaseous substances which are pumped away. By increasing the voltage difference between the plasma and the wafer, ions can be accelerated to the wafer and make cuts with vertical walls. The freedom from undercutting, which occurs when the substrate side walls are attacked by an isotropic etch, makes smaller feature sizes possible. In addition, high purity and good control are possible with plasma etching. Usually chlorine is used to etch aluminum (the vapor pressure of aluminum fluoride is too low), oxygen is used for organic solids including photoresist, and fluorine or fluorine-chlorine combinations are used for everything else, making such gaseous end products as CO_2 , H_2O , and SiF_4 . The sidewalls can be further protected with the use of a gas such as CHF_3 in the plasma, which can produce a teflon-like coat. When alternated with etch cycles using a gas like SF_6 , deep, highly anisotropic structures result, as the downward-directed fluorine cuts through the coat on the bottom [60].

When overlying layers must be deposited after the etch (for instance metal traces coming down into a contact hole), sloped, rather than vertical sidewalls become desirable. Either liquid or non-directed plasma etches can then be used. With both methods, etchants must be chosen that do not significantly etch either the overlying mask or the underlying layer of material. Much care goes into monitoring and process control as well as maintaining high purity. As x-rays from the plasma can cause radiation damage, fabrication steps where they may reach the gate oxide may best be done with liquids.

F. Input Protection

Static charges, particularly when the humidity is low, can generate voltages of several kV or more. Since gate oxides can withstand voltages of only about 1 V per nanometer, and are at most some tens of nanometers thick, chips must be well protected from static charges. In addition, the stored energy in a static charge can be more than enough to melt aluminum traces on the input lines. An entire mini-industry is devoted to the elimination of static charge, especially in fabrication facilities, where not-

yet protected chips must necessarily be handled, as well as assembly factories using chips in which protection networks would use too much area or add enough input capacity to compromise performance.

Techniques used include sources to ionize the air, so the appropriate sign of ions will drift to and neutralize charge on chips, moderately conducting work surfaces, gloves, and clothing, and appropriate grounding of workers and equipment.

Protection networks should be put on every input wherever possible. Any line that goes to a source or drain of a junction isolated circuit is automatically protected against excursions in the forward direction of the diode. Even in the reverse direction, in many cases the diode will break down before any gates on the same line. For lines going only to gates, or where better protection is needed, special structures, described in reference [61] can be used. Fig. 10 shows such a structure, used in the NMOS Microplex II chip [1], [2]. The input runs along a diffusion path. The diffusion-substrate diode shorts negative excursions to the substrate and the series resistance tends to block the associated current. The diffusion also acts as the drain of a transistor with a grounded gate and source that run parallel to it. The gate causes a bunching of the field lines which produces an avalanche breakdown of the transistor that will short input pulses over about 15 V to ground. The transistor width is set by the need to dissipate energy non-destructively as well as by the required resistance of the diffusion path. Another type of device uses a metal gate-field oxide transistor with the gate tied to the input. The thick field oxide makes the threshold high enough so normal positive signals do not short to ground. (Both types were used in series on the Microplex I.) However, the large thickness will cause a large threshold shift if the chip is subject to too much radiation (see section III J).

Circuit faults can also damage chips. These will generally have lower voltages and source impedances than static charges. Fig. 11 shows the Microplex I structure tested to destruction. High impedance pulses (half sine waves with 100% duty cycle) in excess of 75 volts were clamped and caused no harm. As the impedance was lowered, the heat dissipation increased until the aluminum input trace melted. Most of the series resistance was at the metal - diffusion contacts, so increasing their number in a manner that splits the current equally, would produce a more durable device if protection from low impedance sources is important. Many thousands of channels on the Microplex II and III (which use only the grounded gate protection transistor) were used for a number of years with no failures, either during installation or afterwards.

Additional opportunities and problems arise in protecting CMOS circuits. The presence of p^+ regions allows the use of clamping diodes to the V_{dd} bus for positive signals in a manner similar to the n^+ diode clamping to V_{ss} (ground). However, overload currents can now cause latchup if proper structures are not used.

G. Designing Integrated Circuits - Foundries

Fig. 12 shows typical steps in the design, fabrication, and testing of integrated circuits. Before launching into making them, one should decide whether they are necessary. There should be a need for

large numbers of circuits, or for small size (for example, for vertex chamber readout), or for things made possible by small size, such as low power or high speed, since a considerable investment must be made for design, development, testing time, and often a substantial minimum purchase will be required. Considerable effort often must be spent on the box: "Select foundry".

The last two problems can be avoided by fabricating the chips through a multi-project wafer service such as the MOS Implementation System (MOSIS) located at the University of Southern California [62], or the IC Manufacturing Service (ICMS) coordinated by IMEC in Leuven, Belgium [63]. In addition, all the steps of Fig. 12 inside the dotted box are handled by an organization such as MOSIS. With MOSIS, however, you go from a tape describing the mask layers directly to the final silicon chips. Some mask-making software makes automatic adjustments to certain mask dimensions, but does not always check afterwards with enough sophistication to be sure there are no unexpected problems. For example, it is possible, in some cases where dimensions are increased, to produce inadvertent short circuits. If this is likely, use your software to make pen plots directly from your data base, and then manually compare them with equally magnified films from the mask maker on a light table. While painful, this procedure will be less so than receiving circuits that do not work.

Doing it yourself (if you have a large enough order or have fabrication facilities at your disposal), allows more control, more checking, access to more fabrication processes, and may also provide faster turn-around. In choosing a foundry, the most recent survey from Integrated System Design [64] is a good starting point. Listed for each foundry are such items as available processes, minimum feature sizes and metal pitch, wafer sizes, maximum number of metal and poly layers, packaging, minimum order size, and typical turn-around times. Since some companies enter and leave this market frequently, depending on the demand for their proprietary products, earlier surveys might also be useful. Before making a final choice, it will be important to get an accurate and complete list of SPICE (see next section) parameters *and* their errors due to process variation. Otherwise, you will not be able to calculate reliably just how the designed circuit will behave. Surprisingly, some companies do not provide, and may not even know, all the important errors.

If you decide to design an integrated circuit, several differences with discrete circuit design will become evident.

1. The layout normally plays a larger role in the performance of an integrated circuit than it does in a discrete one.
2. More can (and must) be calculated before fabrication starts.
3. There are frequent tests that must be passed before the next design or fabrication step is started.

H. Designing Integrated Circuits – Layout

While some circuit layout work, such as the linking of predesigned cells, can be done automatically, most new circuits will require techniques that cannot be learned by rote. Nevertheless, some general techniques can be usefully described, and are covered, here and there, in references [15] - [22],

particularly [16], [19], and [21]. It is best to start with an overall plan for the flow of signals and power, and to try to make things as regular and neat as possible, with subunits matching each other. For example, a circuit with many repeating subunits can receive its power on metal lines with the interleaving scheme shown in Fig. 13a. If equality of the voltage drop across each circuit is important, the method of Fig. 13b would be better. Bus widths may be varied, and in extreme cases, individual busses for each sub - circuit may be used. (Note: corner squares count as 0.5 squares each, because current lines crowd the inside corner.) The use of mirror-image subunits there may save space if bus widths are set by lithography rather than current capacity, since each bus serves two units. Mask-to-mask registration errors, however, may then introduce differences between the mirrored pairs, since the relative shifts are in opposite directions. Logic circuits may have parallel metal control lines running one way on one level of metal, and the flow of data they control running at right angles on the other, as in Fig. 13c. It is usually easier to make connections between the lower level of metal and poly or diffusion, so that metal level should be selected for lines with many such connections.

Layout rules, fabrication technology, and the underlying device physics resulting from the fabrication process are far more closely linked than with discrete circuits, since all devices are formed together. Diffusion in high temperature steps may cause improperly placed devices to overlap. All share a common substrate which may short them together under some circuit conditions. Even without direct shorts to the substrate, capacitive coupling to it can cause severe problems, particularly with mixed analog-digital circuits. Careful design is necessary to prevent this [65]. The general rule: there are no grounds - just big circuit elements, also holds for chips.

The guide through this maze is a set of design rules from the foundry, which, when carefully followed, should produce circuits that are compatible with the underlying technology. They can range from the seminal but simple set in Mead and Conway [16] to ones having hundreds of parts. The more complex sets, usually coming from foundries rather than textbooks or brokering services such as MOSIS, are tuned to a specific production line, and are designed to allow maximum layout optimization. (In non-ASIC facilities, the layout rules and fabrication steps can be further specialized to the needs of their specific circuits, resulting in a further increase in performance or yield.) Fig. 14 shows several examples from a simple set of layout rules.

In designing devices that use more than one mask, one needs to allow for errors in registration. Fig. 15a-c shows capacitors that will vary; the one in Fig. 15d will not. When possible, process designers try to use self-aligning methods, such as the critical gate to drain-source alignment mentioned earlier. Another example of self-aligned structures can be seen in field shields of the radiation hardened transistors described in reference [66].

In addition to the detailed design rules and the global strategies of power, control and data flow, attention must also be paid to the possibility of process variations causing changes in circuit values due to differences in component placement on the chip. For example, even though the highly corrected

optical systems of mask aligners should keep the dimensions of similar capacitors equal throughout a chip, variations in such things as gas flows and temperatures may produce differences in dielectric thicknesses. The common centroid layout of Fig. 15e will produce four equal capacitances despite a linear variation of dielectric thickness in any direction. Except for the limited regions where vertical and horizontal leads cross, this is also true for the capacity of the connecting links. Components with values to be held to a fixed ratio should be of the same type--same layers, implants, etc.

Interactions with neighboring circuits can also influence layout. For example, long silicon strip detectors will have signals in the sub-millivolt range that must be separated from reset switch noise ten times as large and switching transients (mostly capacitatively coupled from the gate electrode) hundreds to thousands of times larger. The separation can be done using a technique called double correlated sample-and-hold: in one method the amplifier output is placed on two parallel storage capacitors. One is isolated by opening a switch before the event of interest comes in, and the other after. The two capacitors are connected to a differential amplifier and can be read out at a later time. Subtraction of the reset noise and the early switching transients is now automatic. However care must be used in designing the switching network to the storage capacitors. For instance, the line that turns off the second switch must stay away from the isolated and now defenseless first capacitor, yet each must place identical transients on their capacitors. The capacitors and the two sections of the differential amplifier should also be close to each other and of similar shape.

Fig. 16 shows the switching transients and a calibration pulse, alternately positive and negative, and the switching transients that dominate it before (top traces) and after (bottom traces) subtraction. The calibration pulse here is 20 times the size of normal signals, and 400 times the noise level of the circuit. Fig. 17 shows the circuit layout. An off-chip differential amplifier was used. Even better subtraction could be expected for an on-chip one. A second method uses a single capacitor at the amplifier output which is grounded while the circuit is being reset, and then is connected in series with the next stage.

Another example in which large digital signals are cancelled when they couple into analog lines is shown in Fig. 18. An example in which differential techniques can be used to cancel pickup from digital switches, AC ripple on DC voltage busses, temperature variations, and other nonlinearities in time-to-amplitude (TAC) circuits is given in reference [67]. A common gate starts two TAC circuits. One is stopped by the normal "start" signal, the other by the normal "stop", and the outputs are then subtracted.

Layout is normally done using programs that provide easy initial design, copying, and modification of the layout. In addition, they may check for design rule errors, prepare input lists for analysis programs such as SPICE (see next paragraph), and prepare tapes for use by mask makers. One example of such programs is MAGIC [68]. Commercial programs with many additional features are now in common use.

I. Designing Integrated Circuits -- Simulation

The most common circuit simulation program is SPICE (Simulation Program with Integrated Circuit Emphasis), developed at the electrical engineering department at the University of California at Berkeley [69]. To use it, the integrated circuit is broken up into a mesh of discrete components, and an input list consisting of each one with associated node specifications is prepared. Often, wide transistors should be divided into an array of parallel transistors. Traces may need to be replaced with resistors and capacitors (to represent parasitic capacitances which may play a role in cross-coupling as well as rise times). Desired calculations and imposed conditions are then specified.

SPICE has models for different transistor types. It sets up nodal equations of the form $[M] \{V\} = \{I\}$ where $\{V\}$ is the vector of unknown voltages, $\{I\}$ of current sources, and the matrix $[M]$ represents the circuit. For instance, a branch with a resistor and a capacitor in parallel will have terms of the form $1/R + C \, d/dt$ in the appropriate matrix elements. It then inverts the matrix to find the desired voltages. A DC analysis, done first, can be followed by an AC small signal analysis and transient analysis if desired. Even in the DC analysis, repetitive calculations are necessary due to the presence of non-linear elements. For instance, a diode will be approximated by a straight line tangent to the point of greatest curvature. Repeated iterations, each involving a matrix inversion, will move this tangent point.

In a transient analysis, all of this and more must be repeated for each time point. SPICE often has difficulty in converging, and does not give much help in indicating just what is wrong. The information in reference [69] (which is not given in most help files) is vital, but not always sufficient to make it work. Given this state of affairs, it is not surprising that a number of improved and sometimes specialized commercial versions have become available. They include AllSpice, Cadence Spice, CSPICE, DSPICE, HSPICE, I-SPICE, IG-SPICE, IS-SPICE, PSPICE, RAD-SPICE (simulating the operation of circuits subject to ionizing radiation), SmartSpice, SPICE-Plus, SSPICE, and Z-SPICE. There are more than 10 books on SPICE [70]. Some information on SPICE programs is also given in the introduction to references [71] and [72]. Reference [73], while primarily on the semiconductor device models that form an important element of an accurate program, also has a chapter on the overall SPICE program and information on methods for improving convergence.

Large integrated circuit companies often have proprietary versions of SPICE in which some of the above problems are corrected. New generations of circuit simulators, designed to overcome some of SPICE's limitations and problems, and having in some cases, mixed analog-digital capabilities, are now available [72]. A SPICE analysis of a commercial comparator and its use in a constant fraction discriminator is given in reference [74].

If the calculation is any other than an exploratory one, it could be important to run extreme cases as well as the central one. For example, the maximum value of a threshold voltage (which can change

due to many things including implant dose variations), the maximum value of the transistor length, and the minimum value of field oxide thickness could combine, together with other suitably varied parameters, to give the maximum value of rise time for an RC circuit using a field oxide capacitor. Maximum and minimum cases for current drain and rise time are two examples of the sort of grouped parameter variations that may need to be done. As many of these changes, such as diffusion resistivity, may affect many different input elements in a SPICE program, it is often easier to write a Fortran (or similar high level code) shell which takes the inputs, calculates the SPICE input elements, and writes the input file. Such a shell program can also contain software switches that perform the (often) many changes needed to change between DC, AC, and transient analyses. Crucial to all this are the basic parameters. Not too many foundries provide all the necessary information for this type of analysis without prodding.

There are three intermediate forms of application-specific circuits which can be developed with less work than the sort described above, and so should be considered first. In order of increasing circuit density, performance, and (unfortunately) design and fabrication time, they are:

1. field programmable logic arrays
2. gate arrays (digital) and linear arrays (analog)
3. standard cells (both digital and linear).

Field programmable logic array chips contain circuits that are linked by programmable electronics and thus require no design work prior to fabrication. They are described in references [75] and [76], and further papers cited there. Arrays that can be reprogrammed in a millisecond or less are being developed to optimize the *hardware* for the specific problem underway [77].

Gate and linear arrays are fabricated in standard forms but with metal layers left off. These layers are custom designed for each application, and can be fabricated in a few, relatively quick steps. Gate arrays are described in references [19] and [75]. In particular, a flow diagram for gate array design (with 11 boxes instead of the 37 of Fig. 12) is shown on page 246 of reference [19]. While they will not normally have the density of full custom chips, often 75 to 80% of the gates on arrays now available with three levels of metal can be used [78].

Standard cells are pre-designed with standardized dimensions and connection locations, so the assembly of these elements into the design of a completed chip is relatively easy. They cannot be made in advance, so their fabrication time is longer. They are also described in references [19] and [79].

Companies making these chips generally offer extensive documentation and design support. Integrated System Design publishes annual vendor guides listing companies and major products in a number of fields including gate arrays, programmable logic products, standard cells, linear arrays, design programs, design and consulting services, foundries, and even a list of acronyms [80]. Linear arrays may contain individual transistors, resistors and capacitors as well as digital gates. Linear cell libraries may contain such components as op amps, analog to digital converters, voltage references,

current sources, oscillators, filters, switches, and comparators, as well as fixed and programmable resistors and capacitors. Articles on semi-custom bipolar design using arrays of individual components can be found in [81].

J. Effects of Radiation

Radiation damages integrated circuits by two primary mechanisms: (1) damage to the silicon lattice from hadronic interactions or electromagnetic ones with high momentum transfer (usually hundreds of KeV/c and up) and, (2) trapping of ionization from charged particles in insulating structures [28]. In addition, ionization charge deposited rapidly enough can alter the state of a circuit and cause erroneous results even if no permanent damage is done. Dynamic CMOS circuits, whose states are determined by charge stored on insulated nodes, are especially susceptible to such "single event upsets". Such high rates are not common in particle physics experiments except from tracks of heavily ionizing nuclear recoils.

Damaged lattice sites can trap charge carriers, reducing their lifetime. They may also have energy levels in the band gap which can absorb a thermally excited electron from the valence band (in effect, emitting a hole), and then can further excite it to the conduction band. In regions with electric fields, such as the depletion zone between the base and collector of bipolar transistors or in turned-off diodes, this source of alternating holes and electrons produces leakage currents that degrade the performance of the device. Displacement of silicon from lattice sites also causes changes in the effective doping density, changing n type silicon to p type. The exact mechanism is still under study, and may involve a number of things including interactions of dopant atoms and impurities with mobile, radiation - produced vacancies and interstitial silicon [82].

Bipolar chips are usually most susceptible to lattice damage, since their oxide surfaces are mainly used only to passivate the silicon surface, and isolate components from each other. The basic transistor gain is reduced by an increase in the base-to-emitter leakage current resulting from lattice damage in the depleted volume between them. Transistors operating with small volumes and high currents will be least affected.

MOS circuits, on the other hand, are most susceptible to damage from trapped oxide charges. The amount of charge liberated by a single minimum ionizing particle is not large. Much of it in low or zero field regions recombines, and the rest can usually be quickly absorbed by the circuit with no noticeable effect (otherwise silicon strip detectors would be far easier to operate!). However, in oxide layers, particularly those with high fields set up by electrodes such as a gate and the conducting silicon channel under it, the electrons that do not recombine move rapidly (in picoseconds) to the positive electrode (glass is a reasonably good conductor, once the charge is inside past the surface barrier). The positive charge moves more slowly, and dispersively, with the charges appearing at the other surface over a range of times from milliseconds to months. In turned-on NMOS transistors, the charge appears

at the oxide-silicon interface. Fig. 19 shows, schematically, the charge at various times during this process.

Until recently, the mechanism of travel, the details of the interactions at the oxide-silicon surface, and the observed strong sensitivity of radiation hardness to trace amounts of hydrogen in the oxide were not well understood. They still aren't, but recent tests, in which the hydrogen is replaced with deuterium by adding a deuterium anneal (or a hydrogen one for control) to the fabrication steps, show a clear increase, by a factor of about 3, in the times for the charge transfer with deuterium [83]. The times are inversely proportional to the square of the oxide field, so a charged carrier -- normally H^+ -- is involved, and the larger mass of deuterium certainly would be expected to reduce its velocity and increase the collection times [83]. The reason for the specific value, which can range from 2.5 to 4.5, depending on field and fabrication conditions, is not understood.

In radiation hardened circuits, a few percent at most of the charge reaching the vicinity of the oxide-silicon interface will be trapped. The rest will be neutralized by electrons from the underlying silicon. In most non-hardened circuits, however, half or more will be trapped in the oxide at or near the interface. Etch-off and photon assisted tunneling experiments indicate most of the charge is within about 5 nm of the interface [84]. Turned-on PMOS transistor fields will carry charge away from the interface, but even with such fields, some charge from ionization near the interface will diffuse there. Much more will of course, for turned-off PMOS transistors.

The accumulated charge will attract electrons or repel holes in the underlying silicon, and change transistor threshold voltages. Charge trapped some distance from the interface, called oxide-trapped charge, does nothing more than produce a simple voltage shift in transistor current-voltage characteristics. Charge trapped closer to or at the interface can interact with electrons or holes in the silicon and change its charge state, depending on the voltage. These changes also have varying characteristic times, being slowest for locations furthest from the silicon where tunneling through the oxide must occur. The distant ones are sometimes called "oxide traps" or "border traps" and the closer ones "interface traps" or "interface states" [85]. The mechanisms are not yet fully understood, but probably involve trivalent silicon atoms with a dangling bond that, at the interface, is charged negatively below midgap and positively above. Holes, as well as ions, are probably involved. The change tends to reinforce the effect of the oxide-trapped charge for PMOS transistors and cancel it for NMOS ones. The relative rates of interface trap formation and oxide-trapped charge accumulation are not constant, so threshold changes may actually reverse for NMOS transistors with increasing radiation. Fig. 20a shows the change in transistor characteristics around threshold. Fig. 20b shows the changes in n and p channel thresholds as a function of dose.

In addition to threshold changes, the interface states usually change the shape of the current vs gate voltage curves as the changed voltage alters their charge states. They also cause an increase in noise, probably by trapping and releasing charge carriers in the channel, and reduce the mobility of the

charge carriers. Trapping and release of charge carriers in the bulk may also cause noise, particularly in buried channel devices (transistors in which the current flow is kept at some distance below the gate oxide by a potential minimum created by the charge on ionized dopant ions).

Only a small amount of work has been done in studying increases in noise, since they are not crucial for most digital circuits. Figs. 21a and 21b show the increase in noise of a MOS and a JFET transistor [86]. These results are for discrete transistors; the scarcity of data for integrated circuits is just now starting to end with the first research results for LHC and the formerly planned SSC detectors [87].

JFETS, having no gate oxide, do not have such threshold shifts. With their drain-source currents carried in the bulk, rather than along or near an oxide interface, they have less charge trapping and detrapping, and hence less noise. They will have increased leakage from radiation induced generation centers, and some threshold shifts from changes in depletion voltages, but they are generally much harder than MOS transistors made without special hardening steps.

Threshold shifts depend on the fabrication technology as well as the total dose and dose rate. They are proportional to the gate oxide thickness raised to a power that can range from about 1.4 to about 3. One power comes from the amount of ionization generated and one from the separation distance between the sheet of trapped holes (near the substrate) and the gate. The remaining factor comes from the trapping and surface state formation probability of the oxide near the silicon interface which depends, in a complicated way, on the thickness. (For example, thick gates will be stiffer, and will have more differential expansion and contraction during high temperature steps. This will cause more relative motion against the underlying silicon, with larger changes to the structure of the interface region). Very thin oxides ($<10\text{nm}$) have less charge trapping than would be expected from the power law; trapped charge apparently can tunnel out. The threshold shift for thick oxides ($>100\text{nm}$) continues to increase with thickness, but at a lower rate than would be predicted from the power law. Nitridation of the gate oxide by annealing in 100% NH_3 , followed by an anneal in dry oxygen, is known to further increase the hardness of thin gates [88].

An additional effect of trapped hole charge is to lower the threshold of the (parasitic) field transistors. These are structures formed by conducting traces (the "gate") on field oxide over the substrate (the "channel"). They should be designed so they never operate as turned-on transistors. The faster-than-linear thickness dependence of threshold voltage change means, however, that these thick, potential transistors can be an early source of failure if they should turn on. Also, the positive charge, added to the positive fixed oxide charge already there, attracts a layer of induced negative charge in the silicon below. P type silicon can be changed to n type from oxide bulk damage, even in the absence of overlying conductors. This n type silicon can short together n^+ sources and drains.

