COPY 1

REPORT NO. FAA-RD-76-62

AN EVALUATION OF SILICON ON SAPPHIRE TECHNOLOGY

HENRY DOMINGOS
RICHARD ANDERSON

LIBRARY
11 78

MARCH 1976

FINAL REPORT

Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
FEDERAL AVIATION ADMINISTRATION
Systems Research & Development Service
Washington, D.C. 20590

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FAA-RD-76-62		

4. Title and Subtitle	5. Report Date
An Evaluation of Silicon on Sapphire Technology	March 1, 1976
	6. Performing Organization Code RADC/RBC
7. Author(s) Henry Domingos and Richard Anderson	8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Defense U.S. Air Force Rome Air Development Center Rome, NY 13441	10. Work Unit No. (TRAIS) 9567006 11. Contract or Grant No. DOT/FA72WAI-356
12. Sponsoring Agency Name and Address Department of Transportation Federal Aviation Administration Systems Research and Development Service Washington, D.C. 20590	13. Type of Report and Period Covered Final Report March 1, 1975 March 1, 1976 14. Sponsoring Agency Code ARD-350 (215-607)

15. Supplementary Notes

Prepared by Post Doctoral Programs, Rome Air Development Center

Domingos, Clarkson College of Technology; Anderson, Syracuse University

16. Abstract

Silicon on sapphire is a recent microelectronic circuit technology originally devised to eliminate the drawbacks of conventional junction isolation. It has since developed into a high performance MOS-type integrated circuit family. In this report a description of this logic family is given, including such topics as manufacturing processes, device physics, and performance of commercial circuits. The report concludes with an overall evaluation of this technology.

17. Key Words	18. Distribution Statement
Silicon on Sapphire Microelectronic Circuit	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22151.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	151	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

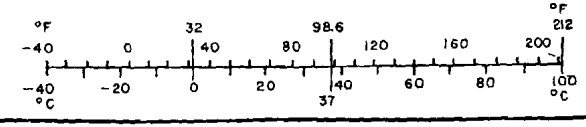
Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.

Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



FOREWORD

The Post-Doctoral Program at Rome Air Development Center is pursued via Project 9567 under the direction of Dr. W.W. Everett, Jr. The Post-Doctoral Program is a cooperative venture between RADC and the participating universities: Syracuse University (Department of Electrical and Computer Engineering), the U.S. Air Force Academy (Department of Electrical Engineering), Cornell University (School of Electrical Engineering), Purdue University (School of Electrical Engineering), University of Kentucky (Department of Electrical Engineering), Georgia Institute of Technology (School of Electrical Engineering), Clarkson College of Technology (Department of Electrical and Computer Engineering), State University of New York at Buffalo (Department of Electrical Engineering), North Carolina State University (Department of Electrical Engineering), Florida Technology University (Department of Electrical Engineering), Florida Institute of Technology (College of Engineering), Air Force Institute of Technology (Department of Electrical Engineering), The Naval Postgraduate School (Department of Electrical Engineering), and the University of Adelaide (Department of Electrical Engineering) in South Australia. The Post-Doctoral Program provides, via contract, the opportunity for faculty and visiting faculty at the participating universities to spend a year full time on exploratory development and operational problem-solving efforts with the post-doctorals splitting their time between RADC (or the ultimate customer) and the educational institutions.

The Post-Doctoral Program is totally customer funded with projects undertaken for Air Defense Command (NORAD), Air Force Communications Service, Federal Aviation Adminstration, Defense Communications Agency, Aeronautical Systems Division (AFSC), Air Force Aero Propulsion Laboratory (AFSC), Rome Air Development Center (AFSC), Electronics System Division (AFSC), NASA, Air Force Avionics Laboratory (AFSC), and Air Force Weapons Laboratory (AFSC).

This effort was conducted via the Reliability and Compatibility branch of RADC for the Federal Aviation Administration. Responsible engineers at RADC are Ed O'Connell, Regis Hilow, Bob Thomas and Clyde Lane. Mr. Fred Sakate is the responsible person at FAA.

The authors of this report are Prof. Henry Domingos of Clarkson College of Technology and Prof. Richard Anderson of Syracuse University.