Several methods can be used to prevent this, among them:

1. The channel stop implant doping can be increased, making it more difficult to invert the field. This method will be limited, eventually by breakdown problems at the field-drain/source junction

due to the larger electric field created by the larger charge concentration. This should work for levels up to 10-100 Krads if deposited over an extended time [89].

2. The bias of the substrate relative to the other transistor voltages can be changed, increasing the size of the depletion zone, thus terminating field lines on fixed atoms rather than mobile charge carriers. Leakage in NMOS (PMOS) circuits can be prevented or at least retarded with negative (positive) substrate bias. Of course, the characteristics of all the other transistors will change, and that, or breakdown may limit the use of this technique.
3. P⁺ guard bands, heavily implanted p⁺ regions in the p substrate around n channel transistors, can be used to prevent inversion of the substrate to n type. In transistors that should be turned off, they can also prevent leakage from the source to the drain by paths in the substrate that circle around either end of the gate, provided that the (turned-off) gate and gate oxide are carried all the way over to the band. In addition, they can be used to block current flows that would cause latchup. However, they must be spaced away from the transistors they protect, and so take significant extra area. Fig. 22 shows a simple layout for a guard band [90].
4. The use of silicon on insulator (SOI) technology (see section III D.), in which each transistor sits, isolated, on an insulating substrate, is free from this problem, since there are no silicon paths connecting them. However, charging effects in the oxide may affect transistor operation.
5. A conducting layer can be placed inside the field oxide and tied to the substrate, shielding it from the conductors on the top of the field. The faces of this conductor must be insulated at the edges of the field, and it takes extra mask steps, but no extra area [66]. Fig. 23 shows a cross section through such a field shield. (Compare with Fig. 3a.) This structure prevents field inversion by providing an effectively thin field oxide (the oxide thickness between the substrate and the shield is 25 nm rather than several hundred). Also, since the shield is tied to the substrate there is no field to separate radiation induced holes and electrons, or to turn on the field transistor. For good ohmic contact, the shield poly has the same doping as the substrate. Circuits made with this technology have operated at 2.5 - 3 GHz after a total dose of 50 - 100 Mrad.
6. Hard field oxides can be fabricated which have only a small hole trapping probability. As in the field shield method, the layout density does not suffer, but the process is more complex. A number of references are given on p.387 of Ma & Dressendorfer [28].
7. A field oxide will be hardened if the threshold shift from trapped holes cancels that from interface states. Special fabrication steps are needed to produce such oxides.

Many other fabrication steps, in various devious ways, can affect radiation hardness. The use of hydrogen, which in soft processes passivates dangling bonds at the interface and produces quieter transistors, can, after the gate deposition, increase interface states and reduce hardness. The reason, as mentioned earlier, is still not fully understood in detail. Processes not designed to change the hydrogen content can, nevertheless, do so. For example, the formation of a polyimide scratch mask, often used to

produce a lower pinhole density than found in glass, involves hydrogen-containing compounds and temperatures of several hundred degrees C, and increases the hydrogen content of lower layers. Implants through gates to adjust thresholds, and the use of equipment that generate x-rays such as plasma etchers and electron beam metal deposition machines, will also reduce hardness. Keeping post-gate temperature excursions low helps, since differential motion of the gate oxide and silicon substrate are reduced. Low melting-point glasses containing boron and phosphorus are useful here. Cleaning wafers is especially important before high temperature oxidation steps. It turns out hydrofluoric acid in the final rinse produces harder circuits than the more commonly used de-ionized water.

Circuit design can also increase hardness [28]. For instance, in amplifiers, negative feedback can be used to reduce the effects of threshold and mobility changes, just as it is often used to reduce the effects of temperature and process variations. Static memories, which do not depend on low leakage, are usually better than dynamic ones which do. Synchronous circuits, in which signals arrive a bit early, then are gated through with a clock pulse, are better than asynchronous ones which may be affected by timing changes due to threshold changes. Techniques, such as that described in the paragraph on transmission gates, are available in CMOS to cancel switching transients. This produces a larger valid to invalid signal ratio, and can be directly translated into decreased sensitivity to radiation-induced parameter shifts. As mentioned in point 2 above, radiation-induced threshold shifts can be compensated by shifting the bias of the substrate relative to the transistor voltages. In general, conservative circuit designs, with large safety margins, will tend to be harder than ones that barely make it in the unirradiated state.

With time, and depending on temperature, some damage will anneal out. This is usually, but not always good. For example, if threshold shifts from trapped holes have cancelled those from interface states, and they anneal at different rates, the shifts can increase with time.

Completed chips are often tested with low to medium (1-20 MeV) energy protons or neutrons, when lattice damage is being studied, and ~ 1 MeV electrons, often Compton electrons made from the gammas of a ^{60}Co source when ionization effects are being studied. These electrons are not energetic enough to do too much damage to the lattice, and yet are close enough to minimum ionizing to mimic high energy particles. When lattice damage does matter, as it may for silicon detectors, even low energy protons may not be adequate because recoils from higher energy particles can produce dense clusters of damage not easily reproduced by several-MeV particles. Since the gammas will, for the most part, go through the silicon without interacting, material should be placed between the source and the silicon. For easily reproducible results, its thickness should be considerably greater than the range of the Compton electrons, but less than the interaction length of the gammas. The dose from backscattered electrons from walls and plugs must also be measured.

Increasingly, ~ 10 KeV x-rays are being used in place of ^{60}Co , since large dose rates can be delivered with little shielding, permitting rapid testing for process control. The method also has its dangers,

particularly for high energy physics. Because of the difference in Z between the oxide and the silicon, the x-rays are absorbed with a spatial distribution which will depend on such things as the poly, field oxide and gate oxide thicknesses, and will differ from the uniform distribution of the cobalt gamma or penetrating particle ionization. In addition, the ionization produced by the non-relativistic electrons ejected by x-rays is far denser than that from minimum ionizing particles, and has a higher recombination rate (and so a lower fraction of holes arriving at the oxide-silicon interface). Worst of all, it has a damage rate ratio to that of minimum ionizing particles that depends on the gate oxide electric field and thus on the electrical state, during irradiation, of every transistor on the chip. Several references are given in [91].

Another sometimes deceptive test is the one run by many high energy physicists using the proper beams, but on unhardened, sample chips from a commercial foundry [92]. Unless the hardness, whatever its value, is tested and controlled by the foundry (and it usually isn't), a change in any of the hundreds of steps, meant perhaps to improve yield, noise, threshold uniformity, etc., will have unknown effects on the hardness. It is necessary to test hardened chips under actual electrical conditions and with radiation selected for type, energy, and rate to produce damage similar to that expected in real service. Even then, some units may fail the radiation tests, while passing all others, and you will probably never know why. Even the experts at Sandia have chips like that.

A general survey by J. Raymond and E. Petersen [93] of hardness in different technologies available in 1987 is given in Fig. 24 which combines data from two figures of their paper. An individual chip would have a certain failure level for ionizing radiation (horizontal axis) and another for neutrons (vertical axis, with the damage being caused by nuclear collisions). It would appear then, at one specific point on the plot. The rectangles enclose a representative set of points for each listed technology.

Synergistic effects and combined failure modes were not considered, i.e. hardness along each axis was measured separately. The decrease in gate oxide thickness that accompanies the industry-wide decrease in minimum feature sizes (see the section on scaling under "V. Technology Trends") makes hard CMOS devices easier to make. The technology described in reference [66] increases the hardness limits of bulk CMOS for ionizing radiation yet another order of magnitude from that shown. Some caution is needed, particularly, for MOS electronics, because of the dependence on the rate of radiation dose, density of ionization, bias on the gates during irradiation, and tolerance of the circuit design, in addition to the total dose listed in the figure.

J.I. TTL and J.I. ECL are junction-isolated TTL and ECL. Oxide isolated bipolar circuits are more sensitive to ionizing radiation than junction isolated ones due to oxide - charging failure modes similar to those of MOS electronics. CMOS/SOS (silicon on sapphire) is somewhat less resistant to ionizing radiation than bulk CMOS with hardened field oxide because of charging effects on the underlying sapphire insulating layer. Of course, CMOS/SOS has other advantages, for instance, lower capacity to

the substrate. CMOS is somewhat harder than NMOS because of greater circuit immunity -- for instance by being less sensitive to radiation induced threshold shifts -- rather than by having smaller threshold shifts.

Protons (and other charged hadrons) cause damage both by nuclear collisions and by ionization, and so appear as diagonal lines. For instance, 90 MeV protons dumping 1 Mrad of ionization cause as much nuclear collision damage as 10^{13} neutrons per sq. cm in the MeV energy range. Protons kill low power TTL chips in the lower right corner of the box by collisions, and in the upper left corner by ionization damage. The neutron/gamma line is for a mix found at a certain reactor. It is an example of how one might combine effects; similar lines might be drawn for hadron showers with a mix of pi zero initiated electromagnetic showers.

Several thousand people had been working in this field in the U.S. for over 20 years. Lists of U.S. foundries making radiation hardened circuits were given in "VLSI Systems Design" [64] and "Defense Science" [94], but many have recently left the field. A substantial amount of the work in the field is classified, and while some of the products might be available on a limited basis, knowledge of the process steps may not. This can have an impact when it is desired to fabricate specialized devices such as monolithic pixel detectors. There are similar projects underway in Europe. Device testing is under way by groups preparing for experiments at the planned Large Hadron Collider at CERN. Reports by those groups are perhaps the best source of up-to-date information. [95]

In summary, circuits and fabrication technologies resistant to radiation damage to 10 Mrads are commercially available now. Ones resistant to 100 Mrads have been developed and may be available. A price has to be paid for them: some combination of extra area and reduced performance, of additional and difficult process steps, or, with thin gate oxides, of a more fragile circuit.

K. Testing

Tests for developing new fabrication technologies use different instruments and techniques from those used with settled technologies where only new circuit designs must be tested. The former include Auger, Rutherford backscattering, secondary ion, and x-ray emission spectroscopy; interference contrast, transmission electron, and scanning probe microscopy; laser desorption/ionization, total reflection x-ray fluorescence, neutron activation analysis, and the measurement of lifetimes and diffusion lengths. There are companies that specialize in making such tests. References [11]-[14] cover these techniques, with chapter 12 of "VLSI Technology" [12] being a good one to start with. Developments in high energy physics, on the other hand, will be mainly concerned with the development of new circuits using existing technology, so that is covered here.

"Troubleshooting Analog Circuits" [96] can be read with profit by anyone testing integrated circuits, analog and digital (which, after all, can have analog problems). Try to ignore the apparatus pictured on the cover.

If the fabrication process is being developed, there are likely to be process splits, where the wafers are divided into several groups in which the step being tested is varied. Even if the process is settled, but the circuits are being developed, some wafers might be held back from further processing at each major mask step. This will allow a faster recovery if a mask error is found as a result of circuit tests.

K 1. Designing for Testability

Testing starts with the initial design, which should be done with the needs of a test program in mind ([53], pp1653-5). Circuits made of repeating units of basic cells will usually be easier to test as well as to expand, and often allow for denser layout. Calibration inputs are generally useful in analog circuits. Effects of noise on incoming calibration lines can be reduced by using large calibration pulses which are heavily attenuated immediately before or after entering the chip. Multiple lines will allow one channel to be pulsed while adjacent ones are off, allowing measurement of cross coupling. Pulsing many channels simultaneously, even though they are not adjacent, can induce cross coupling through common power lines. This effect may be worse than normal if the usual number of simultaneous signals is not so large.

Pulse inputs capable of reproducing what the chip will experience in use are also desirable for digital circuits. Designing a comprehensive testing program for a complex digital circuit is a major undertaking [97]. The circuitry not only must allow the proper combinations of pulses to be introduced to all important nodes, internal as well as external, but the results must be observable, and not be lost deep in the innards of the chip. Often this task is simplified by subdividing the chip into subunits, with some space devoted to getting the test pulses in and the (usually intermediate) results out. Additional space is sometimes used for self-test circuits.

If at all possible, particularly with analog circuits, key points in the interior of the circuit should be made accessible for measurement of pulse shapes. Signals on metal lines can be easily measured if a window in the scratch mask is placed over some part of the line, preferably over a contact to a lower layer. Signals on poly and diffusion layers with no direct contact to metal can be reached if a contact to a minimum geometry metal patch is inserted, again with a window over it. Only minimum geometry is needed, since it is easy to place a fine tungsten probe in the contact depression where the metal dips down into the hole in the insulating oxide, and the side walls keep the tip from sliding away. Extending a metal line slightly beyond its normal end to a test point, if space is available, will allow probing with minimal danger to the metal line.

K 2. Apparatus for Testing

The apparatus to hold and move the wafer or chip under a microscope, provide it with power, and hold the micropositioners that control the probes is called a probe station. Fig. 25 shows such a unit. The ability to probe directly on the wafer as well as on diced chips is important, since significant time can be saved if one does not have to wait for packaging. The unit in Fig. 25 is a manual prober. Automatic ones which locate each chip, contact it, and test it with a preset program are also normally

available in integrated circuit laboratories. They are limited to contacting large pads, typically 50 -- 100 microns in size, but are crucial when many structures on the wafer must be systematically measured. Manual probers allow more flexibility. Just as with automatic probers, they will usually provide power and input signals to the chip through its pads using a set of custom probes mounted on a probe card. But in addition they may use micropositioners with special, fine probes having tip radii of a micron or less. Probes with input impedances of 0.1 pF in parallel with 1 to 10 megohms and rise times of 0.1 to 1 ns are easy to use. Stray capacity of the probe input is reduced by a series resistor at the tip and a shield that is driven to the same voltage as the input line. For high impedance circuits, units with a MOS transistor directly in the probe tip have a capacity of 0.02 pF (but no input protection!), about 10^{14} ohms, and similar rise times [98].

The microscopes are unlike any you may remember from college biology. They have binocular, zoom eyepieces. The objectives have large diameters with clear-space working distances of one cm. or more, even at magnifications of 2000. Objects as small as a micron or two look crisp and undistorted.

Sometimes short circuits can be repaired by cutting the offending trace with an ultrasonic cutter or a laser. Isolating parts of the circuit may also help in diagnosis. Laser systems are available that can cut through a selected region of insulation without harming an underlying metal layer; can cut through that one and further layers of insulation to expose still lower metal levels, and can deposit traces, pads for test probes, and bonding pads of low resistivity metals such as gold and silver [99]. Resolution is about 1 micron. Machines with ion beams that focus to spots of 0.1 micron diameter or less can also do this by sputtering. They form their traces by directing a jet of gas such as $W(CO)_6$ onto the wafer. The ion beam then converts adsorbed molecules to tungsten metal. Fig. 26 shows a repair where one trace was cut and two others were connected. Vertical cuts through several layers can be used for diagnostic purposes. A nice touch: the secondary electrons are used to produce on-line, real-time magnified images [100]. Electron microscopes and mass spectrometers are found in some recent units [101]. Focussed ion beam machines can also be used to repair masks. Both lasers and ion-beam machines are expensive, the later several times more than the former. However, it is possible to have repairs made by some of the companies that make them, or by service companies that own them.

The voltage contrast scanning electron microscope provides another way of examining the entire chip: a high energy beam is scanned across the chip. The number of low energy electrons knocked out and collected will depend on the voltage of the circuit elements in the immediate vicinity of the impact point of the beam.

K 3. Test Structures

The first thing tested on a wafer will be the fabrication process. Many routine tests are performed as wafers are fabricated, even in well developed processes. These may include checking for particulate matter, verifying the thickness of newly applied layers using interference of polarized light, and etching off all layers and treating the surface to reveal defects in the silicon substrate that may have

been introduced by the processing. Special wafers will be included in the run for such sacrificial tests. In addition, every wafer will contain a set of test devices called a process control monitor. A line with a well established process may only have two small areas on the wafer devoted to them, while an experimental one may have them near every chip, and over much of the entire wafer surface. The monitors are probed by the foundry itself, using an automatic prober. Normally the results of these tests are provided to the user. Typical devices in such a block may include:

1. Four-contact structures to measure surface resistivity (ohms per square) of various layers. Several widths in both x and y directions may be used. (See Fig. 27a. The outer electrodes are for current in and out; the inner pair, placed where the current flow lines are parallel, are used to measure the corresponding voltage.)
2. Contact resistance structures. These are designed so measurements can be made to separate the resistances of the contacts and the structures in series with them. See Fig. 27b.
3. All the types of transistors available in the technology in varying sizes so second-order dimensional effects can be measured. In addition, field transistors should be included, even though they are not normally used, to verify they have adequately high thresholds.
4. Capacitors using all possible overlapping combinations of layers. Both breakdown voltages and capacitance will be measured. Even if some layer pairs are not deliberately used as circuit elements, they will be needed for evaluation of the parasitic capacitors in overlap regions.
5. Circuit elements such as inverters, pass gates, and ring oscillators.

A ring oscillator is a series of linked inverters in a ring. With an even number, a fluctuation on any output will come back to the same point in the same sense, inducing saturation, alternately positive and negative around the ring. An odd number can be unstable and oscillate. For example, take $V=0$ midway, and the output of stage $i+1$ to be,

$$V_{i+1}(t) = -V_i(t - \delta) \quad (4)$$

where t is the time and δ is the delay in a single inverter. Applying the relationship twice around a ring with an odd number, n , of stages,

$$V_0(t) \equiv V_{2n}(t) = (-1)^{2n} V_0(t - 2n\delta) = V_0(t - 2n\delta) \quad (5)$$

Since this holds for all times, the cycle time must be $2n\delta$. Measuring this time is a convenient way of measuring fast electronics with not-so-fast equipment, since you can increase the number of linked inverters in the circle, and then only have to measure the relatively long time, $2n\delta$. Fig. 27c shows a photograph of a ring oscillator.

In addition to these elements, it is a good idea to have a special section on the wafer in which each major unit of the chip is broken out and given a set of pads that can provide it with all the power, input

leads and output lines necessary for full functioning of that piece. This will also enable it to be bonded up for radiation hardness tests under power if needed. These separate units may be easier to test. They may be the *only* things you test if a fatal flaw in one part prevents the rest of the chip from working.

K 4. Stress Testing

When working chips are produced, the next step is to stop them from working. Integrated circuits are normally produced in large quantities and are expected to last for a number of years. However, even when process and circuit designs are settled, and the run was successful, individual failures will be found. Often there will be a high early failure rate, usually from random manufacturing defects. All chips should be checked to detect and remove those early failures. Normally the remaining population will have a reasonably long life, and the failure rate will drop.

The expected life of the remaining chips may be monitored using accelerated tests on a fraction of them. The accelerated tests are called "stress tests". The nature of these tests and the acceleration factors achievable will depend on the mechanisms being accelerated and on the expected causes of failure. These have been deduced from long term test results of *non*-accelerated tests which should not only give the mean life, but also the lifetime distribution for the important failure mechanisms. Examples are given in reference [12]. Often the distribution can be well fitted by a log-normal distribution

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp - \frac{1}{2} \left[\frac{\ln \frac{t}{t_m}}{\sigma} \right]^2 \quad (6)$$

where t_m is the median time to failure, and σ is a parameter specifying the width of the distribution. A large community of engineers and scientists works in this field. Much of the work is published in the annual proceedings of the International Reliability Physics Symposium [102].

Mechanisms that have been accelerated include:

1. Temperature -- The acceleration factor is often the ratio of reaction rates derived from the Arrhenius equation:

$$\frac{R_T}{R_{T_0}} = \exp E_{act} \left(\frac{1}{kT_0} - \frac{1}{kT} \right) \quad (7)$$

The activation energy, E_{act} , is usually in the range of $\approx 1/4$ to 4 volts. Depending on the damage mechanism, increasing temperature can either increase or decrease (if annealing of the damage plays a role) the damage rate. In the latter case, E_{act} will be negative.

2. Temperature cycling -- The major purpose of temperature cycling is to increase the rate of mechanical damage due to repeated stress changes from differential thermal expansion. For most electronics, which is normally cycled slowly compared to the time to reach thermal equilibrium,

or cycled so rapidly that a constant, though non-equilibrium, temperature is maintained, it is easy to reach a substantial acceleration factor by using a cycle time that is only slightly longer than the thermal equilibrium time.

3. Humidity -- Water vapor can permeate many kinds of packaging material, transporting contaminants from the surface, possibly leaching impurities from the epoxy, and directly attacking the metal at pad regions not covered by the scratch mask as well as under any cracks that may be present in the mask. Phosphorus is often used as a component in intermediate oxide layers to lower the oxide melting point and improve its flow characteristics to make a smooth substrate for an overlying metal layer. Also, phosphorus will trap sodium ions during high temperature steps, improving device stability (see also sections III B 8 and III L 3) [103]. However, if it is leached out by water, it can dramatically degrade the circuit. Normally this need not be of concern in physics experiments, since the ASIC chips are among the crown jewels of the experiment, and high quality hermetic seals can easily be afforded. However, one important exception occurs at the front end, where multiple scattering considerations may dictate the thinnest possible packaging. Accelerated humidity tests can often be combined with high temperature to further increase the acceleration.
4. Irradiation -- This has already been discussed in section J. The main limitations to the acceleration rate are the intensity of the accelerators and sources (normally not a significant problem at labs specializing in radiation damage), and the need to operate the chip under realistic electrical conditions to produce the sort of damage that will actually be seen. Thus, the chip should not be overwhelmed to the point that it cannot operate normally. Except for integrated detector chips, this also should not be a serious problem. It is also important to keep the effects of temperature in mind. When the radiation comes gradually, a certain amount of temperature-induced annealing goes on. If a detector chip is tested at room temperature but operated at liquid argon temperature, for example, you may find significant errors.
5. Voltage acceleration -- Damage due to dielectric breakdown, radiation damage, hot electron effects (where high energy electrons in the channel damage the gate oxide), and corrosion can be accelerated by increasing the voltage. Dielectric breakdown tends to occur rapidly once some threshold is exceeded, so rather than accelerating the process, one needs only to briefly exceed the highest voltage the chip is expected to see. The other mechanisms tend to increase as

$$R_V(T, V)/R_{V_0} = \frac{R_T}{R_{T_0}} V^{\beta(T)} \quad (8)$$

where, $R_V(T, V)$, the voltage acceleration factor relative to R_{V_0} , is the product

of the temperature acceleration factor $\frac{R_T}{R_{T_0}}$ and the voltage V raised to the $\beta(T)$ power. β ranges from 1 to 4.5 (reference [12], p628). Since most circuits permit only modest changes in operating voltages for proper operation, the amount of acceleration is usually not large. Also, the changes interior to the chip from an externally applied voltage increase may not produce a failure pattern comparable to that in actual use. This is an excellent place to make use of special test structures and pieces of the circuit that have been copied and supplied with their own contacts.

6. The main current dependent failure mode being accelerated is electromigration, with an acceleration factor R_i

$$\frac{R_i(T, I)}{R_{i_0}} = \frac{R_T}{R_{T_0}} I^{\beta(T)} \quad (9)$$

which again is proportional to the temperature acceleration factor and a power β of the current, I , where β ranges between about 1 and 4. As with voltage, it is difficult to vary I and still keep the circuit working, so again, special test structures can come to the rescue. Table 3 (from reference [104]) is a summary of much of this material.