Dr. Henry Domingos received a BEE Degree (1956) from Clarkson College, the MSEE Degree (1958) from the University of Southern California, and the Ph.D. Degree (1963) from the University of Washington. His industrial experience is with Union Carbide Corporation, Hughes Aircraft Co., Atomics International Inc., and the General Electric Co. He has taught Electrical Engineering at the University of Nevada, the University of Washington and Clarkson College of Technology where he is Associate Professor of Electrical Engineering. He is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, the IEEEE, The American Physical Society and the American Society of Engineering Education.

Dr. Richard Anderson received a BEE (1950) and M.S. (1952) degrees from the University of Minnesota and the Ph.D. degree (1960) from Syracuse University. He has worked in the IBM Research Laboratory and has taught Electrical Engineering at Syracuse University where he is Professor of Electrical and Computer Engineering.

He was a Visitng Professor in the Physics Department of the University of Madrid (1960-61) and at the University of Sao Paulo (1967-69), both under the auspices of the Fullbright-Hays program.

He is a member of the American Physical Society, the Brazilian Physical Society, the Electrochemical Society and is a Fellow in the IEEE. He holds an honorary Doctorate from the University of Sao Paulo.

The authors wish to acknowledge the cooperation of the many persons in industrial, educational and government laboratories who helped us in the preparation of this report. We are particularly indebted to Dave Meyer, Bill Laughton, Earl Parks, Ed O'Connell, Regis Hilow, Bob Thomas, Ivars Irbe, and Clyde Lane for their help in collecting and preparing the data presented here, and in evaluating the various technologies involved in SOS.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	vii
LIST OF TABLES	x
SECTION I: INTRODUCTION	1
A. History and Development of SOS Technology	1
B. Characteristics of SOS Circuits	4
C. Scope of This Report	6
SECTION II: SUBSTRATES	8
A. Sapphire vs. Spinel	8
B. Properties of Synthetic Sapphire	11
C. Growth and Machining of Sapphire Wafers	13
D. Trends in Substrate Costs	25
SECTION III: FABRICATION	27
A. Substrate Preparation	27
B. Epitaxial Growth	29
C. Fabrication Methods	35
1. Device Structures Available	35
2. Double Epitaxial Metal Gate Process	35
3. Single Epitaxial Metal Gate Process with Ion Implantation	41
4. Planar Metal Gate Process	46
5. Silicon Gate Process	46
6. Gate Oxidation	50
7. Thick Oxide	51
8. Silicon Etch	52
9. Process Reproducibility	52
10. Dicing	53
11. Mounting and Packaging	53
12. Summary	53
SECTION IV: ELECTRICAL PROPERTIES OF SOS DEVICES AND CIRCUITS	55
A. Bipolar Diodes and Transistors	55
B. Single Channel PMOS Or NMOS	56
C. CMOS Devices	57
1. General Characteristics	57
2. Threshold Voltage	61
3. Drain Current and Gain	63
4. Leakage Current	68
5. Kink Effect	69
6. Edge Effects	72
D. CMOS Logic Circuits	73
E. Transient Performance	75

	Page
F. Power Dissipation	79
G. Bulk CMOS Layout	79
H. SOS CMOS Layout	81
I. Linear Circuits	82
SECTION V: SOS CIRCUITS AND PERFORMANCE CHARACTERISTICS	83
A. Bulk CMOS Dual Complementary Pair Plus Inverter	83
B. SOS Dual Complementary Pair Plus Inverter	96
C. Quad 2-input NOR Gate	105
D. 7-Stage Binary Counter	109
E. Quad Bilateral Switch	111
F. SOS Test Chip	111
G. Decade Counter, and 14-stage Binary Counter	114
H. 256 Bit Static RAM	116
I. SOS Timing Circuit	120
SECTION VI: DISCUSSION	123
A. SOS Processing	123
B. Yield and Area Considerations	124
C. Reliability	125
D. Radiation Effects	126
E. Comparison With Other Integrated Circuit Families	128
F. Market Outlook	131
G. Availability of Parts	133
H. Summary and Recommendations	134
REFERENCES	136