Table 3
Failure Acceleration Factors for Silicon Integrated Circuits

Device association	Failure mechanism	Relevant factors	Accelerating factors	Acceleration (E_{act} = apparent activation energy for temp.)
Silicon oxide and Silicon-Silicon oxide interface	Surface charge accumulation	Mobile ions, V, T	T	$E_{act} = 1.0-1.05$ V (depends upon ion density)
	Dielectric breakdown	E, T	E, T	$E_{act} = 0.2 - 1.0$ eV $E^g, g(T) = 1.0 - 4.4$
Metallization	Electromigration	T, J, A, gradients of T and J, grain size	T, J	$E_{act} = 0.5 - 1.2$ eV $J^g, g = 1 - 4$
	Corrosion (chemical, galvanic, electrolytic)	Contamination, H, V, T	H, V, T	Strong H effect $E_{act} = 0.3 - 1.1$ eV (for electrolysis) V may have thresholds
	Contact degradation	T, metals, impurities	Varied	
Bonds and other mechanical interfaces	Intermetallic growth	T, impurities, bond strength	T	Al - Au: $E_{act} = 1.0 - 1.05$ eV
	Fatigue	Bond strength, temperature cycling	Temp. extremes in cycling	
Hermeticity	Seal leaks	Pressure differential, atmosphere	Pressure	

Note: V -- voltage; T -- temperature; E -- electric field; J -- current density; A -- area; H -- humidity

L. Assembly and Packaging

In the spring of 1956, Fred Kulicke and Albert Soffa, owners of a small engineering and design shop, trying to get some business from Western Electric in Allentown, were asked by the engineers there if they could make a better model of a device used to attach wires to a tiny object. They were shown a tiny soldering iron and scissors mounted on a set of micrometer screws, and the object: the world's first transistor planned for mass production. Experienced in making unusual machines, and recognizing it doesn't take much force to generate a lot of pressure on a small enough object, Kulicke and Soffa made a device that used pressure and a modest amount of heat to *weld* the wires in place. The ensuing development of the first wire bonder started the microelectronics assembly industry, now populated by

hundreds of companies. They also included a few other devices to speed production, such as low backlash, motion-reducing pantographs and a built-in microscope. Modern wire bonders, with automated controls, machine vision, and memories, can learn complex patterns of fiducials, bonding pads, and wiring procedures given only an initial manual pass. They can load and position parts automatically, making up to 20 welds per second on pads less than 80 microns in size.

The large size of the assembly and packaging industry will come as no surprise to anyone who has shepherded a freshly minted chip through to the beginning of its useful life. For example, two people, working part time, completed the design of the Microplex chip in less than two years. The Mark II silicon strip vertex detector that used it was ready for installation after twelve additional people had worked four more years. Although an extreme case, due to the tight spatial constraints and mechanical precision required, you should not be too surprised if the last little chore of housing and feeding your chip takes considerably more effort than making it.

Turning a tested wafer into a mounted, working chip takes four main steps: dicing the wafer, bonding the chips to the circuit board, attaching leads, and encapsulating it so it can be cooled and protected from air, moisture, and mechanical damage. Chapter 13 of reference [12] provides a good introduction to the field. It also has over 60 further references. Academic journals [105] and trade publications [30], [33] also cover this field. Reference [23] is a basic reference for packaging and interconnects. In addition to these topics, it covers the impact of packaging and interconnects on circuit design. Two excellent articles on packaging for main frame computers also have introductory sections containing much useful material [106]. A recent development provides protection and contacts in a package that is barely larger than the chip itself [107].

L 1. Dicing and Die Bonding

Dicing is usually done by mounting the wafer on a film of mylar with a slightly sticky top layer [108], and sawing through it and part of the film using a circular saw blade with embedded diamonds. Modern saws with optical feedback can make 15 cm long cuts straight to ± 3 microns in less than half a second, checking the quality of the cut as they go.

Selecting, placing and bonding the dice on a chip carrier, pin grid array, or printed circuit board in commercial products is normally done by an automatic die attach machine which has an arm that picks up chips that have passed on-wafer tests, and places them. Chips intended for use at high temperatures or with high substrate currents, may be soldered using thin (less than 50 microns) eutectically soldered gold sheets. Silver filled epoxy or polyimide makes a more flexible, less expensive connection where the chips can be heated and cured simultaneously. Recent improvements in conductive epoxies have resulted in over 95% of new devices being made for use with epoxies rather than solder. Fig. 28a shows a semi-automatic bonder capable of dispensing adhesive in a precise pattern, then picking up, placing and bonding several die per minute with an accuracy of ± 25 microns.

Fig. 28b shows one being picked up from its mylar backing, and Fig. 28c shows it being placed in a pin grid array.

L 2. Attaching Leads

Wire bonders connect input and output pads on chips to the corresponding conductors on circuit boards by welding fine wires (17 to 500 microns in diameter for aluminum, and less than 50 microns for gold), using pressure together with additional heating and/or ultrasonic energy. The flow of the melted material disrupts the thin layer of oxide and other surface contaminants. With a precisely controlled vertical force and some but not too much horizontal motion, a strong weld is produced. Information about wire bonding can be found in [109].

While wire bonding is the oldest method, it is still the most common because of its low initial cost, flexibility, and ease of changing from one type of job to another. Two other important methods are tape automated bonding (TAB) [110] and flip chip technology [111]. In the former, the bonding wire is replaced with a tape with metal conductors in a pattern specialized for each application. TAB leads, because of their rectangular cross section, and, if desired, close spacing, can have lower impedance than wires. With their tape backing, they are also more robust than wire bonded connections. Flip chip technology bonds chips, circuit side down, to a substrate or to another chip. TAB and flip chip bonding normally require preparatory steps to form metal bumps, which can have diameters as small as 7 microns, on the pads. Besides the direct costs that requires, there may be some extra loss of yield, although the heavy metal pads will then suffer less damage from probe testing. Ultrasonic bonders can form welds with local material temperatures less than 100 °C above room temperature. With other methods, however, the temperature at the weld must be higher than that caused by any following soldering steps on the board. Current minimum lead-to-lead pitch limits are 70 microns for ball bonds, 55 microns for wedge bonds, 70 microns for TAB and about 30 microns, in a two-dimensional array, for indium bumps (which also bond at room temperature).

Fig. 29a shows a manual wire bonder, intended for lab use, and Fig. 29b, a close-up of the work piece. Figs. 30 show the sequence of steps to make a ball/wedge bond pair. Note the last step in making a connection involves pulling on the wire to break it away from the bond. This also tests the bond, which, if weak, will fail at this step. Tests of first bond strength should also be made. Special devices which test bonds, both destructively and non-destructively, are available. In a wedge bonder, a wedge is made for the first as well as the last bond. A special tool holds the wire so it projects out sideways under a tool which will compress it in the first step.

Ball bonders have more variables to misadjust. They require the use of gold wire for ball formation, which, in turn, requires a holder heated to about 200° C, whereas aluminum wedge bonders work at room temperature. However, ball bonders, starting from a symmetric ball, are faster (8 - 10 wires per second) than wedge bonders (5 - 7 wires per second), since there is no need to rotate either the head or workpiece parallel to the bond direction. Gold is more resistant to weathering than aluminum, but not

as stiff. There are potential problems at the gold-aluminum interface: an intermetallic compound, Au_2Al , which is brittle and poorly conducting, can form. (The Au_2Al can form at the same time as other compounds including the purple, strongly bonding, highly conductive AuAl_2 , which, under the name "the purple plague" once got the blame.) Vacancies caused by differential diffusion and differential expansion of the various intermetallic compounds, can form, merge, and cause the bond to fail.

Our experience has been, for the low production rates commonly needed in high energy physics, aluminum wedge bonding is better. This was certainly the case with the 20,000 wire bonds used in the silicon microstrip vertex detector at the Mark II.

As the number of elements, N , on a chip rise, the number of input and output pads needed, P , also rises and is, roughly [23], [112]:

$$P = aN^b \quad (10)$$

The value of "a" depends on the type of circuit and "b" lies between 0.5 and 0.7. Typical values of "a" can range from 1 to 10. Reference [112] also gives probability distributions for P . While bond pad sizes have been getting somewhat smaller, they haven't kept up with the decrease in transistor sizes. So if transistor sizes shrink by a factor, s , and pads by a larger fraction, S , ($s < S < 1$), a chip of constant size formerly with N elements can now hold about N' , where

$$N' = \frac{N}{s^2} \quad (11)$$

while the number of pads needed, P' is

$$P' = a(N')^b = a\left(\frac{N}{s^2}\right)^b = \frac{P}{s^{2b}} \quad (12)$$

But the number that can be placed along the edges has only increased to

$$P'' = \frac{P}{S} < \frac{P}{s} < \frac{P}{s^{2b}} = P' \quad (13)$$

where the inequalities hold because $s < S$ and $2b > 1$.

The only way to keep the pads on the edge is to make the chip smaller, which is exactly the opposite direction one wants to go. TAB, flip chip, and ball grid array techniques, which can easily bond to pads which are dispersed throughout the area of the chip, are useful here.

The leads for tape automated bonding are normally gold plated copper on a tape, usually of polyimide [113]. The pads on the chip are covered with gold or indium bumps. Under thermocompression, similar to that in wire bonding, a weld is formed between the chip and tape. The tape can carry a dense array of leads without the danger of shorting that might be present if wire

bonding was attempted. (Special cases do exist of high density wire bond arrays, but they are more difficult) [2]. Some TAB methods are of interest primarily to industries making inexpensive, high volume products where chips on one tape, and leads on another are merged and all bonds are simultaneously welded.

In one method bumps are formed on the chip by sputtering a 0.1 micron thick metal layer over the entire chip, and covering that with a layer of photoresist. (Because of their high impact velocity, the sputtered atoms will stick well to all parts of the chip, both insulating and conducting, and so form a continuous conducting layer under the resist.) A lithographic step removes the resist at bump locations, and gold or indium bumps about 25 microns thick are plated on. Then the resist is removed, and an etching step removes the sputtered metal layer between bumps (as well as a small fraction of the bump). In another method, the resist is patterned first, the bump material is sputtered on, and is then lifted off where it overlies the unetched resist.

The other high density method of attaching leads, flip chip technology, also involves plating bumps on the chip, but the chip is then flipped over and bonded to leads on the circuit board [111]. This method provides the highest bond density, the strongest bonds, and allows chips to be placed close to each other, resulting in short, low capacitance leads.

With both elements being rigid, careful bonding technique is needed, although there will be self-aligning forces. If the "circuit board" is silicon, standard techniques can produce fine metal traces similar to those on the chip, making especially dense connections and consequent improvements in performance as well as freedom from effects of differential expansion [78]. If not, there will be a limit to the maximum number of on-off cycles due to fatigue cracks starting at the chip-bump junction, and growing until the bond fails. Proper design of the bumps is necessary to prevent this from causing premature failures. Reference [114] gives a thorough description of an IBM technology which uses tin-lead solder bumps about 100 microns in diameter. The Hughes technology uses indium bumps which can be as small as 15 microns in diameter, bonded to pads 30×30 microns. Arrays as large as 256×256 have been produced [115]. An easier way of forming tin-lead solder bumps is to put special tin-lead wire into a wire bonder and make a ball bond, but without the attached wire lead. The bump is reflowed in a reducing atmosphere to make a round ball as in the IBM process [116]. (See also the section on hybrid pixel detectors, under VII, item 8.)

L 3. Encapsulation and Packaging

Again, chapters 13 of Ref. [12] and 3 of Ref. [23] are good places to start. Various types of plastic and ceramic packages are described with BeO (beryllia) recommended as having both a low dielectric constant and higher heat conductivity than the more common Al_2O_3 (alumina). Heat conductivities are also discussed: typical levels are of the order of 0.01 Watt/sq. cm. for unfinned packages using natural convection, 0.1 Watt/sq. cm. with finned packages and forced air, and 1 Watt/sq. cm. for liquid cooling. Heat pumps using the Peltier effect or heat pipes using evaporative cooling of a liquid that is

transported by hydrostatic and capillary forces [117] may be a good way of removing heat from difficult places such as vertex detectors at colliding beam machines.

We normally think of aluminum as a metal that does not corrode readily, covered as it is with a fraction of a micron of tough aluminum oxide. However the entire aluminum trace in an integrated circuit is only about half a micron thick, and it doesn't take much corrosion before the trace is gone. Unless bumps are plated on for flip chip or TAB bonding, at least part of the traces are unprotected where windows are made for external contacts. Under humid conditions, small amounts of HCl can act as a catalyst, converting $6 \text{H}_2\text{O} + 2 \text{Al}$ to $\text{Al}_2\text{O}_3 + 3\text{H}_2 + 3 \text{H}_2\text{O}$ (see reference [12], p 583). Also Na^+ ions readily diffuse through SiO_2 , and can change the electrical properties of a circuit if they are allowed in the package. (Until good packaging methods were developed, this was a common source of delayed failure, as the ions slowly diffused into the chip.) Thus, encapsulation with epoxies or silicones, or the use of a hermetic package is a very important step in the fabrication of the final product. The figure on page 583 of reference [12] shows what massive differences exist in moisture permeability of some packaging materials.

Epoxy is usually considered to be something that hardens, when mixed in the right proportions, into an inert mass that can be ignored thereafter. We have clean surfaces and if we want to be careful, we consider the effects of differences in thermal expansion coefficients. In microelectronics much more care is required. The ability to conform to micron scale topography is needed for proper sealing. Viscosity as the epoxy is being injected past delicate wire bonds is also. It must be free of trace contaminants such as Na^+ and Cl^- ions. With metal traces at different voltages separated by only a few microns, corrosion by galvanic action can increase if the wrong epoxy is used [118]. These problems and many others are met by using special compounds formulated for integrated circuit encapsulation.

It may be important to be sure there are no static charges on the insulating part of the chip surface before it is potted. If there are, you will have *lots* of time thereafter to study how they change the way it works. For example, it is not difficult to accumulate enough static charge to invert and make conducting, the silicon between adjacent strips in a silicon strip detector. From spark lengths, we are all painfully aware that thousands of volts can be generated by an ungrounded person in dry air. Inverting the underlying silicon with static charge sitting on the field oxide between strips is similar to inverting it with the charge on metal lines over the field oxide of an integrated circuit, and the threshold there is only tens of volts. If that doesn't ruin the chip, there is another snare waiting if the chip is irradiated with power on. There probably will be at least some places in the epoxy with an electric field. Charges liberated there by the radiation may drift along the lines of force as they do in the gate oxide, be trapped, and again change the circuit properties.

M. Yield

Chapter 14 of reference [12] provides an introduction to the methods of estimating and improving the yield of good chips. Some results from the fabrication of the Microplex III, a 128 channel silicon strip amplifier chip provides some useful examples.

Fig. 31 shows the number of working channels for each die on two 10 cm diameter wafers. Several kinds of defects are present. Regions along the left edges of both chips have metal traces that are too thin or missing due to over etching everywhere near the wafer edges. (The placement of the chips on the wafer put the metal closer to the edge on the left side.)

Both wafers have regions where the gain of one specific channel, number 101, is low, generally 20% to 50% of normal. Such a defect could only be due to a dimensional error in the channel 101 region on one of the master reticules that was used to make the masks by a step and repeat process. There also had to be a process variation that confined the problem to a limited region of each wafer. There were no visible defects, and the chips were usable for test assemblies, so we did not investigate further. If another run had been needed, of course, the reticules and masks that were stored at the foundry would have been checked. Both chips at 9J, in a generally good region of the wafers, have only the first few channels working. This is likely to be due to a defect in a mask, but not in the master reticule, since the other 149-150 positions do not have this problem.

In addition to these repeating defects, there are also some random ones. For instance, at position 3F of wafer B, channels 102 - 107 have zero gain. All 128 channels are working at that position on wafer A.

The relatively high yield seen here is due to the combination of:

1. a conservative design (5 micron rules were used for a 4 micron process and extensive modeling was done using SPICE)
2. a mature process (all the machines were well maintained; all the variables were carefully tuned).

The yield will drop if larger chips, finer geometry, processes with more fabrication steps, or recently developed processes are used -- all things that will be needed in the future. Improved manufacturing will certainly help, but understanding the yield will remain an important part of chip design.

The defects seen above have come from processing problems, from mask defects, from random point defects, and from combinations of them. Some are due to inherent limitations of the current process; others to excessive variations in it.

Some examples of process defects are:

1. Edge to center and wafer to wafer variations in chemical vapor deposition system gas flows cause variations in the thickness of the deposited layers. If parts of the layer are masked and the unprotected parts etched away by a liquid, time must be allowed to etch to the bottom of the thickest part. Meanwhile lateral etching has narrowed the thin parts, but not the bottom of the thicker ones, which the liquid reaches only at the end of the etching process (see Fig. 32). For

polysilicon gates, this will result in length variations. Replacing liquid etchants with reactive ion etching using directed beams provides a vertical cut that eliminates this problem.

2. Metal does not deposit easily on such vertical surfaces, however. To provide sloped ones leading down to contact holes, the glass insulator must be heated and reflowed. Here a too-thick glass layer can narrow contact regions.
3. As mentioned earlier, oxidation will cause wafers to expand. Etching the oxide on one side will add a warp which must be flattened with a vacuum chuck during lithographic steps. All this causes mask alignment problems, especially on the outer parts of the wafers.
4. Particulate matter landing on wafers or masks may lead to pinholes in insulating layers and to photoresist defects which then lead to problems in the following process steps.
5. Scratches, often on the outer parts of the wafer, can be a problem, particularly with manual handling.

Mask defects can be reduced by scanning the mask and making an on-line comparison with the data base. That was done with the Microplex masks. The defects thus probably came afterwards, perhaps from particulate matter. It is not normally practical to reinspect them between each use.

Point defects can come from particulate matter and also from defects in the silicon wafer that, for example, cause junction leakage. As the chip size is increased to reduce both handling and interconnect problems, these become more serious. A uniform density of $1/A_0$ reduces the yield of chips of area A by a factor of $\exp(-A/A_0)$. Fortunately, the defects are often not uniformly distributed, resulting in multiple defects in already dead chips and an increase in the number of survivors.

Devices that detect micron-sized particles in the air and on unpatterned wafers are in universal use in fabrication facilities. Detectors have been developed that use scattered laser light to detect defects and particles on patterned wafers, and so can frequently monitor wafers during the production process. They search for pattern variations that do not repeat with the known chip-to-chip dimensions [119].

Table 4 shows typical sources of wafer loss in a 2 micron CMOS fabrication facility, and, on surviving wafers, typical sources of die loss [120].

Table 4. Typical Sources of Yield Loss

Wafer breakage	2%	Particle Contamination	30%
Wafer misprocessing	7%	Design margin	6%
Process equipment failure	4%	Photolithographic errors	9%
Gross contamination	2%	Material defects	6%
		Process variation	9%
Total wafer loss	15%	Total die loss	60%

Thus the average yield expected for this process would be $85\% \times 40\% = 34\%$. (Multiplication within each column, rather than the addition of probabilities used in [120] gives lower losses.)

IV. A Survey of Application-Specific Circuits in High Energy Physics

This survey is given in the form of Table 5 directly below. For completeness, each chip is associated with its name and description, number of channels, its technology expressed in microns and its type. The designers and their addresses are included to allow access to the origins of the chips, and references allow the earliest publications to be consulted.

Table 5

The first ASIC chips for high energy physics were made in the mid-1970's for proportional chamber readout systems [121]. The first circuits, where the apparatus could not possibly work without custom VLSI-readout chips were for colliding-beam silicon strip vertex detectors and were developed in 1982-1984 [1], [2], and [122]. Since 1987, there has been an explosive growth in the development of specialized chips: amplifiers, comparators, high speed sample-and-hold chips, logic chips, and data analysis chips among others. Table 5 (The First Fifteen Years), immediately below, gives a partial list of such chips as of August, 1990. This list is limited to circuits that have actually been fabricated. Digital gate arrays and non-specialized CCDs are not included. A review of fast analog integrated circuits and technologies can be found in reference [123].

Name	Description	Nbr. Chan.	Tech. Type (um)	Designers and/or Authors	Addresses	References (journal: vol.page(s)'year) or (NS Session # at IEEE conf year) (unpublished) or (report: lab-number)
Microplex	silicon microstrip amplifier	128m	5 NMOS	J. Walker, S. Parker +	Hawaii, CIS	NIM 226.200'84; NIM A243.153'86; NS 33.57'86;
CAMEX-DCPLEX	silicon microstrip amplifier	60m	3.5 CMOS	W. Buttler, G. Lutz +	MPI, Fraun+	NIM A253.444'87; NS 34.491'87; NS 35.166'88
CAMEX-SCPLEX	Double correlated multiplexer silicon microstrip amplifier CMOS analog multiplexer	60,64,128m	3.5 CMOS	W. Buttler, G. Lutz +	MPI, Fraun+	NIM 226.196'84; NIM A253.439'87; NIM A257.587'87; NIM A263.163'88
JAMEX	4 Switched Capacitors silicon microstrip amplifier	64m	4 NJFET- 4 PMOS	W. Buttler, G. Lutz +	MPI, Fraun+	NIM A263.163'88; NIM A273.778'88; NS 36.246'89; XXIV ICHEP.1249'88; 14th ESSCIRC.171'88; Fermi-TM-1604; ICFA Instrum. Bull. 5.32'88
MX1	silicon microstrip amplifier	128m	5 CMOS	P. Seller	RAL	NIM A264.391'88; NIM A279.204'89; NIM A288.140'90;
MX2	silicon microstrip amplifier	128m	3 CMOS	J. Stanton	RAL	NIM A292.435'90; NS 38.69'91; 4th San Min.394'90
MX3	silicon microstrip amplifier	128m	3 CMOS	J. Stanton	RAL	NS 35.176'88
MX4	silicon microstrip amplifier	128m	3 CMOS		RAL	NIM A279.189'89; RAL-89-028
MXRP					RAL	NS 36.522'89
MX5	silicon microstrip amplifier	128m	1.5 CMOS	P. Seller, M. Tyndel+	RAL	
MX6	silicon microstrip amplifier	128m	3 CMOS		RAL	
ADE	silicon microstrip amplifier	64	5 CMOS	A. Olsen, E. Nygard+	SI	NIM A273.565'88
BVX3	silicon microstrip amplifier	8	2 CMOS	T. Zimmerman, C. Britton	Fermi, ORNL	Fermi TM 1754; Fermi-Conf-92-314'92
PADC	silicon microstrip preamp, comparator	64m	1 BipolSS	H. Ikeda, N. Ujje+	KEK, NTT	CF; NS 36.502'89; KEK 90-47 '90; NIM A300.335'91
SVX	silicon microstrip amp, comparator	128m	3 CMOS	S. Kleinfelder, H. Spieler	LBNL	NS 35.171'88; NS 36.35,477'89; NS 37.434,1120'90; NIM A289.388'90
Balder	silicon microstrip amp, comparator	48m	3 CMOS	A. Olsen, E. Nygard+	SI	CF
—	silicon microstrip amp, comparator	64	BipolTekS	E. Barberis, N. Cartiglia+	UCSC	NS 40.740 '93
AMPLEX	silicon pad detector amplifier	4,16m	3 CMOS	P. Jarron	CERN	14th ESSCIRC 660'89; CERN/EF-89-14; NIM A279.388'89; NIM A288.157,240'90
Bipoltest	fast silicon detector amplifier		3 SACMOS	P. Jarron	CERN	CERN/EF-89-14; CERN/ECP 90-13
AMPLEXF	silicon detector amplifier	4m	3 SACMOS	P. Jarron, T. Vallon-Hulth	CERN	CERN/EF-89-14
AMPLEX-SICAL	silicon calorimeter amplifier + sum	16m	3 CMOS	E. Beuville +	CERN, IMEC, INFNP, Surrey	NS 39-766'92
VIKING	silicon detector amplifier	128	1.5 CMOS	E. Nygard +	CERN	NIM A301.506'91; CERN-PPE-SSDD Jan. 15'92 (Milan)
QPA02	silicon detector amplifier	4	BipolQCL	T. Zimmerman, D. Christian +	Fermi, Penn	NS 36.507'89; NS 37.342,439'90, Fermi TM 1606
TRA403	silicon detector amplifier (obs)	4	Bipolar	B. Cake	LeCroy	LeCroy data sheet
—	silicon detector amplifier	1	VTCbga	C. Britton, Jr.	ORNL	
—	silicon pad preamplifier	1	VTCbga	E. Kennedy, C. Britton, Jr. +	ORNL	NS 38.58'91
2HICII	pad preamplifier (silicon?)	2	CMOS	C. Britton, Jr.	ORNL	pvt. comm. Britton, ORNL
DRO	(test) pixel amplifier, comparator	9 x 12m	3 SACMOS	F. Krummenacher, P. Jarron+	CERN, Lau, ETH	NIM A288.176'90; NIM 290.149'90
—	pixel amp, comparator, delay	16 x 63m	3 SACMOS	F. Krummenacher, P. Jarron+	CERN +	CERN/ECP 91-26; NS 39.654 '92
—	pixel detector readout 120x120um	10 x 64m	MOS	S. Gaalema	Hughes, UCB- SSL, SLAC	NIM A275.580'89; SLAC PUB 4942 '89, 5211'90

—	pixel detector readout 30 x 30um	256 x 256m		NMOS	S. Gaalema +	Hughes, UCB-SSL, SLAC	NIM A275.580'89; SLAC PUB 4942 '89, 5211'90
—	pixel detector readout 30 x 30um	256 x 256m		PMOS	T. Collins, S. Worley	Hughes, UCB-SSL, SLAC	SLAC PUB 5357'90
—	pixel detector readout 50x150um	64 x 32m		CMOS	O. Barkan, E. Atlas	Hughes, UCB-SSL, SLAC	SLAC PUB 5518 '91
—	pixel detector readout 300x320um	7 x 8m		CMOS	D. Raymond, A. Lewis +	RAL, Imp	NIM A310.552'91
—	charge preamp (high resistivity Si)			JFET-hr	P. Rehak, S. Rescia	BNL	NIM A288.168'90
—	microstrip gas chamber amplifier	8	3	SACMOS	F. Krummenacher, E. Enz+	SSS, INFNP	NIM A313.483'92
—	25 samples/chan LAr readout	16m	2	CMOS	U. Schoeneberg	Fraun, MPI	NIM A288.191'90
—	liquid ionization calorimeter preamp			JFET	S. Rescia +	BNL, Interfet	EFC2.91
—	pad readout preamp	4		BipolHVHF	C. Britton, Jr., E. Kennedy +	ORNL, UT	NS 39.1302'92
TRA-400	MWPC amplifier	4		Bipolar	B. Yamrone	BNL, LeCroy	
	(obsolete/unsupported)						
TRA-401	MWPC amplifier	4		Bipolar	B. Yamrone	BNL, LeCroy	NS 30.149'83; NIM 217.140'83
	(obsolete/unsupported)						
TRA-402	MWPC amplifier	4	4	Bipolar	B. Yamrone	LeCroy	
TRA-1000	detector (MWPC, Nal,...) preamp	1	10	Bipolar		LeCroy	
RBA	MWPC amplifier, comparator	2		ECL	M. Boussicut+	CEN-SAC	NS 25.144'78
RBB	MWPC delay, readout	8		PMOS	M. Boussicut+	CEN-SAC	NS 25.144'78
FILAS	MWPC amplifier	8		PMOS	J. Borel, G. Merckel+	CEN-GREN	NIM 131.583'75; NS 23.274'76; PSSCC 124 '74
—	MWPC amplifier	1	4	Bipolar	D. White	RAL	
QPA03	MWPC amplifier	4		BipolQCL	T. Zimmerman, J. Oliver	Fermi, Harvard	CF (unpublished)
VTX	MWPC amplifier	6	2	BipolQCL	T. Zimmerman, R. Yarema	Fermi	NS 39.742'92
MB43458	MWPC amplifier	4		Bipolar	K. Kondo	Tsu, Fujitsu	NS 33.933'86; NIM A268.75'88
MB43468	MWPC amplifier (inverting)	4		Bipolar	F. Suekane, T. Kondo+	KEK, Fujitsu	NS 33.73'86
—	straw ch preamp	1		ATT LA	M. Newcomer	Penn	NS 37.68 '90
ASD8	straw ch preamp, shaper, discriminator	8		BipolTekS	M. Newcomer, R. Van Berg	Penn	ECF2.105
ICON	current mode preamplifier	1	1.5	CMOS	M. Campbell +	CERN, IMEC	NS40.271 '93
—	TPC front end amplifier/line driver	1		CMOS	W. Buttler, B. Hosticka +	Fraun, MPI, Dort	Bari, Europhysics, High Energy 595 '85
—	charge amp, Aleph EM calorimeter	16m	2	CMOS	J. Veillet, S. Richer +	LAL	NS 39.762'92
—	charge amp, analog sum	16m	5	CMOS	J. Lewis +	RAL	NIM A288.197'90
ALOG 0 0	logarithmic amplifier	1		Bipolar	V. Goursky, G. Gaucher	CEN-Sac	CF, NS 35.181'88
—	shaper amplifier	2		Bipol A400	H. Ikeda, M. Ikeda +	KEK	NIM A292.439'90
—	transimpedance preamp for photodiodes	1	2	CMOS	D. Binkley, J. Rochelle+	CTI, UT	NS 39.747'92
MVL-100	amp, comparator	1		Bipolar		LeCroy	
	(obsolete/unsupported)						
MVL-400A	comparator (obsolete/unsupported)	4		Bipolar	B. Yamrone	BNL, LeCroy	NS 30.149'83; NIM 217.140'83
MVL-407	400 MHz comparator	4	4	Bipolar	B. Yamrone	LeCroy	LeCroy data sheet
—	comparator (for Opal, Aleph)	1	4	Bipolar		RAL	P. Sharp, Snowmass '88, NIM A286.109 '90
—	comparator	1	3	CMOS	A. Olsen, P. Jarron	CERN	CERN-EF 89-14 '89
—	constant fraction comparator	1	2	CMOS	D. Binkley, M. Simpson +	CTI, EG&G, UT	NS 38.1754'91
DCFD	differentiation-const. frac. comparator			Bipolar	M. Tanaka, H. Ikeda+	KEK	NS 39.1321 '92
ICO5	comparator, latch	4		BipolQCL	M. Haldeman, S. Holm	Fermi	EFCI.169
IC01	2 channel sum, comp, quad latch	2		BipolQCL	M. Haldeman +	Fermi, Penn	CF, NS 36.507'89; NS 37.342'90
IC02	logic, analog sum, latch driver	3,4,1		BipolQCL	M. Haldeman, B. Merkel	Fermi	CF, NS 37.342'90
IC04	5 chan logic, number of hits (1-8)	5,1		BipolQCL	M. Haldeman, J. Hoff	Fermi	CF, NS 37.342'90
IC03	4 DAC threshold set, 1 ADC readback	4,1		CMOS	C. Swoboda, M. Hawkey+	Fermi	CF, NS 37.342'90
MDD500A	logic delay	4+ref	5	CMOS		LeCroy	
C10115	programmable trigger logic array	16x16		NMOS	E. Platner, J. Gould+	BNL, Synertek	NIM 140.549'77; NS 24.225'77
MLL400	256 bit 330 MHz shift register	4m		CMOS-S	R. Stewart, E. Platner	BNL, RCA, LeCroy	NIM 217.140'83; NS 30.149'83

MLL400A	256 bit 250 MHz shift register	2	4	CMOS-S		LeCroy	
-	32x32 bit Opal trigger processor	1		CMOS	M. French, F. Storch	RAL	
-	256x24 bit histogrammer	2	2	CMOS	F. Storch, J. Alsford	RAL	NS 35.209'88
CAP	CCD cluster array processor-sparsifier	1	2	CMOS	D. Phillips, A. Gillman	RAL	RAL 89-064'89; NS 37.305'90
DMT	Digital Mean Timer		1.2	CMOS	J. Mann, J. Chapman	Michigan	EFC1.149'91; EFC2.43'92
Microstore	256 cell sample and hold	1	3	NMOS	J. Walker, S. Chae +	CIS, SLAC	NS 32.616,622'85; NS 33.81 '86; NS34.142 '87
AN201	(commercial version of Microstore)					Analytek	
CDU	4 cell sample and hold	32	3	NMOS	J. Walker, G. Haller +	CIS, SLAC	NS 33.221'86; NS 34.142'87
-	2048 cell (128/chan.) sample and hold	16	3	CMOS	S. Kleinfelder	LBNL	NS 35.151'88
SCA	4096 cell (256/chan.) sample and hold	16	2	CMOS	S. Kleinfelder	LBNL	NS 37.1230'90
MIQ401	integrate and hold (QMUX)	4m	5	Bipolar	J. Mueller, L. Levit +	LeCroy	NS 33.897'86
MVV200	320 cell/channel waveform recorder	2	5	CCD	W. Farr, L. Levit +	LeCroy	NS 32.636'85
-	1 GHz analog sampler			GaAs		Triumf	
-	TVC, 8 intervals	1	1.6	CMOS	A. Stevens +	ATT, Penn, C	NS 36.517'89
SAPE	64 cell swc analog pipeline + buffer	4	3	CMOS	A. Olsen, P. Jarron +	SI, CERN	ECFA Study, Barcelona 678-9, Sept'89; CERN/ECP 90-13; NSConf; 543'90; ECFA 90-133 Vol. III, 84 Dec'90
SAPE2	64 cell, 60 MHz analog pipeline	4	1.5	CMOS	A. Olsen, F. Anghinolfi +	CERN, SI	SSC Symp. on Det. Research, Ft. Worth, 465'90; NS Conf 543'90
CSI	charge sampling integrator (swc)	4	1.5	CMOS	F. Anghinolfi, P. Aspell +	CERN	ECF1.135
-	58 cell swc analog pipeline	4	2.5	CMOS	J. Moeschén + (Zeus)	Fraun	NIM A265.321'88; NS 36.465'89; CF; NIM A277.217'89; NIM A288.180'90; EFC2.117
-	12 chan x 8 deep buffer + 13 nonbuffer	25m	2.5	CMOS	J. Moeschén + (Zeus)	Fraun	NIM A265.321'88; NS 36.465'89; CF; NIM A277.217'89; NIM A288.180'90; EFC2.117
-	amp+25 cell swc analog pipeline + trigger	16m	2	CMOS	U. Schoeneberg + (H1)	Fraun, MPI	NIM A288.191'90
AACC	analog amplifier, comparator	64		DIBipol	J. DeWitt	UCSC	NIM A279.186'89; NIM A288.209'90; NIM A310.557'91
ALP	address list processor	256	1.2	CMOS	S. Kleinfelder	LBNL	EFC1.219
DTSC	64 bit deep -> 16 bit deep FIFO	64	1.2	CMOS	J. Dewitt	UCSC	NIM A288.209'90; NIM A310.557'91
MQT200F	charge to time converter	1	10	Bipolar		LeCroy	
QAC-TAC	12 bit charge, time to amplitude	2		AMS	H. Ikeda, T. Tanimori	KEK	KEK 88-85'88; NIM 350.305 '94
	analog sampler, 32 cells/chan	32	1.5	CMOS	H. Ikeda, S. Inaba	KEK	KEK 88-85'88; NS 38.135 '91
CSX-E	fastbus cable segment transceiver	5	1.5	Bipol	H. Walz, D. Gustavson	SLAC	NS 36.532'89; SLAC-TN-86-7
TMC1004	1024 bit x 1ns time memory cell (TDC)	4	0.8	CMOS	Y. Arai, T. Matsumura+	KEK, NTT	CF; KEK 88-78'88; NS 36.528'89; NS 39.784'92; JSSC 27.359'92
TCC/AMU	time to charge / analog memory unit	1	1.2	CMOS	E. Gerds, J. Van der Spiegel +	Penn	EFC2.105
KGL5202	1.2 GHz clock driver	1	0.8	GaAs	H. Mori, O. Sasaki	OKI, KEK	CF
FTDC 01	32 ns TDC, 0.5 ns resolution	3m		CMOS	J.-F. Genat, G. Hantelle+	U. Paris	CF; EFC1.229
CDN 400	TDC	4		ECL	V. Goursky, G. Hantelle	CEN-SAC	CF; NS 35.181'88
MTD110	TDC 10 bit	1	4	Bipolar		LeCroy	
MTD132	TDC 0.75ns, 16-bit, 16 hit/chan, 2 hit<15ns	8m	1.2	CMOS	S. Kleinfelder, T. Majors+	LeCroy	NS 38.97'91
CTR 400	TDC or scalars	8		ECL	V. Goursky	CEN-SAC	CF; NS 35.181'88
KGL4101	256 bit, 1.2 GHz shift register TDC	1	0.8	GaAs	H. Mori, O. Sasaki +	OKI, KEK	NS 36.512'89; GaAs IC Symp. Tech. Digest, 10.135'88; CF
QT100C	charge to time converter	1		Bipolar		LeCroy	
Amchip	associative memory for pattern recog.	1	1.5	CMOS	S. Amendolia, S. Galeotti +	INFNP, ESS, +	NS 39.795'92
MB43669	TAC	1		AMS	M. Tanaka, H. Ikeda	KEK	NS 38.301'91
-	TAC	2		Bipol/LGA	T. Tanimori, H. Ikeda +	KEK, Tok, Ts	NS 36.497'89, CF
-	current integrator	2		Bipol/LGA	T. Tanimori, H. Ikeda +	KEK, Tok, Ts	NS 36.497'89

CQT 01	gated integrator	1		Bipolar	V. Goursky +	CEN-SAC	CF; NS 35.181'88
APL 01 R	linear gate	1		Bipolar	V. Goursky +	CEN-SAC	CF; NS 35.181'88
INP 20	image processor	2		CMOS	D. David	CEN-Gren	CF
EPA4	image processor	4		CMOS	J. Kaiser	CEN-SAC	CF
DCU	digital correction unit	1	2	CMOS	S. MacKenzie, L. Palfraath +	SLAC	NS 34.250'87
FPADDER	floating point adder	1	1.2	CMOS	J. Hoff, W. Foster	Fermi	Fermi TM 1779
—	controller for event builder	1	2	CMOS	S. Newett	RAL, EP, CERN	NIM A269.538'88
—	data chip for event builder	1	2	CMOS	R. Marbot +	EP, RAL, CERN	NIM A269.538'88
SFM25	fastbus, Sparc RISC interface		1.2	CMOS	A. Marchioro, M. Letheren	CERN	CF
—	line driver	2		Bipolar	M. French, J. Alsford	RAL	IEEE-NS, San Francisco'87, unpub
PiPAD	12 bit 1 MHz ADC		3	SACMOS	V. Valencic, F. Anghinolfi,	Mead, CERN	CERN/ECP 90-13, ISCAS 90
Amplex-SiCAL	extended dynamic range amplifier	16	3	CMOS	E. Beuville, P. Jarron	Saclay, CERN	NPB 23A.198'91; NS 39.766'92
CURC	(test) fast amplifier		1.5	CMOS	P. Jarron, S. Bryfaert +	CERN, IMEC	
DPC	pipelined integrator, encoder	1	2	BiCMOS	T. Zimmerman, W. Foster +	Fermi	EFC2.131'92
—	12 bit ADC	1	2	CMOS	O. Milgrome, S. Kleinfelder+	LBNL	NS 39.771'92

Abbreviations:

amp	amplifier
AMS	Analog Master Slice (semicustom bipolar, A600 Fujitsu, Tokyo)
ATT LA	AT&T Corp. Linear Array (LA200)
BiPoIA400	Fujitsu Analog Master Slice A400
BipolHVHF	Harris VHF dielectrically isolated bipolar
BipolLGA	Bipolar linear gate array
BipolQCL	Quick Chip 2 (a Tektronix bipolar linear array)
BipolSS	Bipolar, NTT Super Self-aligned Technology
BipolTekS	Bipolar, Tektronix SHPI (vert. npn, lat. pnp, p chan JFET)
CF	Survey of integrated circuits by Christian Fabjan, CERN
CMOSGA	CMOS Gate array
CMOS-S	CMOS-Silicon on Sapphire
comp	comparator (=discriminator)
dbl	double (as in double hit resolution time)
DIBipol	dielectric-isolated bipolar
Germany	
diff.	differential
EFC1	Proc. 1st Ann. Conf. on Electronics for Future Colliders (LeCroy)
EFC2	Proc. 1992 Conf. on Electronics for Future Colliders (LeCroy)
ESSCIRC	Proc. European Solid State Circuits Conference
FIFO	first in - first out buffer
ICFA	International Committee for Future Accelerators
ICHEP	Proc. International Conference on High Energy Physics
ISCAS	International Symposium on Circuits and Systems
J-bipol	JFET-bipolar
JFET-hr	JFET on high resistivity silicon
m	multiplexed output
NIM	Nuclear Instruments and Methods in Physics Research

Institutions:

Analytek	Analytek, 10261 Bubb Rd. Cupertino, CA USA 408-725-2560
ATT	AT&T Bell Laboratories, Whippany, NJ, USA
BNL	Brookhaven National Lab, Upton, NY, USA
C	Catholic University of Leuven, Heverlee, Belgium
CEN-SAC	CEN/Saclay, Gif-sur-Yvette, France
CEN-GREN	CEN/Grenoble, Grenoble, France
CERN	CERN, Geneva, Switzerland
CIS	Center for Integrated Systems, Stanford Univ, Stanford, CA USA
CTI	Computer Technology and Imaging, Knoxville, TN USA
Dort	Univ. Dortmund, Dortmund, Germany
EP	Ecole Polytechnique, LPHNE, Paris, France
ESS	European Silicon Structures, Grenoble, France
ETH	Inst. for High Energy Physics, ETH, Zurich, Switzerland
Fermi	Fermilab, Batavia, IL USA
Fraun	Fraunhofer Institut für Mikroelektronische Schaltungen, Duisburg,
Fujitsu	Fujitsu Microelectronics Ltd, Tokyo, Japan
Hawaii	U. of Hawaii, Honolulu, HI USA
IMEC	IMEC, Leuven, Belgium
IMP	Imperial College, London, UK
INFNP	Istituto Naz. Fiz. Nucleare, Pisa, Italy
Int	Interfet Corp, 322 Gold St. Garland, TX 75042 USA
KEK	Nat'l. Lab for Hi. En. Phys., Tsukuba-shi, Ibaraki-ken, Japan
LAL	Lab. de l'Accelérateur Lineaire, Orsay, France
Lau	Electronics Lab LEG, EPFL, Lausanne, Switzerland
LBNL	Lawrence Berkeley National Lab, Berkeley, CA USA
LeCroy	LeCroy Corp., Chestnut Ridge, NY, USA
Michigan	U. of Michigan, High Energy Physics Dept. Ann Arbor, MI USA

NJFET	N channel JFET
NPB	Nuclear Physics B (Proc. Suppl.)
NS	IEEE Transactions on Nuclear Science (abstract number may be given if not published)
NSConf	Conference Record of the IEEE Nuclear Science Symposium
obs	obsolete or unsupported
PSSCC	Proc. Solid-State Circuits Conf.
SACMOS	Self-Aligned Contacts MOS (Faselec A6, Zurich) (3um SACMOS=1.5um CMOS)
USA)	
San. Min.	San Miniato Topical Seminar
SSCir	solid state circuits
sram	static random access memory
swc	switched capacitor
TAC	time to amplitude (voltage or current) converter
TDC	time to digital converter
trig	trigger (usually a merged output from >1 channel)
TVC	time to voltage converter
VTCbga	VTC bipolar gate array

Obsolete or unsupported LeCroy units not listed above

MIL	200	current comparator
MQT	200, 201	Q→T (but MQT 200F not obsolete)
MSC	300, 301, 302	scaler
MVT	200	V→T
QT	100C	Q→T

MPI	Max Planck Institut fur Physik und Astrophysik, München, Germany
NLHEP	Natl. Lab. for High Energy Physics, 1-1 Cho, Tsukuba, 305, Japan
OKI	OKI Electric Industry Co. Ltd., Tokyo, Japan
ORNL	Oak Ridge National Laboratory, Oak Ridge, TN, USA
Paris	U. of Paris, Paris, France
Penn	U. of Pennsylvania, Philadelphia, PA, USA
RAL	Rutherford Appleton Lab, Didcot, UK
RCA	Radio Corp. of America (now part of the General Electric Corp.,
SI	Senter for Industriforskning, Oslo, Norway
SLAC	Stanford Linear Accelerator Center, Stanford, CA, USA
SSS	Smart Silicon Systems, S.A., Lausanne, Switzerland
Surrey	U. of Surrey, Guildford, Surrey, UK
Synertek	Synertek, 2050 Coronado Dr. Santa Clara, CA 95051 USA
Tok	Inst. for Cosmic Ray Research, U. Tokyo, Tokyo 188, Japan
Ts	Toshiba Co. Ltd. Tokyo 183 Japan
Tsu	Univ. of Tsukuba, Sakura, Niihari, Ibaraki 305 Japan
UCB-SSL	U. California, Berkeley, Space Sci. Lab, Berkeley, CA, USA
UCSC	Univ. of California, Santa Cruz, CA, USA
UT	Univ. of Tennessee, Knoxville, TN, USA
VTC	VTC Inc., Bloomington MN

V. Technology Trends

A plot familiar to most high energy physicists shows the beam energy (and now at the upper end, for colliders, the equivalent single beam energy) achieved by particle accelerators [124]. Relentlessly exponential since 1930, it covers twelve orders of magnitude. A detailed plot shows many curves, with the baton being passed from Cockcroft-Walton and Van de Graaff accelerators to cyclotrons, betatrons, synchrotrons, strong focusing synchrotrons, and colliders. Similar curves exist for the microelectronics industry, also the result of many changes, both in design and fabrication technology. Fig. 33 shows the minimum feature size in production chips as a function of time [125]. There is a factor of 2 drop every 5 to 6 years.

A. Scaling

During the early 1970's, while commercial processes were in the 5-10 micron range, experimental 1 micron devices were being fabricated as part of a program to change many aspects of MOS transistors in a coordinated way to solve problems that arose from the reduction in scale [126]. For example, depletion distances depend primarily on doping and voltage levels, not on device sizes. When the non-shrunk source-substrate and drain-substrate depletion regions start to occupy a large part of the channel length, it becomes difficult to turn transistors off.

In the scaling program, all vertical and horizontal dimensions as well as supply voltages, are reduced by a common factor, s ($s < 1$), while the substrate doping is increased by the same factor. Then the depletion width,

$$w = \left\{ \frac{2\epsilon_{si}(\Psi_b + V_{sub})}{qN_a} \right\}^{\frac{1}{2}} \quad (14)$$

becomes

$$w' = \left\{ \frac{2\epsilon_{si}(\Psi'_b + sV_{sub})}{(qN_a/s)} \right\}^{\frac{1}{2}} \quad (15)$$

$$w' \cong ws$$

where $\Psi_b = (2kT/q) \ln(N_a/n_i)$ is the built-in junction potential, V_{sub} is the source to substrate voltage, N_a is the substrate doping concentration, and n_i is the intrinsic carrier concentration. The built-in potential term prevents exact scaling, but can be mostly canceled out by another term that will come from the polysilicon gate work function. To this approximation then, the depletion distance, w' , is then reduced to match the new dimensions.