LIST OF FIGURES

Figure		Page
I-1.	Processing steps for silicon on sapphire device fabrication.	5
II-1.	Structure of synthetic sapphire.	12
II-2.	Temperature variation of the thermal conductivity of sapphire.	16
II-3.	Variation of the specific heat of sapphire.	17
II-4.	Transmittance of sapphire as a function of wavelength.	18
II-5.	Modified Bridgman method of growing single crystals of sapphire.	19
II-6.	Modified Czochralski method for growing sapphire ribbons.	21
III-1.	Electron mobility in silicon on sapphire films.	33
III-2.	Hole mobility in silicon on sapphire films.	33
III-3.	Change in electron mobility due to oxidation of film.	34
III-4.	Change in hole mobility due to oxidation of film.	34
III-5.	Devices which can be fabricated with SOS technology.	36
III-6.	Fabrication sequence for the double epi metal gate process.	37
III-7.	Fabrication sequence for the single epitaxial metal gate process with ion implantation.	42
III-8.	Planar metal gate process.	47
III-9.	Circuit processing by the silicon gate, ion implantation method.	48
IV-1.	Drain characteristic curves for aluminum gate SOS transistors with 20 mil channels.	58
IV-2.	Drain characteristic curves for silicon gate SOS transistors with 2 mil channels.	60
IV-3.	Threshold voltage for p-channel aluminum gate SOS transistors.	64
IV-4.	Threshold voltage for n-channel aluminum gate SOS transistors.	65

		Page
IV-5.	Threshold voltage for p-channel silicon gate SOS transistors.	66
IV-6.	Threshold voltage for n-channel silicon gate SOS transistors.	67
IV-7.	Switching path in the device which exhibits a kink effect.	71
IV-8.	Basic CMOS inverter.	71
IV-9.	Transfer characteristics for a typical CMOS inverter.	74
IV-10.	Basic CMOS NAND gate and NOR gate.	74
IV-11.	Basic configuration of one inverter driving another for purposes of calculating delay time.	77
IV-12.	Bulk CMOS inverter layout.	80
IV-13.	Aluminum gate SOS CMOS layout.	80
V-1.	Schematic of a bulk CMOS dual complementary pair plus inverter.	84
V-2.	Photomicrograph of a bulk CMOS dual complementary pair plus inverter circuit.	85
V-3.	Layout and cross section of the p-channel transistor.	87
V-4.	Layout and cross section of the n-channel transistor.	88
V-5.	Diagram and layout of the gate protection circuit.	89
V-6.	Electrical characteristics of the gate protection circuit.	91
V-7.	Equivalent circuit for a bulk CMOS complementary pair.	92
V-8.	Drain characteristic curves for transistors in the bulk CMOS dual pair.	94
V-9.	Kink effect in bulk CMOS transistors with floating substrates.	95
V-10.	SOS CMOS dual complementary pair plus inverter.	97
V-11.	Gate protection device for the SOS CMOS dual pair plus inverter.	99
V-12.	Electrical characteristics of the SOS gate protection device.	100

		Page
V-13.	Electrical characteristics and construction of p-channel SOS transistor.	102
V-14.	Electrical characteristics of an SOS n-channel device.	103
V-15.	Photomicrograph of a quad 2 input SOS NOR gate.	106
V-16.	Schematic diagram of an SOS quad 2 input NOR gate.	107
V-17.	Electrical characteristics and layout of gate protection device used in SOS NOR gate.	108
V-18.	Photograph and block diagram of an SOS 7-stage binary counter.	110
V-19.	Photograph of an SOS quad bilateral switch.	112
V-20.	Photograph of a complex SOS evaluation chip.	113
V-21.	Photographs of SOS counters.	115
V-22.	Photomicrograph of an SOS 256-bit RAM.	117
V-23.	Block diagram of a 256-bit SOS RAM.	118
V-24.	Photograph and diagram of a memory cell in a 256-bit SOS RAM.	119
V-25.	Photograph of an improved SOS 256-bit RAM using a silicon gate, ion implant process.	121
V-26.	Photomicrograph of an SOS timing chip.	122

LIST OF TABLES

TABLE		Page
II-1.	Physical Properties of Silicon, Sapphire, and Spinel.	9
II-2.	Physical Properties of Synthetic Sapphire.	14
III-1.	Silicon-Sapphire Orientation.	29
V-1.	Characteristics of the Bulk CMOS Process Used for the Dual Complementary Pair Plus Inverter.	90
V-2.	Electrical Characteristics and Specification of Bulk CMOS Devices in the Dual Complementary Pair Plus Inverter.	93
V-3.	Structural Characteristics of an SOS Dual Complementary Pair Plus Inverter.	104
V-4.	Electrical Characteristics of an SOS Dual Complementary Pair Plus Inverter.	104
V-5.	Typical Characteristics of an SOS and Bulk CMOS Quad 2-input NOR Gates From Two Manufacturers.	109
V-6.	Characteristics and Performance of 7-Stage Binary Counters	111
V-7.	Characteristics of Decade Counters.	114
V-8.	Characteristics of 14-Stage Binary Counters.	114
VI-1.	Performance Characteristics of Logic Families.	129
VI-2.	Characteristics of MSI Integrated Circuit Families.	129
VI-3.	Free World Markets for Integrated Circuits in Millions of Dollars.	132