Similar substitutions show the threshold voltage and drain-source currents also decrease by s . Vertical scaling requires thinner gate oxides and drain/source doping distances, made possible by

carefully controlled ion-implants. Table 6, immediately below, lists some of the resulting parameter changes:

Table 6 – Scaling Changes

<u>Imposed Changes</u>		<u>Derived Changes</u>	
All lateral dimensions	s	Electric fields = voltage / distance	1
All vertical dimensions	s	Internal voltages	s
(including implant depths, gate oxide thickness, ..)		(including threshold voltages)	
Substrate doping concentration	1/s	Currents	s
Supply voltages	s	Capacitances = Area / thickness	s
		Delay time / circuit = VC / I	s
		Power / Circuit = VI	s ²
		Power density = VI / Area	1

So to the expected increase in circuit density and decrease in cost, we can add higher speed and lower power per device, two things that do not usually go together. The electric field magnitude and direction are the same at matching points on the two circuits. A number of difficulties arise in fabricating circuits on ever smaller scales, in addition to the obvious ones associated with fine-line lithography. Aluminum from overlying contacts is more likely to spike through the thinner sources and drains, shorting out the diodes that isolate them from the substrate. In the region of 0.1 micron feature size, the scaled gate oxide thickness of a few tens of nanometers becomes thin enough so electrons may tunnel through its potential barrier. The decrease in transistor channel area causes an increase in 1/f noise. If this becomes a limiting factor in the performance of an analog circuit, another form of scaling, in which the voltage is reduced by the square root of the scaling factor, may be desirable [127].

Exact scaling can not be achieved [51], [53], and [128]. For example, subthreshold currents, work function differences, and contact resistances do not scale. Often the supply voltage can not be reduced because of the need to be compatible with other types of chips. This produces higher values of electric fields which may cause breakdown, and may project high energy carriers ("hot carriers") into the gate oxide where they can produce damage and threshold shifts similar those from radiation damage. To prevent this, supply voltages are sometimes lowered within each chip. Also, variations of the set of rules of Table 6 have been used in an attempt to overcome some of these problems.

B. Progress and Limits

Minimum sized features are usually in poly structures; metal lines, on top in hilly terrain, must be wider. Fig. 34, the minimum line-to-line pitch, shows poly, metal 1, and widest of all, the top level metal 2, all dropping at the same rate [129]. Of course, this relentless decrease must eventually come to an end, probably starting within the next 10 or so years, perhaps around 0.1 micron. (Be aware though, the literature is littered with surpassed limits). The inherent limits are discussed by R. Keyes [130]. Figs. 35 and 36 are from that paper. Fig. 35 shows, as a function of time, the area used to store a bit of information. Fig. 36 shows the number of atoms used to dope the base of a logic bipolar transistor. It's now less than 100,000, an impressively small number, although still somewhat more than the number of xenon atoms (35) used to spell out the letters "IBM" by a scanning tunneling microscope [131]. (Information retrieval time by the microscope may be a bit slow, however.)

With a similar rate, but with less publicity, chips have been getting larger, as shown in Fig. 37, [129]. (The most advanced steppers now have fields of about 2×2 square cm.) In a way, this is even more remarkable: the increase in area requires a lower density of fatal defects at the same time that the smaller feature sizes makes even the same density more difficult to sustain. Defect density, of course, is strongly dependent on the nature of the fabrication facility, but is observed by some to increase with the inverse square of the feature size [132].

Numerous attempts have been made to use the entire area of the wafer for a single chip--wafer scale integration. Such integration is pointless if the feature size is much larger than for state of the art VLSI chips, and at these necessary sizes, defect free wafers cannot now be made. So self-testing and substitution of flawed parts with redundant cells must be employed. Extra area is needed for them as well as for the larger power busses and the usually longer and more numerous lines carrying internal signals. Thus far wafer scale integration has not been able to compete with VLSI, although the development of test and replacement technology will likely prove useful to both [132]. There is one current application, however: entire wafers containing metal traces have been used as a substrate for other chips, serving as a super-high resolution circuit board with a perfectly matching coefficient of thermal expansion.

Fig. 38 shows the number of transistors per chip for microprocessors and DRAMs as a function of time [125]. It is about what might be expected for an increase, scaled by area and divided by the square of the feature size. Some things which do not get much smaller, for example, output pad sizes, may be compensated by others such as multiple level metal layers, which allow denser use of the chip area even without a decrease in feature size. Fig. 39 shows a similar rate of device number increase for bipolar chips [133]. Multiple levels of metal are particularly important for semi-custom circuits such as gate and logic arrays, due to the increasing difficulty in connecting a large fraction of the prepositioned elements as the number on an individual chip increases. The reduced energy dissipation per operation as a function of time is shown in Fig. 40 [130].

Bipolar circuits have traditionally been faster than MOS circuits, but used to be far less compact. However, new bipolar technology has reduced circuit sizes and capacitance, and increased speed. In addition to benefiting from the general improvement in fabrication equipment, new, self-aligned processes are used. For example, a single cut, made by one mask, may delineate the region for the implanted region of the base, and in latter steps, a heavily doped poly emitter contact. The (single crystal) emitter is then made by out-diffusion of some of the contact dopant, putting a layer of doped poly between any metal lines and the single crystal emitter [134]. It came as a surprise that this made a dramatic increase in current gain from prior values in the 50 - 100 range to ones in excess of 1000 [135].

Space also has been saved by isolation devices using a trench made by an anisotropic etch [133]. Current leakage along the less than perfect trench faces may be a problem. Reference [123] contains information on bipolar technology and as well as additional references.

Fig. 41 shows the decrease in ring oscillator delay time per stage as the circuit is changed from emitter coupled logic (ECL) to non-threshold logic (NTL) using a self-aligned process [133]. Fig. 42 shows delay times vs power for a number of technologies including scaled NMOS and CMOS [136]. Fig. 43 shows another delay time: the ever increasing time to define and design an integrated circuit once you decide you want one [137]. Note the last point plotted is for 1980. During the 1980's new programs such as Magic were developed that speeded up routine layout and circuit parameter extraction, so that even the increasing complexity of the circuits did not increase design time. The use of gate arrays and standard cells further decreased the design time. Reference [138] gives typical times and costs for such arrays, as well as for full custom chips, as a function of device size. Despite these improvements, the time to define the chip--its functions, performance, size, architecture, etc.--has continued an unabated climb [139]. Research (and work by new companies) is now underway on this problem.

So we see that almost everything is not only getting better and better in every way, it is doing so exponentially. Fig. 44 shows one change that has managed to avoid this function: the cost of wafer exposure machines, which is increasing *faster* than exponentially [140]. No physical laws are known to limit this kind of improvement.

VI. New Technologies

Detectors capable of working under the brutal conditions at the coming generation of colliders will surely have to use new technologies; several candidates will be mentioned here. Nevertheless, it will be well to keep in mind a quote from reference [137]:

"A common trap is replacing a known technology with too high a cost, or known limitations, with a new technology having unknown cost and limitations."

Some changes, such as the decrease in feature size or increase in speed of silicon VLSI chips, seem to roll along as predictably as ticks from a Swiss watch (which are also now made with integrated circuits). Other changes, e.g., the widespread use of semiconductors made from elements in columns III

and V of the periodic table, for example gallium arsenide, or semiconductors using band-gap engineering, are much less certain. Less is known about them, and they must compete with the relentless (for now) moving target of silicon technology. Nevertheless, they will surely find some niche markets, and their physics is interesting in its own right.

A. Gallium Arsenide

Gallium arsenide has several advantages over silicon: a low-field electron mobility more than five times greater, the ability to make efficient light emitting diodes and lasers (the electron energy change and momentum change crossing the band gap matches that of a photon--phonons are not needed to take up excess momentum), and greater radiation hardness. A larger natural band-gap than that of silicon (see Table 1) makes possible a higher resistivity, nearly insulating substrate. In addition, it is easier to modify the band-gap, for example, by substituting aluminum or indium for gallium and phosphorus for arsenic.

The higher maximum speed provided by its mobility can also be traded off for less power at any given circuit speed. This mobility also reduces the effective transistor channel resistance which is the principle source of white noise in low-noise amplifiers (but the $1/f$ noise is worse). The electron velocities in GaAs and silicon are shown in Fig. 45. The *maximum* high field electron velocity is actually less than that of silicon: 5 cm per microsecond vs 10 for silicon at electric fields over 0.5 MV per cm. In addition, GaAs hole mobility is slightly less than that of silicon.

A major disadvantage is that it does not form a useful native oxide. SiO_2 , for example, is used on silicon for surface passivation (bonding mobile electrons from dangling bonds at the surface), for an insulating and dielectric layer, for a mask during many fabrication tasks, and for a capping layer to reduce evaporation of dopant atoms during high temperature processes. Device isolation is not as good, and there are no complementary devices. It is difficult to anneal GaAs as effectively as Si; the arsenic tends to evaporate first. Yet, just as in the case of silicon, high temperature anneals are important in activating ion-implanted dopant atoms--both to place them in vacant lattice sites, and to reduce damage caused by the incoming ions.

The greater radiation hardness, in large part, is due to things GaAs does *not* have (and which could be left off of silicon circuits were they not so useful):

1. no gate oxides, so no trapped-hole induced threshold shifts (JFETs with reverse biased, metal gate-semiconductor diodes are used instead)
2. no field oxides, so the field does not get inverted and cause leakage (if an insulating layer is used instead, there are so many trapping states at the interface, even massive radiation doses do not cause conduction)
3. an already short minority carrier lifetime, bad for bipolars, but radiation then does not make it much worse

4. that large band gap, so midgap states made by nuclear recoil do not readily cause leakage in depletion regions.

Reference [141] provides a good introduction to GaAs. Reference [27], pp 63-66 discusses radiation effects and gives further references to the literature.

B. Band-Gap Engineering

Molecular beam epitaxy, a method for laying down precisely defined layers of different substances, was invented at Bell Labs in the late 1960's. It provided the ability not only to modify the band gap, but to make novel structures, called superlattices, in which the band gap varies significantly in distances comparable to the carrier thermal de Broglie wavelength -- tens to hundreds of atoms thick. For example, alternating the doping in adjacent layers between n and p type will produce alternate layers of fixed positive and negative charge, as the dopant atoms ionize and the electrons and holes annihilate. Electrons, injected in such a structure, can be attracted to the positively charged layers. Using a single layer of graded doping in the base of a bipolar transistor, a built in electric field can produce a drift velocity ten times greater than the diffusion velocity in normal transistors. An oscillation frequency of 45 GHz has been achieved using that technique.

High speed field effect transistors have also been made with a single layer of doped AlGaAs to donate electrons which diffuse to an underlying layer of undoped GaAs. These electrons have a low scattering probability with a mobility of 8000 cm²/volt sec., double that of the doped GaAs used in normal MESFETs[136]. As in MESFETs, transistor control is by an overlying gate. Naming is by acronym, two in this case: MODFET (MODulation Doped FET) and/or HEMT (High Electron Mobility Transistor).

One device that may play an important role in the detection and amplification of weak light signals such as those from scintillating fibers, is the avalanche photodiode. Standard ones, while more efficient than photomultipliers, are subject to larger fluctuations if the electrons make holes which travel backwards, gain energy, and by impact ionization, make more electrons. Band gap engineering can make boundaries with larger steps in the conduction band than the valence band. Tilting these steps with an applied electric field can produce a structure in which only the electrons, traveling along the bottom of the conduction band, see a field (at the descending steps) large enough to produce multiplication.

Many other possibilities exist, including lasers in which population inversions are made when electrons tunnel from high to low voltage levels and structures that exhibit negative differential resistance. Three introductory articles are given in references [136], [142] and [143].

C. Fiber Optics

The use of fiber optics, with its submillimeter fibers, high band widths, and freedom from interference, promises to be one of the key technologies in data transmission to and from large collider detectors with their conflicting needs of hermetic calorimeters and large tracking channel counts. They already play an important role in the data acquisition system of the SLD detector at SLAC. This is a rapidly developing field. Reference [144] lists some examples, mostly from high energy physics that represent only a tiny fraction of the literature. References [6], chapter 12, and [10], chapter 7 also have material on photonics.

One future problem, heat dissipation (only a few percent of the energy of solid state laser or light emitting diode systems goes into useful light), may be solved by new, highly efficient microlasers or by modulating light brought in on fibers from external sources. Other uses are possible, for example, pattern recognition networks using signals from bundles of fibers in parallel, that might be used in triggering systems [145]. Several additional examples will be mentioned in the next section.

VII. Some Suggestions for Future VLSI Development

Given the great scope and rapid pace of developments in integrated circuits, and the cost of most really new technologies, there is only a relatively modest amount that high energy physicists and engineers working with them are likely to be able to do. For example, some of the needs of experiments at the new colliders would include:

1. faster rise and fall times
2. reduced generation of and/or improved removal of heat
3. increased digital functionality in each circuit
4. improved reliability; easier diagnosis and repair
5. chips with multichannel analog inputs and low impedance, multiplexed, digital outputs
6. increased radiation hardness for front-end chips
7. wider, faster busses and improved modular electronics systems
8. space-point detectors with, in many cases, time-tagged, sparse field readout
9. high quantum efficiency light detectors with submillimeter channel pitch
10. high-speed, radiation-hard semiconductor detectors.

Free help may be available for a number of these items. Thus a large part of what we might do could be to closely follow developments in these fields, adapting appropriate ones for our use. Here are some comments on these items (especially the free part):

Items 1 through 4:

These are of great importance to the computer industry, which is devoting large resources to technologies which high energy physics will also be able to use.

Item 5:

This is important to the automotive and aerospace industries, to name just two.

Item 6:

A community comparable in size to that working in high energy physics had developed radiation hard electronics for over two decades for military applications. The pace of this work, and the number of companies in the field is now reduced, but much valuable information has been published (often in the issues of the IEEE Transactions on Nuclear Science covering the annual conference on nuclear and space radiation effects). Much, however, is still secret that could be of great use in physics experiments. New work may be needed to understand and solve the problems coming from the increase of noise under radiation. More will certainly be needed to understand the mechanisms of bulk damage from high-energy collisions, but that is of importance mainly for detectors rather than for VLSI circuits. Many of those working on radiation hardening are quite interested in high energy physics, and it should be possible to get help from some of them.

Item 7 - - Modular electronics:

The first generation of standardized electronics, using NIM modules, was easier to assemble and maintain than the individually soldered circuits it replaced. Some decrease in that ease, but a great increase in capability, came with CAMAC. Fastbus, in turn, provided far larger boards with more capabilities and a wider, faster, more flexible bus system. The bus made use, in one dimension anyway, of that larger board. The price has been a more expensive, mechanically awkward system with components which are difficult to access, replace and repair. Two special tools are required for something as simple as changing a module. Many Fastbus boards use chips with only modest levels of integration. Fastbus is unlikely to play a significant role in the future, but a backward-compatible upgrade of CAMAC, FASTCAMAC (www.yale.edu/fastcamac) proposed by Satish Dhawan of Yale and a number of CAMAC manufacturers, will have comparable data transmission speeds, even without counting the time it takes to change the boards.

A significant reduction in board area would be possible with the use of programmable logic chips such as gate arrays, chips assembled from predesigned cells, and fully custom chips. Electrical connections could be limited to a few sturdy power and signal lines, while intermodule data and instruction busses could take advantage of the small size and large band width of fiber optics. Since distances are short, and optical power readily available, simple, reliable proximity connections to a back plane fiber system would be adequate.

It is beyond the scope of this article to even suggest a specific system, since detailed surveys of needs as well as competing technologies must be evaluated. It would be a good idea to seek maximum compatibility with some existing industry standard. For instance, the power or mechanical standards might match that used in VME crates, allowing downward compatibility with many other modules [146].

It might be desirable to use an intermediate level of modularity. Just as standard cells for semicustom chips use a standard scheme for power busses and cell-to-cell interconnections, so might the *external* contacts of the chips designed for use in this system, permitting use of a number of standardized boards. With more potential users, it could be worthwhile to put in the effort to develop sets of standardized high performance chips such as ADC's, TDC's, scalers, etc. The custom chips for such a modular system could make use of standard technologies and an array of computer programs such as MAGIC, SPICE, and their successors. These could eventually reduce the design time from years to months or even weeks.

Item 7 - - Special help for a very special case of modular electronics. One of the more difficult electrical/mechanical problems around involves the construction of the precision support, bussing, and cooling system for silicon trackers at colliding beam machines. To prevent distortion, the supports for the Mark II vertex detector did not over-constrain the detector, and so did not hold it as rigidly as possible. To reduce external forces, the electrical connections used highly compliant, flexible material. The fine lines and connections needed to carry all the control and data signals were surprisingly difficult to fabricate on this material. This is another place where the "DC on copper, (multiplexed) pulses in glass" approach might be useful for future detectors. One problem to be overcome in using copper wires for DC current supplies only is that silicon does not easily generate visible light so separate light sources must be used. In addition only a fraction of the energy used appears as light captured by the optical fiber. Removing the heat produced by the rest may be a problem, although light sources continue to become more efficient.

Techniques of the microelectronics industry may be helpful, even in solving purely mechanical problems. For example, the 6 - 9 cm long silicon strips of the Mark II detector had to be accurately aligned parallel to a common z axis. The assembly and encapsulation in epoxy required surgical precision and took more than one day per module. Better, faster techniques will be needed to complete the large silicon trackers proposed for future colliders. Designing for ease of fabrication, as well as for heat removal and positional stability will be needed.

Providing:

(1), intrinsic two dimensional readout is available, either from double sided strip detectors or from pixel detectors, and

(2), the chips are to be soldered on the substrate, help is at hand from an industry standard: the die bonder mentioned in section III L 1.

The small laboratory machine shown in Fig. 28 can pick, place, and bond a chip with ± 25 micron accuracy in less than 15 seconds. It may be possible, on a machine with optical feedback, to reach 5 micron accuracy. The design of such detectors should take into account, from the very start, assembly and testing as well as the needed performance.

Item 8 - - Space-point detectors

Another class of problems needs custom fabrication technologies. Silicon drift detectors are one example [147]. Another is the integration of detectors and circuits to make pixel devices that detect ionization in the bulk of the chip and guide it to an array of low capacity pickups on the surface. The potential benefits will be high spatial resolution, multiplexed, sparse field readout, and minimal thickness.

Space - point detectors --Pixel

While they are the current detectors of choice when high precision and good double track separation are required, strip detectors have a number of disadvantages:

1. Two sets of strips are necessary to recover the position information thrown out when charge is collected on a long strip.
2. Even with the two sets, n simultaneous hits within the overlap area of the two results in $n!$ possible solutions (the somewhat misnamed " n^2 ambiguity").
3. With intense beams, each strip has a high occupancy.
4. The large strip capacity produces small signals, and a relatively low signal-to-noise ratio.

These conditions are dramatically improved when strips are replaced with pixels. The first use of pixel detectors was by Damerell who showed CCDs could detect minimum ionizing particles [148]. They proved to have great advantages in pattern recognition, but were not widely used in other experiments due to their slow readout and thin sensitive region of only a few tens-of microns. These problems do not occur in the pixel detectors now under development which can have sparse field readout and in which nearly the full thickness of a chip is sensitive. The simplest uses a PIN diode with an array of collection electrodes bump-bonded to a readout chip [149].

Fully depleted CCDs have also been fabricated [150], including one version which uses features of drift chambers for readout [151]. The devices described in reference [150] collect ionization charge from the bulk and bring it to a potential minimum below the surface. Electrodes on the surface manipulate the location of the minimum to bring the charge to readout electrodes along the edge of the sensitive region. Reference [152] shows how field effect transistors can be placed on the surface so that the charge in the minimum affects their current flow. The ionization charge, staying in place until cleared, can be read out at any desired frequency set by noise and beam rate considerations, and does not get carried out to the edge of the chip on necessarily high capacity structures.

References [153] and [154] describe methods of fabricating transistors on high resistivity silicon. The proposed pixel devices, however, bring the ionization charge out to the edge of the chip and so have higher input capacity. Transistors are either confined to the array borders [153] or, if in the pixels

[154], do not have any mechanism to ensure that all of the ionization charge is guided to a collection electrode, rather than to other parts of the readout electronics such as the bottom side of transistor drains, sources, or channels.

The fabrication methods of [150], [153], and [154] use the low temperature planar technology pioneered by Kemmer [155]. High temperatures are avoided after the field oxidation, since they might drive in, by thermal diffusion, surface impurities that would ruin the detector diodes. However, the ion implants used to make the two electrodes can not be fully annealed, and there are high temperature steps needed later (such as further oxidation steps) for optimal transistor fabrication. S. Holland, in an important set of papers [156] has shown that with a backside getter of phosphorus implanted polysilicon, regular high temperature processes can be used to produce transistors simultaneously with low leakage diodes on high purity silicon.

Monolithic pixel detectors

Monolithic detectors, with readout circuits in each pixel were first fabricated in 1990-1991 [47], [157]. They are thinner and have (currently) five times lower input capacity than the lowest capacity bump-bonded devices.

Fig. 46 shows a schematic cross section of a monolithic detector. Ionization charge goes to an array of collection electrodes, each several microns across, which are in contact with the substrate. The electrodes are connected to the gates of readout transistors. These, and the rest of the pixel circuit are in a single implanted n well that surrounds each collection electrode and occupies the top few microns of the wafer. It provides an optimum environment for the transistors, shields the substrate from transistor switching transients, and, together with a diode covering the bottom of the chip, sets up the collection field which spans the wafer and focuses the charge on the small collection electrodes. The charge remains on the nearby MOS gates where it controls currents flowing to the edge of the chip. It can send current in more than one direction, can be read out at once, and can be stored and read out later as well. Signals from minimum ionizing particles in silicon strip detectors, before amplification, are about 0.01 to 0.1 mV; here they are about 0.05-0.1 V. Additional readout electronics are in n and p wells along the chip edges. The p wells are formed from the same p implant and drive in that (together with a p^+ contact implant) form the collection electrodes. Backside gettering similar to that in [156] is used.

Infrared tests show more than 99.995% of the positive ionization charge generated under the n well, away from the array edges, goes to the collection electrodes, rather than to the n well which occupies over 90% of the area. Fig. 47 shows two typical events from a test run in a 500 GeV/c muon beam, where one of the four planes used is tilted with its normal at 56 degrees to the beam direction. With a signal to single-channel noise ratio of 65 to 1 (with much of the noise due to off-chip electronics), each of the many struck pixels has a clear signal, even at this large angle. The accuracy as a function of pulse height for 90 degree tracks is shown in Fig. 48. Large pulse heights have a large error due to Landau

fluctuations, and small ones due to their reduced signal to noise ratio. The 63% in the middle have a spatial error in the 1.3 to 1.5 micron range.

Angled tracks have poorer resolution, typically in the 3 to 6 micron range, but, from the separation of their entrance and exit positions, can give a direction vector having an accuracy of 1 to 3 degrees (see Fig. 49), though usually with a two-fold pointing ambiguity. There can be exceptions. For example, in the bottom event of Fig. 47, charge spreading to an adjacent column from one end of the track, but not the other, indicates the former is near the face away from the collection electrode, since ionization charges coming from that face have longer collection times.

Fig. 50 shows how charge sharing near the border of a large pixel (in this case the long direction of a 34x125 micron pixel) can improve its position accuracy, although the errors here have significant non-Gaussian tails. If long pixels are used in a staggered "brick" like arrangement, charge spreading to adjacent rows provides three sharing regions in each length: two at the ends of each pixel and one in the middle next to the dividing line in the adjacent rows. The central non-sharing region is, in addition, split into two shorter ones, so the overall resolution in the long direction is greatly improved. Alternately, square pixels with sides of 60 - 70 microns or less would provide sets of space points with small errors in both directions and, for angled tracks, with a pointing vector attached uniquely to each point. This could provide significant help in tracking complex events.

These first-generation detectors have also been used to detect and measure a 0.1 mm wide calcification in a tissue sample from a human breast tumor, about half the diameter of the smallest ones normally seen in clinical practice [158].

Second generation detectors, now under test, have sparse-field readout, and a bottom-side diode structure that does not require guard rings. They are made with three fewer masking steps and over 50 fewer steps over all. A planar diode is fabricated and terminated around the device boundaries on the bottom with a vertical side-wall etch and then an oxidation and field implants of the etched surfaces. These diodes have similar electrical properties, and, covered with poly and/or tungsten during much of the fabrication and afterwards, are also scratch resistant.

Hybrid pixel detectors

The main advantage of hybrid detectors is that two known, commercial components are combined with a method that is commercially available, if not commonly so. Bumps have been made with solder, gold, tin, indium, and other materials. They have been made with ball bonders that do not leave the usual wire on the ball, by plating, and by thermal evaporation followed by lift-off above the masked regions around the bumps. Indium bumps, made this way by the Hughes Aircraft Company, were used to bond one of their readout chips to a detector made by Micron Semiconductor [149]. They have a great advantage: small indium bumps, 25 microns or less in diameter can be made and reliably bonded using commercial aligner bonders [159] without needing high temperatures. Bump bonding services are available from a number of sources which are now being checked by the Atlas collaboration [160].

A SLAC / UC Berkeley Space Sciences Lab / Hughes collaboration tested indium bumped hybrid detectors in a beam at Fermilab. An analysis by G. Lynch of the Lawrence Berkeley Lab showed a resolution of 2.6 microns [149]. Prototype data-push readout circuits which will not need to be triggered have been successfully fabricated and tested. Hybrid detectors also are used in Omega/WA97 and will be used in the LEP-Delphi experiment at CERN [161].

Several groups are now designing readout chips for use in the LHC which will uniquely identify tracks from one of perhaps 20 events in one beam crossing and also avoid confusion from the tracks of 80 or more following beam crossings before the arrival of a level 1 trigger. Due to the expected high rates, all keep the information and readout control within individual columns which can operate independently of each other prior to the level 1 trigger.

The CERN-Omega group plans to store hit information within the pixel while delays within each pixel, which can be set by an adjustable current, identify those hits with times corresponding to the level 1 trigger [161].

The Marseille group plans to shift the address of each hit pixel down its column, one shift per beam crossing, and calculate the hit time from the address and the arrival time at the column end [162]. They have also placed integrated busses for the readout chips on the long detector chips, each of which spans a group of readout chips.

The LBNL group, which first suggested the column readout scheme, plans to set a hit latch, storing a buffer address and analog charge information in each hit pixel [163]. The buffers, located at the end of each column, in turn only store the hit times. All three groups will use buffers at the column ends, and all will have sparse field readout of the hit information. All will have one form or another of dead time which is expected not to be too serious. (Radiation damage is expected to be the main limitation on how close they can be placed to the beam pipe).

A proposed quantum mammography system, using a detector with a higher Z than silicon, would not only be fully digital, with higher resolution and efficiency than present systems, but would count each x-ray quantum to produce the maximum possible information from the available radiation. Such chips would use a high-speed data-push readout [158].

Item 9 - - Light detectors

A VLSI circuit that could detect photons with high (at least 50%) efficiency would be of great importance for scintillating fiber trackers. Work is underway on devices with 1 to 8 channels [164], but the lack of grouped, fast timing outputs and sparse-field multiplexed outputs that would be available with IC technology makes the channel cost very expensive.

Item 10 - - High speed, radiation hard semiconductor detectors

Although the dominant trend of the VLSI industry is to shrink transistor sizes while keeping them and everything else well within the top part of the chip, a start is being made on a few uses of the remaining 99.9% of the wafer, other than just providing enough thickness to reduce the breakage in

fabrication. DRAMs have, for some time, used storage capacitors set edge-on in trenches. Research has started in placing transistors on silicon pillars, with the channel current running up or down and with the gate wrapped around the pillar [165]. The field of micromachining, in which the techniques of VLSI fabrication are used to make mechanical as well as electrical structures in silicon [166], now makes use of machines and techniques to fabricate structures with large vertical aspect ratios [167].

One possible application of this technology, now underway for high energy physics, is the development of solid-state PIN particle detectors in which the p and n electrodes are narrow columns that penetrate all or most of the chip thickness and which are separated by distances significantly smaller than that thickness [168]. Collection distances, times, and depletion voltages can drop by an order of magnitude. Resistance to the effects of bulk radiation damage can increase by similar amounts. This technology can also be used with the monolithic pixel structures described earlier. The collection electrodes of Fig. 46 now become columns. The superior collection power of these columns means n^+ electrodes can be used simultaneously with p^+ ones. Even though they would be run at nearly the same voltage as the n wells, most of the ionization electrons from the bulk would now go to the columns rather than to the wells.

VIII. Non-conclusion

This article started out as a requested review of custom VLSI circuits fabricated or under development for use in high energy physics, but quickly turned into an introduction and beginners guide to a rapidly changing field. The review, frozen at the 15 year point (1992), survived as Table 5, while the guide has expanded to include short, inadequate sections on deciding whether a custom circuit is really needed, a possible set of reference books in case you decide it is, the parts of an integrated circuit, their noise, the types of integrated circuits they can make, and how the whole thing is fabricated. It continues, briefly covering input protection, some design and layout considerations, radiation hardening, chip testing, assembly and packaging, technology trends, and several suggestions for the future. Hopefully some or all of this will help *you* finish the final, and important section covering the changes in the field since this section was written.

Glossary

Anneal A heat treatment to sufficient temperatures (at least about 400° C, but usually 800° to 1200° C) to produce changes in the bulk crystal structure. Two examples are the removal of radiation damage, and the electrical activation of implanted dopant atoms that occurs when they leave the interstitial positions where they stopped, and take up positions on the lattice.

BiCMOS A fabrication process that produces bipolar as well as complementary PMOS and NMOS transistors on the same monolithic substrate.

Channel stop implant In many fabrication technologies, adjacent transistors are separated by a thick field oxide that ties up otherwise mobile electrons on the silicon surface and in addition protects the substrate from impurities. To prevent inversion of underlying p type silicon by fixed positive charge in the oxide near the silicon boundary or by voltage levels that may be found on conductors running on top of the oxide, thus making it conducting, this region is normally implanted with boron.

Class 1000, 100, etc. A class xxx clean area will have fewer than xxx particles per cubic foot larger than 0.5 micron. For class 10, the size limit drops to 0.2 micron or larger. This is not the full story of course. For example, chemicals and water must be pure, and temperature and vibration must be controlled.

ECL Emitter coupled logic: a logic family using bipolar transistors with inputs connected to the bases and with the emitters of the unit connected to a common current source. It is one of the fastest logic families.

Epi Short for epitaxial (epi -- upon; taxial -- arrangement), a layer that is deposited on a substrate, frequently from the gas phase, and has the same crystal structure. For example, placing wafers in a furnace with silane gas and an admixture of dopant gas, will make specially doped layers of silicon.

Epitaxial See **Epi** immediately above.

Getter Part of a device that gets (and holds) impurities. Placing the getter on the backside -- the side away from the integrated circuits -- allows a large area to be used for the getter.

HEMT High electron mobility transistor.

Ion implantation A method of implanting dopant ions using an ion source, an accelerator, and an ion selection system. Control of both the dose and the depth profile is better than doping using thermal diffusion, but the latter can usually reach greater depths than the usual 100 - 200 KeV ion systems.

II Junction isolated. The isolation of a circuit element by surrounding it with a reverse-biased diode junction.

IFET Junction field effect transistor. A transistor in which the thickness of the conducting path between a source and drain, and hence its resistance, is controlled by reverse-biased diode junctions above and below it.

Latchup A condition in CMOS circuits that use reverse biased diodes to isolate the drains and sources from their substrates as well as the PMOS substrate from the NMOS one. They can form parasitic pnp and npn transistors in which positive feedback causes large, destructive currents to flow. See Fig. 7.

LHC An 8 x 8 TeV very high luminosity proton-proton collider planned to be built in the LEP tunnel at CERN. It should make life quite miserable for its detectors and their designers.

MAGIC A computer program used in the layout of integrated circuits. In addition to serving as a drawing tool, it can check that the design does not violate an installed set of design rules, can create an input list for SPICE, and can make an output file that can be used to control machines that make appropriate sets of masks for fabrication of the circuit.

MESFET Metal Semiconductor Field Effect Transistor. A field effect transistor in which the electric field set up by a metal gate electrode controls the carrier density in an underlying semiconductor, and hence its resistance. The gate is isolated from the semiconductor by a reverse biased Schottky diode.

MODFET MODulation Doped FET. A field effect transistor in which the channel consists of thin layers in which the doping, and so bandgap, changes significantly within the de Broglie wavelength of the charge carriers, producing novel properties including high drift velocities.

MOSIS MOS Implementation System. A service for providing shared-wafer fabrication runs at various foundries.

NTL Nonthreshold logic, a nonsaturated logic family using bipolar transistors, with inputs going to the bases, and with one common resistor connected to the emitters and others of similar value connected to the collectors. (See Journal of Solid State Circuits 10 (1975) 524.)

Passivation Rendering a surface nonconducting. A bare semiconductor surface would have many dangling bonds with loosely bound electrons that would readily move in the presence of an electric field. For silicon, these are normally tied up with oxygen, making SiO_2 . Additional roles of this glass layer are to protect the underlying single crystal silicon from impurities and mechanical damage and to provide an insulating substrate for interconnections.

Photoresist The basic photosensitive material that transfers a pattern to a wafer during integrated circuit fabrication. It is normally applied as a liquid to a slowly rotating wafer, which is then rapidly accelerated to a high speed for some predetermined time. The wafer is then baked to harden the resist, exposed, usually to ultraviolet light passing through a mask, and developed. The remaining resist is hard baked, used to protect the underlying semiconductor from some process such as etching or ion implantation, and finally stripped off the wafer.

PISCES A powerful computer program that is used to calculate electric field, voltage, current, and charge distributions in semiconductors. It often makes massive demands on computer memory, CPU cycles, and user tolerance.

Plasma etching Etching of a substance using a plasma. The ions normally react with the substance to form a gaseous product which diffuses away. If a voltage difference is established between the plasma and the substance, the ions will be accelerated to high speeds and can etch anisotropically.

Poly Polycrystalline silicon. By itself, it is a poor conductor due to carrier trapping at the grain boundaries, but when heavily doped so that all those traps are filled, it can be used as a conductor. It is also used to stop implanted ions, as a getter (particularly when doped with phosphorus or boron) to trap impurities, and to protect the underlying substrate from impurities.

Probe station A device to make electrical contact to interior nodes of integrated circuits. It has a platen to hold either unpackaged wafers or packaged chips, probes and their manipulators, and mechanisms to move both the probes and wafers under a microscope. Many other accessories are available including computer-controlled motion control, camera systems, closed circuit TV, and laser cutters.

Punch-through A condition in which the voltage difference between two like type semiconductors, separated by an opposite type, is sufficient to deplete the opposite type. Then field lines directly connect the like types, resulting in current flow between them.

Ring oscillator A sequence of inverters connected in a ring, with each output going to the next input. It is often used as a test structure to measure the chip speed.

Sample-and-hold, double correlated A circuit that saves the output of another one (usually an amplifier) at two different times and, on command, outputs a difference signal. Usually the two times are chosen so that the desired signal is present at only one of them, while presumably identical DC levels and switching transients are present at both and so are subtracted out.

Scratch mask An insulating layer placed on top of an integrated circuit to protect it from mechanical damage and degradation from ambient substances such as water, (which can corrode aluminum), and sodium (which can move under the influence of an electric field and cause long-term shifts in the electrical properties of the circuit).

SOI Silicon-on-insulator, a family of technologies in which islands of single-crystal silicon are fabricated on an insulating substrate. This substrate serves to isolate adjacent circuit elements, just as the reverse biased junctions around drain and source regions and the field oxide between them, do in junction-isolated technologies.

Sparse field readout Readout, usually of a large array, in which only those elements having a non-zero signal are read out.

SPICE Simulation Program with Integrated Circuit Emphasis, the first program designed to simulate integrated circuits. Both SPICE and the many commercial versions are still the main programs in use.

Spiking The short circuiting of a diode by a metal contact (usually aluminum). It occurs when silicon along a defect dissolves in the overlying metal during metal deposition. (The system is usually at an elevated temperature during this process.) The metal then replaces the dissolved silicon, forming a conducting spike through the junction.

SSC A 20 x 20 TeV high luminosity proton-proton collider, canceled by Congress when the machine was in the earliest construction phase. With its higher energy, it would not have needed as much

luminosity as the LHC to do equivalent or better physics research. This would have made the custom integrated circuits needed for its innermost detectors slightly less impossible to design.

Superscripts to indicate doping levels Lightly doped silicon (roughly with resistivities in the range of hundreds to thousands of ohm-cm) is designated as p - or n -, while heavily doped silicon (with resistivities in the milohm-cm range) is designated p⁺ or n⁺.

Superlattice A semiconducting structure in which the band gap varies significantly in distances comparable to the carrier thermal de Broglie wavelength, a distance which is only tens to hundreds of atoms thick.

SUPREM A computer program which can simulate the effects of many semiconductor fabrication steps including ion implantation, oxidation, and diffusion.

Surface passivation See "Passivation" above.

TTL Transistor-transistor logic. A versatile and highly developed logic family in which the input device is a multiple emitter transistor serving the multiple inputs. The transistors that follow saturate in operation and consequently have long storage times. TTL is obsolescent nowadays mostly because of speed limitations and large voltage swings.

Twin tubs In CMOS, both PMOS transistors, which need n type silicon channels and NMOS transistors, which need p type, are used. One method of fabricating them is to implant a properly doped region of n type silicon (called an n well) into a p type substrate, locally over-doping the p type material so it becomes n type. Another is to use p wells. The twin tubs process uses both type of wells. It requires more fabrication steps, but permits greater flexibility in tailoring the well properties.

Vias Conducting connections between different conducting layers of a circuit.

Additional References

- [1] J. Walker, S. Parker, B. Hyams, and S. Shapiro, "Development of High Density Readout for Silicon Strip Detectors", Nucl. Instr. and Meth. 226 (1984) 200.
- [2] G. Anzivino, et al., "First Results From a Silicon-Strip Detector with VLSI Readout", Nucl. Instr. and Meth. A243 (1986) 153.
- [3] to [37] See text III. A., page 5.
- [38] F. Fischer and F. Neppel, "Sputtered Ti-doped Al-Si for enhanced interconnect reliability", 22nd Annual Proceedings on Reliability Physics, (1984) 190.
- [39] N. Lu, L. Gerzberg, C. Lu, J. Meindl, "Modeling and Optimization of Monolithic Polycrystalline Silicon Resistors", IEEE Trans. Elec. Devices, ED 28 (1981) 818; N. Lu, L. Gerzberg, J. Meindl, "Scaling Limitations of Monolithic Polycrystalline-Silicon Resistors in VLSI Static RAMs and Logic", IEEE J. Solid-State Circuits, SC 17 (1982) 312; T. Ohzone, et al., "Ion-Implanted Thin Polycrystalline-Silicon High-Value Resistors for High-Density Poly-Load Static RAM Applications", IEEE Trans. Elec. Devices, ED 32 (1985) 1749; R. Saito, Y. Sawahata, and N. Momma, "A Novel Scaled-Down Oxygen-Implanted Polysilicon Resistor for Future Static RAM's", IEEE Trans. Elec. Devices, 35 (1988) 298.
- [40] R. Laxman, "Low ϵ Dielectrics: CVD fluorinated silicon dioxides", Semiconductor International, 18 (May 1995) 71; P. Singer, "Low k dielectrics: The search continues", Semiconductor International 19 (May 1996) 88; S. Murarka, "Low dielectric constant materials for interlayer dielectric applications", Solid State Tech., 39 (March 1996) 83; M. Murakami and S. Matsushita, "SiOF Films: The Problem in Logic Devices", Semiconductor International, 19 (July 1996) 291; P. Singer, "Low k nanoporous silica to be commercialized", Semiconductor International 19 (Nov. 1996) 60.
- [41] M. Bawa, E. Petro, and H. Grimes, "Fracture strength of large diameter silicon wafers", Semiconductor International, 18 (Nov. 1995) 115.

- [42] Bipolar transistor action had been discovered by Bardeen, Brattain, and Shockley in the course of trying to understand what had kept Shockley's original MOS transistors from working. When Andrew Grove, Robert Noyce, Gordon Moore and others formed Fairchild Semiconductor, and once again tried to make MOS transistors, they found their threshold voltages would change uncontrollably by tens and even hundreds of volts. After making, by chance, a stable device, they investigated, and found that the difference had been the use of electron beam evaporation to produce their metal. (An electron beam strikes a metal target and some of the metal vapor condenses on the wafers.) Earlier attempts had used tungsten filaments containing trace amounts of sodium. Evaporated ions contaminated their gate oxides, and, over a period of time, under the influence of electric fields, moved through it, with their Coulomb field shifting the externally required potential needed for turn on (Gordon Moore, Intel Corp., Santa Clara, CA, private communication, Oct. 1, 1990). See also E. Snow, A. Grove, B. Deal, and C. Sah, "Ion Transport Phenomena in Insulating Films", J. Applied Physics, 36 (1965) 1664.
- [43] A. Gnaedinger and H. Talley, "Quantum Mechanical Calculation of the Carrier Distribution and the Thickness of the Inversion Layer of a MOS Field-Effect Transistor", Solid-State Electronics, 13 (1970) 1301.
- [44] C. Jacoboni, C. Canali, G. Ottaviani, "A Review of Some Charge Transport Properties of Silicon", Solid-State Electronics, 20 (1977) 77-89; A. Sabnis and J. Clemens, "Characterization of the Electron Mobility in the Inverted <100> Si Surface, IEEE International Electron Devices Meeting, Technical Digest, (1979) 18; S. Sun and J. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces", IEEE Trans. on Electron Devices, ED-27 (1980) 1497; J. Watt and J. Plummer, "Universal Mobility-Field Curves for Electrons and Holes in MOS Inversion Layers", Symposium on VLSI Technology, Digest of Technical Papers (1987) 81; T. Nishida and C. Sah, "A Physically Based Mobility Model for MOSFET Numerical Simulation", IEEE Trans. on Electron Devices, ED-34 (1987) 310.
- [45] A. Jordan and N. Jordan, "Theory of Noise in Metal Oxide Semiconductor Devices", IEEE Trans. Electron. Devices ED-12 (1965) 148; F. Klaassen and J. Prins, "Thermal Noise of MOS

- Transistors", Philips Res. Repts 22 (1967) 505; F. Klaassen, "On the Influence of Hot Carrier Effects on the Thermal Noise of Field-Effect Transistors", IEEE Trans. Electron. Devices ED-17 (1970) 858.
- [46] A. Van der Ziel, "Noise in Solid State Devices and Circuits", Wiley, 1986; V. Radeka, "Low-noise techniques in detectors", Ann. Rev. Nucl. Part. Sci., 38 [1988] 217.
 - [47] S. Parker, "A Proposed VLSI Pixel Device for Particle Detection", Nucl. Instr. and Meth. A275 (1989) 494.
 - [48] R. Hofmann et al., "Development Of Readout Electronics For Monolithic Integration With Diode Strip Detectors", Nucl. Instr. and Meth. 226 (1984) 196.
 - [49] S. Kleinfelder, et al., "A Flexible 128 Channel Silicon Strip Detector Instrumentation Integrated Circuit With Sparse Data Readout", IEEE Trans. Nucl. Sci., NS-35 (1988) 171.
 - [50] T. Nashashibi and G. White, "A Low Noise FET With Integrated Charge Restoration For Radiation Detectors", IEEE Trans. Nucl. Sci., NS-37 (1990) 452.
 - [51] Antonio Alvarez, editor, "BiCMOS Technology and Applications", Kluwer Academic Publishers, 1989.
 - [52] For two exceptions to this, see Saburo Muroga, VLSI System Design, Wiley, 1982, p. 225. The technologies described there, DMOS and VMOS, are not frequently used, however.
 - [53] W. Holton and R. Cavin, III, "A Perspective on CMOS Technology Trends", Proc. IEEE, 74 (1986) 1646.
 - [54] "IC Photoresist Process is Dry", The Institute (IEEE) (May/June 1993) 1; T. Weidman and A. Joshi, "New photodefinable glass etch masks for entirely dry photolithography: Plasma deposited organosilicon hydride polymers", Appl. Phys. Lett. 62, (25 January 1993), 372.
 - [55] D. Antoniadis and R. Dutton, "Models for Computer Simulation of Complete IC Fabrication Process", IEEE Trans Electron. Dev. ED-26 (1979) 490; C. Ho, J. Plummer, S. Hansen, and R. Dutton, "VLSI Process Modeling--SUPREM III", IEEE Trans. on Electron Devices, ED-30 (1983) 1438.

- [56] A package of programs that simulates fabrication steps and can also predict device performance is described in A. Strojwas and S. Director, "VLSI: Linking Design and Manufacturing", IEEE Spectrum 25 (Oct. 1988) 24.
- [57] M. Pinto, C. Rafferty and R. Dutton, "Pisces II: Poisson and Continuity Equation Solver", Integrated Circuits Laboratory, Stanford University (1984); M. Pinto, C. Rafferty, H. Yeager and R. Dutton, "Pisces IIB Supplementary report", Integrated Circuits Laboratory, Stanford University, Stanford, California (1985).
- [58] Technology Modeling Associates, 595 Lawrence Expressway, Sunnyvale, CA, 94086, and Silvaco, 4701 Patrick Henry Drive, Santa Clara, CA 95054, sell such programs.
- [59] H. Jeong, et al., "The Future of Optical Lithography", Solid State Technology 37 (April 1994) 39; P. Burggraaf, "Stepping to Mix-and-Match I-line Lithography", Semiconductor International 18 (Feb. 1995) 47; P. Dunn, "X-ray's Future: A Cloudy Picture", Solid State Technology 37 (June 1994) 49; M. Levenson, "Extending Optical Lithography to the Gigabit Era", Solid State Technology 38 (Feb. 1995) 57. Besides the standard techniques of the use of large numerical aperture lenses and short wavelengths, some examples are the use: (1) of near-monochromatic lines to reduce chromatic aberrations, (2) of lenses corrected just along a linear region, not the whole surface, and then scanning that line across the reticle and mask, and (3) altering the pattern on the reticle so interference effects sharpen the image.
- [60] F. Lärmer, Patent DE 4241 045 C1 (Germany) (1944); S. Grigoropoulos, et al., "Highly Anisotropic Silicon Reactive Ion Etching for Nanofabrication Using Mixtures of SF_6 / CHF_3 Gasses", J. Vac. Sci. Technol. B 15(3) (May/June 1997) 640.
- [61] W. Carr and J. Mize, "MOS/LSI Design and Application", McGraw Hill, New York, (1972) 97. Reference [17], pp 295-301 also describes CMOS protection structures. Articles in the literature include: B. Unger, "Electrostatic Discharge Failures of Semiconductor Devices" Reliability Physics (1981) 193; A. Goel, "Procedure For Testing Electrostatic Discharge Susceptibility of MOS Devices", Reliability Physics (1981) 200; A. Berman, "Time-Zero Dielectric Reliability Test by a Ramp Method" Reliability Physics (1981) 204; M. Shatzkes and M. Av-Ron,

- "Statistics of Defect Related Breakdown", Reliability Physics (1981) 210; D. Fisch, "A New Technique for Input Protection Testing" Reliability Physics (1981) 212; R. Antinone, "How to Prevent Circuit Zapping", IEEE Spectrum 24 (April 1987) 34; C. Duvvery, R. Rountree, and O. Adams, "Internal Chip ESD Phenomena Beyond the Protection Circuit", Reliability Physics, (1988) 19; S. Aur, A. Chatterjee, and T. Polgreen, "Hot Electron Reliability and ESD Latent Damage", Reliability Physics, (1988) 15; W. Boxleitner, "How to Defeat Electrostatic Discharge" IEEE Spectrum 26 (Aug. 1989) 36 (emphasis mostly on systems, not ICs).
- [62] The MOSIS Service, 4676 Admiralty Way, Marina del Rey, CA 90292-6695; www.isi.edu/mosis.
- [63] See www.imec.be (Interuniversitair Micro-Elektronica Centrum, Kapeldreef 75, 3030 Leuven, Belgium), and www.imec.be/europractice/europractice/html.
- [64] R. T. Maniwa, "Focus report: foundries", Integrated System Design, 9 (Feb. 1997) 30. See also [34], [65], and [80]; "Focus report - ASIC foundry services, survival in a submicron seller's market", ASIC & EDA, (Nov. 1992) 20; "Focus report - the foundry game", ASIC & EDA, 1994 Vendor Guide (Dec. 1993) 20; "ASIC Foundries", ASIC & EDA (Dec/Jan 1995) 44 (EDA, translated from Acronym to English is Electronics Design Automation.); "ASIC Foundries", Integrated System Design, Vendor Guide 1996 Supplement (Jan. 1996) 96. Still earlier surveys can be found in the now discontinued publications, VLSI Systems Design and the earlier VLSI Design, published by CMP Publications, 600 Community Drive, Manhasset, NY 11030.
- [65] Tom Zimmerman, "The use of low resistivity substrates for optimal noise reduction, referencing, and current conduction in mixed signal ASICs", Fermilab Report TM 2035, Nov. 1997; R. Gharpurey and R. Meyer, "Modeling and Analysis of Substrate Coupling in Integrated Circuits", IEEE J. Solid-State Circuits, 31 (1996) 344-353.
- [66] L. Manchanda et al., "A High-Performance Directly Insertable Self-Aligned Ultra-Radiation-Hard and Enhanced Isolation Field-Oxide Technology for Gigahertz Si-NMOS/CMOS VLSI", IEEE Electron Device Letters 10 (1989) 17; L. Manchanda et al., IEEE Trans. on Electron Devices 36 (1989) 651.

- [67] M. Tanaka, H. Ikeda, M. Ikeda, and S. Inaba, "Development of Monolithic Time-to-Amplitude Converter for High Precision TOF Measurement", Conference Record of the 1990 IEEE Nuclear Science Symposium, p558.
- [68] W. Scott, et al., "1986 VLSI Tools; Still More Works by the Original Artists," UCB/CSD Technical report No. 86/272, University of California at Berkeley, Dept of Electrical Engineering and Computer Science, Berkeley CA 94720; Stuart Kleinfelder, "Custom MOS VLSI Design: Tools and Methods", IEEE Trans. Nucl. Sci., NS-36 (1989) 484.
- [69] L. Nagel, "SPICE2: A computer program to simulate semiconductor circuits", Memorandum No. ERL-M520, May 9, 1975, Electronics Research Laboratory, College of Engineering, Univ. of Calif., Berkeley, CA 94720; T. Perry, "Donald O. Pederson", IEEE Spectrum 35 (June 1998) 484.
- [70] Two of the many are: Andrei Vladimirescu, "The SPICE Book", John Wiley & Sons, New York (1994); Ron Kielkowski, "Inside SPICE", McGraw-Hill, New York (1994).
- [71] P. Tuinenga, "Spice - A guide to circuit simulation and analysis using PSpice", Prentice Hall, Englewood Cliffs, N.J. (1988).
- [72] "Saber reference manual", Analogy Inc., P.O.Box 1669, Beaverton, OR, 97075, www.analogy.com/default.htm.
- [73] G. Massobrio and P. Antognetti, "Semiconductor device modeling with SPICE", 2nd ed., McGraw-Hill, Inc., New York, 1993.
- [74] D. Binkley and M. Casey, "Performance of Fast Monolithic ECL Voltage Comparators in Constant-Fraction Discriminators and Timing Circuits", IEEE Trans. Nucl. Sci., NS-35 (1988)226.
- [75] J. Read, Ed., "Gate Arrays: Design Techniques and Applications", McGraw Hill New York (1985); N. Okuda, M. Sugai, and N. Goto, "Semicustom and Custom LSI Technology", Proc. IEEE, 74 (1986) 1636.
- [76] G. Brandenburg, U. Gorschutz, K. Muller, "Poor Man's Gate Arrays: Logic Cell Arrays", IEEE Trans. Nucl. Sci., NS-35 (1988) 213.
- [77] J. Villasenor and W. Mangione-Smith, "Configurable Computing", Scientific American, 276 (June 1997) 67.

- [78] P. Chatterjee, "The 1990s Take Shape", IEEE Spectrum 26 (January 1989) 48.
- [79] S. Kleinfelder, "Custom MOS VLSI Design: Tools and Methods", IEEE Trans. Nucl. Sci. NS-36 (1989) 484.
- [80] Integrated System Design, Vendor Guide 1998 Supplement (January 1998).
- [81] R. Sparkes and W. Gross, "Recent Developments and Trends in Bipolar Analog Arrays", Proc. of the IEEE, 75 (1987) 807; D. Dorfan, "Semi-Custom Bipolar Design", IEEE Trans. Nucl. Sci., NS-36 (1989) 493. See also D. Christian, M. Haldeman, R. Yarema, and T. Zimmerman, "The Development of Two ASIC'S for a Fast Silicon Strip Detector Readout System", IEEE Trans. Nucl. Sci., NS-36 (1989) 507.
- [82] This is an active field with many publications. Two recent ones with references to others are K. Gill, G. Hall, and B. MacEvoy, "Bulk damage effects in irradiated silicon detectors due to clustered divacancies", J. Appl. Phys. 82 (1997) 126, and R. Wunstorf, "Radiation hardness of silicon detectors: current status", IEEE Trans. on Nucl. Sci. 44 (1997) 806.
- [83] N. Saks and R. Rendell, "The time-dependence of post-irradiation interface trap build-up in deuterium-annealed oxides", IEEE Trans. Nucl. Sci. NS-39 (1992) 2220. The motion of protons from one oxide surface to the other has actually been proposed as the basis for a new type of memory in K. Vanheusden, et al., "Non-volatile memory device based on mobile protons in SiO₂ thin films", Nature 386 (10 April, 1997) 587.
- [84] J. Aitken, D. DiMaria, and D. Young, IEEE Trans. Nucl. Sci., NS-23 (1976) 1526; J. Peel and R. Eden, IEEE Trans. Nucl. Sci., NS-18 (Dec., 1971) 84; S. Chang and S. Lyon, "Location of Positive Charge Trapped Near the Si-SiO₂ Interface at Low Temperature", Appl. Phys. Lett. 48 (1986) 136.
- [85] W. Warren, et al., "Microscopic Nature of Border Traps in MOS Oxides", IEEE Trans. on Nucl. Sci., NS 41 (1994) 1817; A. Lelis and T. Oldham, "Time Dependence of Switching Oxide Traps", IEEE-NS 41 (1994) 1835.
- [86] M. Chiwaki and T. Tomimasu, "The Effects of Radiation Damage on the Noise Performance of FETs", IEEE Trans. Nucl. Sci., NS-22 (1975) 2696.

- [87] R. Yarema et al., "Radiation Effects on the SVX Silicon Strip Readout Chip", IEEE Trans. Nucl. Sci., NS-37 (1990) 434.
- [88] G. Dunn, R. Jayaraman, W. Yang, and C. Sodini, "Radiation Effects in Low-pressure Reoxidized Nitrided Oxide Gate Dielectrics", Appl. Phys. Lett. 52 (1988) 1713.
- [89] P. Dressendorfer, Sandia National Laboratories, Albuquerque, NM, private communication.
- [90] J. Schroeder, et al., "An Advanced, Radiation Hardened Bulk CMOS / LSI Technology", IEEE Trans. on Nucl. Sci. NS-28 (1981) 4033.
- [91] C. Dozier and D. Brown, "Effect of Photon Energy on the Response of MOS Devices", IEEE Trans. Nucl. Sci., NS-28 (1981) 4137 and "The Use of Low Energy X-Rays for Device Testing - A Comparison with ^{60}Co Radiation", IEEE Trans. Nucl. Sci., NS-30 (1983) 4382; E. Stassinopoulos et al., "The Damage Equivalence of Electrons, Protons, Alphas, and Gamma Rays in Rad-Hard MOS Devices", IEEE Trans. Nucl. Sci., NS-30 (1983) 4363; T. Oldham and J. McGarrity, "Comparison of ^{60}Co Response and 10 KeV x-ray Response in MOS Capacitors", IEEE Trans. Nucl. Sci., NS-30 (1983) 4377; D. Fleetwood et al., "Accounting for Dose-Enhancement Effects with CMOS Transistors", IEEE Trans. Nucl. Sci., NS-32 (1985) 4369; R. Hamm, "Dose Calculations for Si-SiO₂Si Layered Structures Irradiated by x-rays and ^{60}Co Gamma Rays", IEEE Trans. Nucl. Sci., NS-33 (1986) 1236; and J. Benedetto and H. Boesch, Jr. "The Relationship Between ^{60}Co and 10-KeV X-ray Damage in MOS Devices", IEEE Trans. Nucl. Sci., NS-33 (1986) 1318.
- [92] M. Shaneyfelt, et al., "Hardness Variability in Commercial Technologies", IEEE-NS 41 (1994) 2536.
- [93] J. Raymond and E. Peterson, "Comparison of Neutron, Proton and Gamma Ray Effects in Semiconductor Devices", IEEE Trans. Nucl. Sci., NS-34 (1987) 1622.
- [94] "6th Annual Directory of Radiation Tolerant ICs", Aerospace and Defense Science, (Summer 1991), Indialantic, FL.

- [95] "Technical Proposal Backup Document for the Semiconductor Tracker (SCT) for the Atlas Detector at the Large Hadron Collider at CERN", Report Atlas Indet - NO - 085, 25 Jan, 1995, pp 58-74; Atlas Inner Detector - Technical Design Report 2 (30 April 1997) 323-327.
- [96] R. Pease, "Troubleshooting Analog Circuits", Butterworth-Heinemann, Boston 1991.
- [97] E. Murphy, "Design Automation", IEEE Spectrum 26 (Jan. 1989) 34; W. Holton and R. Cavin, III, "A Perspective on CMOS Technology Trends", Proc. IEEE 74 (Dec. 1986) 1646; T. Williams and K. Parker, "Design for Testability -- A Survey", Proc. IEEE 71 (Jan. 1983) 98.
- [98] Greg Boll, GGB Industries, PO Box 10958, Naples, Florida 34104, private communication. See also www.pico.probe.com.
- [99] R. Waters and J. Logan, "Methods of Trace Cutting For Diagnostic Probing of Semiconductor Devices", Proc. of the 1984 Int. Symposium For Testing and Failure Analysis, (1984) 136. Examples of companies making cutting lasers are Alessi Industries, Irvine CA 92718 and Florod Corp., Gardena, CA, 90247.
- [100] T. Cambria and N. Economou, "Mask and Circuit Repair with Focused-Ion Beams", Solid State Technology 30 (September, 1987) 133; Y. Nakagawa, et al., "Recent Progress On Etching Technology With FIB In Photomask Repair", Proc. of the SPIE, 923 (1988) 114; J. Glanville, "Focused Ion Beam Technology for Integrated Circuit Modification", Solid State Technology 32 (May 1989) 270; K. Nikawa, et al., "New Applications Of Focused Ion Beam Technique To Failure Analysis And Process Monitoring Of VLSI", 27th International Reliability Physics Symposium 27 (1989) 43; J. Bond, "E-Beams Test, Ion Beams Repair", Test and Measurement World 11 (May 1991) 59.
- [101] FEI Company, 7451 NE Evergreen Parkway, Hillsboro, OR 97124.
- [102] "Reliability Physics", proceedings of the annual International Reliability Physics Symposium, sponsored by the IEEE Electron Devices Society and the IEEE Reliability Society.
- [103] D. Kerr et al., "Stabilization of SiO₂ Passivation Layers with P₂O₅", IBM J. Res. and Dev., 8 (1964) 376.

- [104] D. Peck, "Practical Applications of Accelerated Testing -- Introduction", Reliability Physics 13 (1975) 253.
- [105] IEEE Trans. Components, Hybrids, Manuf. Technol., which carried packaging articles, has been split into two parts, one of which is now Part B, "Advanced Packaging". The ASME now publishes the "Journal of Electronic Packaging". Also see "Proc. Int. Electron. Packaging Society".
- [106] W. Grobman, "High-end Computer Packaging--VLSI Scaling and Materials Science", J. Vac. Sci. Technol. A3 (May/June 1985) 725; A. Blodgett, Jr., "Microelectronic Packaging", Scientific American, 249 (July 1983) 86.
- [107] J. Vardaman, "Chip-scale Packages: Cleared for Takeoff", Chip Scale Review (published by Tessera, 3099 Orchard Dr., San Jose, CA 95134) 1 (May 1997) 8; P. Thompson, "Chip Scale Packaging", IEEE Spectrum 34 (Aug 1997) 36.
- [108] Don Moore, "Sticky issues with semiconductor processing tape", Semiconductor International 19 (May, 1996) 139.
- [109] G. Harman, "Reliability and Yield Problems of Wire Bonding in Microelectronics", International Society For Hybrid Microelectronics, Reston VA 22090 (1989); L. Levine, "The Ultrasonic Wedge Bonding Mechanism: Two Theories Converge", SPIE 2649 (1995) 242.
- [110] John Lau (editor), "Ball Grid Array Technology", McGraw-Hill, New York (1995).
- [111] John Lau (editor), "Flip Chip Technology", McGraw Hill, New York (1995).
- [112] W. Vilkelis, "Lead Reduction Among Combinatorial Logic Circuits", IBM J. Res. Develop., 26 (1982) 342.
- [113] I. Milosevic, et al., "Polyimide Enables High Lead Count TAB", Semiconductor Int. 11 (Oct. 1988) 122.
- [114] P. Totta and R. Sopher, "SLT Device Metallurgy and its Monolithic Extension", IBM J. Res. Dev., 13 (1969) 226, and following articles. Most of the issue is devoted to articles on solder bump technology. Articles in IBM J. Res. Dev., 26 (1982) numbers 1 and 3 and D. Seraphim and

- I. Feinberg, "Electronic Packaging Evolution in IBM", IBM J. Res. Develop., 25 (1981) 617 cover similar packaging technology used for more advanced computers.
- [115] G. Kramer, et al., "Silicon PIN diode hybrid arrays for charged particle detection: building blocks for vertex detectors at the SSC", 1989 International Industrial Symposium on the Super Collider, New Orleans, LA, (Feb. 8-10, 1989). Also issued as SLAC Pub 4942. The Hughes group is not now active in making these detectors.
 - [116] Tanaka Kikinzoku Kogyo K.K. Overseas Trade Dept., 2-6-6 Nihonbashi Kayabacho, Chuo-ku, Tokyo 103, Japan.
 - [117] M. Levine, "Solid State Cooling with Thermoelectrics", Electronic Packaging and Production, November 1989, 74. W. O. Miller, et al., "The Super-Conducting Supercollider Silicon Tracking Subsystem: Research and Development", Report LA-12029 (Dec. 1990) 50-68; W. O. Miller, et al., "Mechanical Engineering and Design of Silicon Tracking Devices", Symposium on Detector Research and Development for the Super-Conducting Super Collider, Ft. Worth, TX, World Scientific, Singapore (1990) 739.
 - [118] L. Gallace and M. Rosenfield, "Reliability of Plastic-Encapsulated Integrated Circuits in Moisture Environments", RCA Review 45 (June 1984) 249; K. Fitzgerald, et al. "Microelectronic ghost story", IEEE Spectrum 25 (Dec. 1988) 13; D. Dahringer, "Some Aspects of Adhesive-Induced Corrosion in Electronic Devices", Proc. 11th Electrical Insulation Conf., IEEE (1973) 333; M. Kahn, et al. "Effect of High Thermal Stability Mold Material on the Gold-Aluminum Bond Reliability in Epoxy Encapsulated VLSI Devices", Reliability Physics 26 (1988) 40.
 - [119] R. Browning, I. Lincoln and P. Stonestrom, "Recent Advances in Automated Patterned Wafer Inspection", SPIE Integrated Circuit Metrology, Inspection, and Process Control III, 1087 (1989) 440.
 - [120] R. Hughes and J. Shott, "The Future of Automation for High-Volume Wafer Fabrication and ASIC Manufacturing", Proc. IEEE 74 (Dec. 1986) 1775.
 - [121] P. Bareyre et al., "The Use of a New PMOS Monolithic Integrated Circuit for the Electronic Equipment of a Large Multi-wire Proportional Chamber (MWPC) Detection System", Nucl.

- Instr. and Meth. 131 (1975) 583; P. Bareyre, et al., "A New Monolithic Integrated Circuit for Multiwire Proportional Chamber (MWPC) Read-out System", IEEE Trans. on Nucl. Sci. NS 23 (1976) 274; E. Platner, "Programmable Combinational Logic Trigger System for High Energy Particle Physics Experiments", IEEE Trans. on Nucl. Sci. NS 24 (1977) 225; E. Platner, et al., "Programmable Combinational Logic Trigger System for High Energy Particle Physics Experiments", Nucl. Instr. and Meth. 140 (1977) 549.
- [122] R. Hofmann, et al., "Development of Readout Electronics for Monolithic Integration with Diode Strip Detectors", Nucl. Instr. and Meth. 226 (1984) 196.
- [123] H. Williams, "Fast Analog Integrated Circuits for Particle Physics", IEEE Trans. Nucl. Sci. 35 (1988) 146.
- [124] W. Panofsky, "From Accelerators To Storage Rings To ?", IEEE Trans. Nucl. Sci., NS-30 (1983) 3689.
- [125] Early microprocessor data in Fig. 33 is from S. Xenakis, Intel Corp., Santa Clara, CA 95052, and Marty Levy and Roy Druian, Motorola Corp., Austin, TX, private communications. DRAM data is from F. Carrubba, Hewlett Packard Corp., Palo Alto, CA, private communication. These same sources provided information on the increasing number of transistors per chip, Fig. 38. Unlabeled points in Fig. 33 are from W. Sansen, "Technology Tradeoffs in High-Frequency Detector Integrated Circuits", presented at the Workshop on Silicon Pixel Detectors, IMEC, Leuven, Belgium, (1988), unpublished.
- [126] R. Dennard, et al., "Design of Ion-Implanted MOSFETs with Very Small Physical Dimensions", IEEE J. of Solid-State Circuits, SC-9 (1974) 256.
- [127] T. Enomoto, T. Ishihara, M. Yasumoto, and T. Aizawa, "Design, Fabrication, and Performance of Scaled Analog IC's", IEEE J. of Solid-State Circuits, SC-18 (1983) 395.
- [128] G. Declerck, "Trends in VLSI Processing", Nucl. Instr. and Meth., A275 (1989) 475.
- [129] G. Myers, A. Yu, and D. House, "Microprocessor Technology Trends", Proc. of the IEEE, 74 (1986) 1605.

- [130] R. Keyes, "Miniaturization of Electronics and Its Limits", IBM J. Research and Development, 32 (1988) 24, "The Future of the Transistor", Scientific American (June, 1993) 70.
- [131] Scientific American, June 1990, p 26. For an general introductory article see H. Wickramasinghe, "Scanned-Probe Microscopes", Scientific American (October 1989) 98. An early use of this technology for electronics will likely be in the study, on an atomic scale, of surface properties in integrated circuit structures.
- [132] R. Carlson and C. Neugebauer, "Future Trends in Wafer Scale Integration", Proc. of the IEEE, 74 (1986) 1741.
- [133] T. Ning and D. Tang, "Bipolar Trends", Proc. of the IEEE, 74 (1986) 1669. Fig. 39 has been adapted from Fig. 1 of this reference. One point has been added and two points corrected. Points marked (o) are from E. Rymaszewski et al., "Semiconductor Logic Technology at IBM", IBM J. Res. Develop. 25 (1981) 603; (solid triangle) from M. Usami et al., "Status and Prospects for Bipolar ECL Gate Arrays", Proc. IEEE Int. Conf. on Computer Design, Dig. Tech. Papers (1983) 272; (∇) from S. Lee and A. Bass, A 2500-gate Bipolar Macro Cell Array With 250 ps Gate Delay", Int. Solid States Circuits Conf. Dig. Tech Papers (1982) 178; (solid square) from E. Gonauser et al., "A Bipolar 230 ps Masterslice Cell Array with 2600 Gates", IEEE J. Solid State Circuits SC-19 (1984) 299; (o) S. Brenner et al., "A 10,000 Gate Bipolar VLSI Masterslice using Four Levels of Metal", IEEE International Solid-State Circuits Conf. Dig. Tech. Papers (1983) 152.
- [134] S. Konaka, et al., "A 30 ps Si Bipolar IC Using Super Self-Aligned Process Technology", IEEE Trans. on Elec. Dev. ED-33 (1986) 526; S. Konaka, M. Ugajin, T. Matsuda, "Deep submicrometer super self-aligned Si bipolar technology with 25.4 ps ECL", IEEE Trans. on Elec. Dev. ED-41 (1994) 44.
- [135] P. Van Halen and D. Pulfrey, "High-Gain Bipolar Transistors with Polysilicon Tunnel Junction Emitter Contacts", IEEE Trans. on Electron Devices, ED-32 (1985) 1307.
- [136] H.Morkoc and P. Solomon, "The HEMT: a superfast transistor", IEEE Spectrum, 21 (Feb., 1984) 28.

- [137] Gordon Moore, "VLSI: some fundamental challenges", IEEE Spectrum, 16 April (1979) 30.
- [138] C. Fey and D. Paraskevopoulos, "A Techno-Economic Assessment of Application-Specific Integrated Circuits: Current Status and Future Trends", Proc. IEEE 75 (June 1987) 829.
- [139] Gordon Moore, Intel Corporation, Santa Clara, CA, private communication, August 29, 1990.
- [140] R. Fabian Pease, Department of Electrical Engineering, Stanford Univ., Stanford CA 94305, private communication.
- [141] M. Brodsky, "Progress in Gallium Arsenide Semiconductors", Scientific American, 262 (Feb. 1990) 68.
- [142] G. Doehler, "Solid-State Superlattices", Scientific American, 249 (Nov. 1983) 144.
- [143] F. Capasso, "Band-Gap Engineering for New Photonic and Electronic Devices", Nucl. Instr. and Meth., A265 (1988) 112.
- [144] P. Manfredi, S. Rescia, and V. Speziali, "Improving the Performances of Silicon Detector Electronics in High Energy Experiments by Optoelectronic Transmission of Analog Signals Between Beam Line and Counting Room", IEEE Trans. Nucl. Sci. NS-32 (1985) 640, (two earlier references are cited in this paper); E. Cisneros and G. Burgueno, "Analog Data Transmission Via Fiber Optics", IEEE Trans. Nucl. Sci. NS-34 (1987) 175; R. Skegg, "A High Speed Fiber Optic Serial Networking System for Data Acquisition and Control", IEEE Trans. Nucl. Sci. NS-37 (1990) 293; M. Wong, et al., "Use of Fiber-Optics for Real-Time Control and Data Acquisition", IEEE Trans. Nucl. Sci. NS-37 (1990) 336; B. Leskovar, "Optical Data Transmission at the Superconducting Super Collider", IEEE Trans. Nucl. Sci. NS-37 (1990) 271.
- [145] Mark Lowry, Lawrence Livermore Lab, Livermore, CA, private communication.
- [146] After this part was written, an email was distributed announcing the formation of the VME International Physics Association, whose goal is to work with an industry group, the VMEbus International Trade Association, to extend the VME standard to meet the needs of the physics community in the areas of data acquisition, instrumentation, and controls. Further information can be found at www-ese.fnal.gov/vipa/ and www.vita.com/.

- [147] E. Gatti and P. Rehak, "Semiconductor Drift Chamber--An Application of a Novel Charge Transport Scheme", Nucl. Instr. and Meth., 225 (1984) 608; E. Gatti, P. Rehak, and J. Walton, "Silicon Drift Chambers - First Results and Optimum Processing of Signals", Nucl. Instr. and Meth. 226 (1984) 129; P. Rehak et al., "Semiconductor Drift Chambers for Position and Energy Measurements", Nucl. Instr. and Meth. A235 (1985) 224; P. Rehak, et al., "Progress in Semiconductor Drift Detectors", Nucl. Instr. and Meth. A248 (1986) 367. The first two references also describe a method for making fully depleted CCDs.
- [148] M. Algranati, et al., "Use of Charge Transfer Device for Particle Detection", Nucl. Instr. and Meth. 164 (1979) 615; C. J. S. Damerell, et al., "Charge-coupled Devices for Particle Detection with High Spatial Resolution", Nucl. Instr. and Meth. 185 (1981) 33; C. J. S. Damerell, "CCD Vertex Detectors in Particle Physics", Nucl. Instr. and Meth. A342 (1994) 78; C. J. S. Damerell et al., "Design and Performance of the SLD Vertex Detector, a 120 Mpixel Tracking System", Proc. XXVI Int. Conf. on High Energy Physics, Dallas 2 (1992) 1862; C. J. S. Damerell, "Vertex Detectors at Linear Colliders (Present and Future)", RAL-94-096 (Sept. 1994).
- [149] S. Gaalema, "Low noise random-access readout technique for large PIN detector arrays", IEEE Trans. Nucl. Sci. NS-32 (1985) 417; S. Shapiro, et al., "Silicon PIN diode array hybrids for charged particle detection", Nucl. Instr. and Meth. A275 (1988) 580; S. Shapiro, et al., "Progress on the design of a data push architecture for an array of optimized time tagging pixels", SLAC-Pub-6249, June 1993; see also references [161], [162], and [163].
- [150] J. Kemmer and G. Lutz, "New Detector Concepts", Nucl. Instr. and Meth. A253 (1987) 365; L. Struder, et al., "Device Modeling of Fully Depletable CCDs", Nucl. Instr. and Meth. A253 (1987) 386; L. Struder, et al., "Development of Fully Depletable CCDs for High Energy Physics Applications", Nucl. Instr. and Meth., A257 (1987) 594; H. Soltau, et al., "Performance of the pn-CCD x-ray Detector System Designed for the XMM Satellite Mission", Nucl. Instr. and Meth., A377 (1996) 340. See also the first two references in [147].
- [151] A. Castoldi, et al., "Conception and design criteria of a novel silicon device for the measurement of position and energy of x rays", IEEE Trans. on Nuclear Science, 44 (1997) 1724.

- [152] J. Kemmer and G. Lutz, "New Detector Concepts", Nucl. Instr. and Meth., A253 (1987) 365; J. Kemmer et al., "Experimental Confirmation of a New Semiconductor Detector Principle", Nucl. Instr. and Meth. A288 (1990) 92; P. Rehak, et al., "Feedback Charge Amplifier Integrated on the Detector Wafer", Nucl. Instr. and Meth. A288 (1990) 168; P. Klein, et al., "Design and Performance of Semiconductor Detectors with Integrated Amplification and Charge Storage Capability", Nucl. Instr. and Meth, A305 (1991) 517.
- [153] S. Audet and E. Schooneveld, "High-purity Silicon Radiation-sensor Array", Nucl. Instr. and Meth. A275 (1989) 517; S. S. Wouters, S. Audet, M. Kim and E. Schooneveld, "Integrating Transistors on High-Ohmic Silicon", Nucl. Instr. and Meth. A288 (1990) 44.
- [154] G. Vanstraelen, I. Debusschere, C. Claeys and G. Declerck, "New Concepts for Integrated Solid State Detector Electronics", Nucl. Instr. and Meth. A273 (1988) 625; G. Vanstraelen, I. Debusschere, C. Claeys and G. Declerck, "Fully Integrated CMOS Pixel Detector for High Energy Particles", Nucl. Instr. and Meth. A275 (1989) 574.
- [155] J. Kemmer, "Fabrication of Low Noise Silicon Radiation Detectors in the Planar Process", Nucl. Instr. and Meth. 169 (1980) 499; "Development of Detector Fabrication by the Planar Process", Nucl. Instr. and Meth. 226 (1984) 89.
- [156] S. Holland, "An IC-compatible Detector Process", IEEE Trans. Nucl. Sci., NS-36 (1989) 283; S. Holland, "Fabrication of Detectors and Transistors on High-Resistivity Silicon", Nucl. Instr. and Meth., A275 (1989) 537.
- [157] C. Kenney, et al., "A prototype monolithic pixel detector", Nucl. Instr. and Meth. A342 (1994) 59; W. Snoeys, et al., "PIN detector arrays and integrated readout circuitry on high-resistivity float-zone silicon", IEEE Trans. on Elec. Devices 41 (1994) 903. Earlier papers are listed in these two references.
- [158] S. Parker, et al., "Breast cancer calcification measurements using direct x-ray detection in a monolithic silicon pixel detector", IEEE Trans. on Nucl. Sci. 41 (1994) 2862.
- [159] Research Devices, 121 Ethel Road West, Piscataway, N.J. 08854.
- [160] Atlas Inner Detector Technical Design Report, 2 (30 April 1997) 340-342.

- [161] M. Campbell, et al., "Development of a pixel readout chip compatible with large area coverage", Nucl. Instr. and Meth. A342 (1994) 52; E. Heijne, et al., "LHC1: A semiconductor pixel detector readout chip with internal tunable delay providing a binary pattern of selected events", Nucl. Instr. and Meth. A383 (1996) 55-63.
- [162] P. Delpierre, et al., "Large scale pixel detectors for DELPHI at LEP200 and ATLAS at LHC", Nucl. Instr. and Meth. A342 (1994) 233.
- [163] M. Wright, J. Millaud, and D. Nygren, "A pixel unit-cell targeting 16ns resolution and radiation hardness in a column read-out particle vertex detector", Nucl. Phys. B (Proc. Suppl.) 32 (1993) 468.
- [164] P. Webb, R. McIntyre, and J. Conradi, "Properties of Avalanche Photodiodes", RCA Review 35 (1974) 234; M. Salomon and J. Kitching, "Preliminary Results on the Use of Avalanche Photodiodes for Scintillating Fiber Readout", IEEE Trans. Nucl. Sci., NS-38 (1991) 140; M. Petroff, M. Stapelbroek, and W. Kleinhans, "Detection of Individual 0.4-28 Micron Wavelength Photons Via Impurity-Impact Ionization in a Solid-State Photomultiplier", Appl. Phys. Lett. 51 (1987) 406; M. Petroff and M. Stapelbroek, "Photon-Counting Solid-State Photomultiplier", IEEE Trans. Nucl. Sci., NS-36 (1989) 158; M. Petroff and M. Atac, "High Energy Particle Tracking Using Scintillating Fibers and Solid State Photomultipliers", IEEE Trans. on Nucl. Sci., 36 (1989) 163; M. Atac et al., "Scintillating fiber tracking for the SSC using visible light photon counters", Nucl. Instr. and Meth. A314 (1992) 56; M. Atac et al., "Tracking with scintillating fibers and visible light photon counters" Nucl. Instr. and Meth. A320 (1992) 155.
- [165] H. Takato, et al., "High performance CMOS surrounding gate transistor (SGT) for ultra high density LSIs", IEDM (1988) 222; H. Pein and J. D. Plummer, "A 3-D sidewall flash EPROM cell and memory array", IEEE Electron Device Letters, 14 (1993) 415; S. Watanabe, et al., "A Novel Circuit Technology with Surrounding Gate Transistors (SGT's) for Ultra High Density DRAMs", IEEE J. of Solid-State Circuits, 30 (1995) 960; S. Maeda, et al., "Impact of a Vertical

- ϕ -Shape Transistor ($V\phi T$) Cell for 1 Gbit DRAM and Beyond", IEEE Trans. on Electron Devices, 42 (1995) 2117.
- [166] J. Angell, S. Terry, and P. Barth, "Silicon micromechanical devices", Scientific American, 248 (1983) 44.
 - [167] J. Bryzek, K. Petersen, and W. McCulley, "Micromachines on the march", IEEE Spectrum, 31 (May, 1994) 20.
 - [168] S. Parker, C. Kenney, and J. Segal, "3D - - a proposed new architecture for solid-state radiation detectors", Nucl. Instr. and Meth. A395 (1997) 328-343.

Acknowledgments

This article was written at the request of Robert Kenney and Heinrich Leutz. They made useful editorial suggestions and participated in the preparation of the manuscript for the press. It would not exist had it not been for their interest and hard work. I would also like to thank Terry Walker, the gifted electrical engineer with whom I first worked, and who taught me my first lessons in integrated circuit nursery school and Bernard Hyams who started our joint project. Walter Snoeys and Julie Segal, Stanford graduate students in the Magicians Castle known as the Integrated Circuits Lab, continued my education. Christopher Kenney, with whom I work, Michael Wright, and Walter Snoeys read this manuscript and suggested changes. I am grateful to all of them.

Figures

1. Cross sections of (a) a conventional vertical, junction isolated, npn bipolar transistor, (b) a lateral pnp transistor on the same substrate. (figure from Muller & Kamins, pp 305,364). Plus and minus superscripts indicate heavily and lightly doped regions.
2. Schematic cross section of a JFET with (a) low current flow, (b) high flow at pinch-off voltage, and (c) beyond pinch-off in saturation region. The shaded region is depleted. The solid lines show the depletion regions and the p^+/n boundaries. Arrows indicate electric fields.
3. Schematic cross section through a MOS transistor (a). Below, the left column shows devices with the gate off, the right, on. The top row is for small drain-source voltages, the bottom, large. All but (b) conduct. Their conditions are called, (c) unsaturated, (d) punch through, and (e), saturated. Depleted silicon regions are indicated with a "d". The field implant prevents conduction between adjacent transistors. The field oxide passivates the surface, tying up dangling silicon bonds and also separates overlying conductors from the silicon. The phosphosilicate glass provides a smooth, gradually sloped surface for metal deposition.
4. Current - voltage characteristics for an enhancement transistor, (top) above threshold, and (bottom) below threshold. The gate oxide is 70nm thick, and the width/length ratio, W/L is 4/4 (microns). Characteristics are labeled by the substrate voltage (arrows) and, in groups of 3, by the gate voltage (right side). The source voltage is defined to be 0. Except for submicron edge effects, current is approximately proportional to W/L .
5. Variation of electron mobility as a function of the average transverse field (E -effective) and temperature. Neither substrate bias, V_{BB} (a), nor doping (b), changes the mobility. (see Sabnis and Clemens, Ref. [44])
6. NMOS - CMOS comparison. (a) NMOS inverter. (b) CMOS inverter. (c) cross section through a CMOS inverter using a p well. The p and n-plugs provide ohmic contacts to the p well and n substrate. Other structures, such as guard bands to prevent latchup are not shown here. (d) NMOS transmission gate (on for $V_{Control}$ high). (e) CMOS transmission gate (on for $V_{Control}$ high, $\bar{V}_{Control}$ low). The arrows from the substrate to the channel represent the diodes formed when the channel is inverted (and the transistor is "on"). They are often omitted when there is no ambiguity or when $V_{source} = V_{substrate}$. Similarly, the arrows at the source ends of the transistors are often omitted when all transistors are of the same type or when the source end is not uniquely defined (as in the case of the transmission gate.) The NMOS and PMOS transistors require substrates of opposite type, so for CMOS circuits, either, and sometimes both, are set in special wells.

7. Latchup currents. In normal operation, with $V_{dd} = V_{\text{substrate}} > V_{sd,\text{pmos}} > V_{sd,\text{nmos}} > V_{p\text{well}} = V_{ss}$, all diodes are back biased. However, electrons from the source - drains of NMOS transistors may diffuse down through the p well. If they escape recombination there, they will flow through the substrate to the (ohmic) V_{dd} contact. The IR drop across R_n will reduce the PMOS drain - source back bias, increasing hole diffusion through the substrate to the p well and its $p^+ V_{ss}$ contact. Current from holes that do not recombine in the substrate will increase the voltage across R_p , increasing the electron injection. If the increase is large enough, the positive feedback will cause a $V_{dd} - V_{ss}$ breakdown. Extending the V_{ss} and V_{dd} ohmic contacts (shown dotted) to form guard bands around the transistors greatly reduce R_p and R_n and thus the probability of latchup.
8. The basic photolithography sequence.
9. Fabrication steps for a CMOS circuit. Additional metal layers are used in many circuits. The same basic steps (but more of them) and used for BiCMOS chips. "FOX" labels regions of field oxide.
10. A grounded gate input protection transistor: (a) schematic diagram, (b) cross section, (c) plan view.
11. Photograph of the Microplex 1 input protection circuit tested to destruction. (The Microplex 1 had both a grounded gate transistor, partly visible at the right, and a metal-field oxide transistor, visible in the middle). The internal circuit was protected throughout the test; the aluminum input trace finally melted due to heat dissipation at the aluminum to diffusion contact.
12. Typical design, fabrication, and test steps for an integrated circuit. MOSIS (in the USA) will do all the steps in the dotted box. Many of the steps shown here are not needed, or will be done by the manufacturer, for gate arrays, standard cells, and logic arrays.
13. Examples of a general layout for (a), (b), power busses, and (c), control lines.
14. Examples of design rules for diffusion (often called "active") and poly layers. They are expressed in process independent units (lambdas). It is assumed the maximum error in any layer is less than lambda. Thus, with the required two lambda overlap of gate and diffusion, if the right diffusion border moved less than one unit to the right, and the poly gate less than one unit to the left, the gate would still overlap the diffusion path under it, preventing a short circuit of the transistor they form. The spacing requirement on the left of one lambda is less because an overlap here cuts the width of the diffusion path somewhat, increasing the resistance of that part of the source/drain, but does not disable it the way a short would a transistor.
15. (a)-(c) Three capacitors that will vary due to mask-to-mask registration variations, and, (d) one that won't. The pattern for both layers is shown in the overlap regions. (e) A common-

- centroid layout of four capacitors which produces equal capacities despite linear variations in dielectric thickness.
16. Isolation of a signal from switching transients using double correlated sample and hold techniques. (a) Top photograph, top trace pair: a calibration signal, alternately positive and negative with superimposed switching transients. Bottom pair: the signals after subtraction. To improve visibility, the calibration signal has been made 20 times normal signal height, which causes saturation of the output amplifier in the positive direction. (b) Bottom photograph, top trace: the signal that was subtracted. Bottom trace: same as above.
 17. Layout of the double-correlated sample and hold circuit used for Fig. 16.
 18. A layout in which the large, but differential, signals on the vertical crossing lines have greatly reduced coupling into the output of the horizontal differential amplifier. The use of differential analog signals will also help, particularly in reducing coupling into other analog lines via the digital ones crossing them.
 19. Diagram showing the generation, recombination, drift, and trapping of electrons and holes in a MOS transistor.
 20. (a) Subthreshold drain-source current before and after irradiation. The data is from P. Winokur et al., IEEE-NS 31 (1984) 1453. (b) Variation in thresholds (determined from plots similar to 20a) as a function of total dose. The 100KHz curve, for voltages cycling between the "off" and "on" ones, does not lie between either. The time is comparable to some hole transport times, and hole capture and interface trap formation can change in complicated ways. The top plot is for NMOS, and the bottom plot is for PMOS. The data are from P. Dressendorfer et al., IEEE-NS 28 (1984) 4281.
 21. (a) Frequency dependence of the noise voltage squared in a MOSFET for several total electron fluxes (one rad is approximately 39 million electrons per square cm). (b) Same as (a), but for a JFET. The dashed lines and filled circles are for noise with the thermal noise subtracted.
 22. Cross section (a) and plan view (b) of a guard band. (The relative orientation of the n and p channel transistors in the two views differs.) The extension of the thin gate oxide out to the guard band prevents source-drain leakage paths due to the radiation induced threshold shift of the transistor formed by the poly gate and the field oxide.
 23. Cross section view of NMOS and PMOS transistors (a) without and (b) with a field shield.
 24. Radiation hardness available (in 1987) in various technologies.
 25. (a) A manual probe station. (b) A view through the microscope showing a probe in place. (c) The same view at higher magnification (and shallower depth of field).
 26. A photograph of a trace cut with a focused ion beam and two traces joined with a tungsten patch. (Photo courtesy J. Glanville, FIB Applied Semiconductor Technology, San Jose, CA, 95131)

27. Test structures to measure (a) surface resistivity, and (b) contact resistance. (c) Photograph of ring oscillator.
28. (a) A semi-automatic die bonder. (b) Detail of the die pickup. (c) Detail of the die placement. Photographs courtesy of Kulicke and Soffa Industries, Inc.
29. (a) A manual wire bonder. (b) An enlarged view of the bonding tool and hybrid. Photographs courtesy of Kulicke and Soffa Industries, Inc.
30. Sequence of steps in making a ball/wedge bond pair. The last figure (g) shows the tool and initial step for making the next bond pair. Drawings courtesy of Kulicke and Soffa Industries, Inc.
31. Number of working channels on two 10 cm wafers containing Microplex 3, 128 channel chips. The regions marked "PCM" contain the process control monitors.
32. (a) Undercutting by a liquid etch produces structure size variations if the etched layer varies in thickness, while, (b) a directed plasma etch does not.
33. Minimum feature size in some commercial microprocessors and DRAMs as well as two custom chips (the Microplex and the first monolithic pixel detector) made in a university laboratory (Stanford) as a function of time. The two points with error bars represent the range of minimum line widths in production at those times. The circle labeled "IBM CMOS 7S" is for a technology using copper conductors. Power PC and IBM mainframe chips are expected, in 1998, to use this technology.
34. Minimum line-to-line pitch for polysilicon, bottom level metal (metal 1), and top level metal (metal 2).
35. Area of a planar surface used to represent a bit of information.
36. The decreasing number of dopant impurities in the base of bipolar transistors used for logic.
37. The increasing size of microprocessors. "P6" is the "Pentium Pro".
38. Number of transistors per die for dynamic random access memories and for microprocessors as a function of time. "P6" is the "Pentium Pro". The triangle labeled "IBM CMOS 7S" is for planned chips expected in 1998 using the IBM CMOS 7S technology.
39. Number of circuits on bipolar masterslice chips as a function of time. The ability to connect ever larger numbers on a single chip is greatly aided by the availability of additional levels of metal interconnects. See [131] for references for each symbol.
40. The decreasing energy dissipated per logic operation.
41. Ring-oscillator speed for bipolar emitter coupled logic (ECL) and non threshold logic (NTL) as a function of emitter stripe width. Self-aligned devices can be seen to be much faster than non-self-aligned ones.

42. Delay times vs power for a number of transistor technologies (in 1984). The diagonal lines show constant power-delay products. Josephson junction, MODFET, and GaAs heterojunctions are not now in commercial production. Times have continued to drop for those that are.
43. Time in person-months required to define and design integrated circuits as a function of calendar time.
44. The cost of wafer exposure machines as a function of time. DUV steppers use deep ultraviolet light to minimize the focal spot size, and step a small (1 - 2 cm in size) reticle across the wafer, adjusting its height at each site, to compensate for wafer variations.
45. Electron velocities for silicon and GaAs as a function of electric field. Low field mobilities are 1000 and 4000 volt-cm/sec² for silicon at 300°K and 77°K, and 7000 volt-cm/sec² for GaAs at 300°K. (But see also Fig. 5.)
46. Schematic drawing of a pixel detector showing electron-hole drift lines for one pixel.
47. Typical events in a 4-detector telescope. Bar lengths are proportional to the pulse height after pedestal subtraction. Only positive noise signals are shown. The normal to the third detector is slanted at 56° to the beam direction.
48. Spatial resolution as a function of pulse height. Resolution is degraded for low pulse heights due to electronic noise, and at large pulse heights due to delta rays.
49. Track angle determined from cluster width for 19°, 33°, and 56° tracks. Gaussian curves with rms sigmas of 1.6°, 2.7°, and 1.4° are fitted to the central peaks.
50. A cross sectional view of a pixel indicating the spatial resolution in the two types of charge-collection regions within a cell.

Figure 1

Figure 2

Figure 3

Figure 4

(a)

(b)

Figure 5

Figure 6

Figure 7

The Basic Process (repeated many times)

Figure 8

Figure 9

(a)

(b)

(c)

Figure 10

Figure 11

Figure 12

Figure 13

Figure 14

(a)

(b)

(c)

(d)

Common-centroid layout of the storage capacitors.

(e)

Figure 15

Figure 16

Figure 17

Figure 18

Figure 19

Figure 20a

Figure 20b

(a)

(b)

Figure 21

Device Cross Section

Device Layout

Figure 22

Figure 23

Figure 24

Figure 25a

Figure 25c

FEI FIB 610

— 1 μ m

BEAM ENERGY: 25 keV

Figure 26

Figure 27a

Figure 27b

Figure 27c

Figure 28a

Figure 28b

MODEL 6491 SEMIAUTOMATIC DIE BONDER FOR SILVER GLASS DIE ATTACH

FEATURES

- ★ PROGRAMMABLE ADHESIVE PATTERNS
- ★ SOFTWARE CONTROLLED BOND LINE THICKNESS
- ★ QUICK CHANGEOVER
- ★ PROGRAMMABLE FORCE CONTROL
- ★ DIE EDGE ALIGNMENT FOR ACCURATE LARGE DIE PLACEMENT
- ★ AUTOMATIC COMPENSATION FOR VARIATIONS IN DIE AND PACKAGE THICKNESS

Figure 29a

Figure 29b

Figure 30

(a)

(b)

Figure 32

Figure 33

Figure 34

Figure 35

Figure 36

Figure 37

Figure 38

Figure 39

Figure 40

Figure 41

Figure 42

Figure 43

Figure 44

Figure 45

Figure 46

Figure 47

Figure 48

Figure 49

Figure 50