SECTION I

INTRODUCTION

A. History and Development of SOS Technology

When small scale integrated circuits were being developed in the early 1960's, the only available method of electrically insulating components on the same chip was by reverse-biased p-n junction diodes. This technique possesses well-recognized disadvantages. The drawbacks fostered the development of several alternative schemes, one of which was silicon on sapphire technology. It was clear that the fabrication of circuit components directly onto an insulating substrate would eliminate the problems associated with junction isolation.

The epitaxial growth of single crystal semiconductor films on insulating substrates has been the subject of much research effort in the past decade or so. A comprehensive review of the subject has been published recently (1). For theoretical and practical purposes only the heteroepitaxy of silicon is commercially utilized for integrated circuits, and only sapphire substrates will be treated at length in this report. A wide variety of other substrate materials, including spinel, quartz, silicon carbide, beryllium oxide, chrysoberyl, and garnet, have been investigated and discarded, perhaps prematurely. Only spinel remains as a subject of current investigation, but sapphire alone is commercially utilized, for reasons which will be discussed.

The first successful epitaxial growth of silicon on sapphire was reported in 1963 by Manasevit and Simpson (2), (3). They utilized the reduction of silicon tetrachloride by hydrogen at elevated temperatures to deposit single crystal layers on oriented sapphire substrates. The films were p-type, of high resistivity. The authors noted that good epitaxial films depended on the apparent condition of the substrate, which was cleaned and etched by hydrogen at 1260°C. There was also evidence of reaction between the silicon and sapphire during initial stages of deposition. Growth occurred by nucleation, as opposed to formation of monolayers. The observations of Manasevit and Simpson are basically unchanged after extensive research in a number of laboratories since 1964

(4) - (9).

Practically all device processing is done with (100) silicon on $(1\bar{1}02)$ sapphire. It has been clearly established that the surface finish of the substrate is critically important to good film growth, and various mechanical and chemical etches plus some form of hydrogen or air prefiring are commonly, but not universally used. Silane is used exclusively instead of silicon tetrachloride, which causes excessive etching of the substrate. Aluminum autodoping of the silicon films, by outdiffusion from the substrate or transport from the backside of wafers via the carrier gas, can be controlled or counteracted.

The most undesirable characteristic of silicon films heteroepitaxially deposited on sapphire is the low lifetime of minority carriers, roughly 1 ns. In contrast, minority carrier lifetimes in bulk silicon devices are several orders of magnitude higher. The short lifetime makes it difficult to fabricate good bipolar devices, and leads to high leakage in MOS devices. The cause of the low lifetime is the highly disordered interface layer and the presence of impurities introduced from the substrate. Rai-Choudhury, et al (10), have shown that the lifetime can be increased by a factor of one hundred by first depositing an n^+ layer of silicon, followed by a lightly doped n layer.

Hall mobilities are found to be 20-50% of the values in bulk silicon, and increase with the thickness of the silicon film (9). Once the film thickness is greater than a few tenths of a micron, the silicon quality is very good. Since the channel of an enhancement mode device is only several hundreds of angstroms thick, good quality MOS devices can readily be fabricated in such films.

Although bipolar transistors have been fabricated in SOS films (11), (12), the low minority carrier lifetime requires special care and novel processing steps to achieve useful h_{fe}. Consequently, only MOS transistors are important in SOS integrated circuits today. The earliest devices were, indeed, MOS p-type enhancement mode transistors (13). Since the electrical characteristics of MOS transistors depend mainly on the silicon properties near the surface of the film where quality is superior, performance of the devices is excellent. Boleky has reviewed the performance of the

complementary enhancement mode devices and deep-depletion devices on insulating substrates (14). Deep-depletion components make it possible to achieve CMOS performance with simpler processing steps (15). Normally, both n and p-type high resistivity layers must be fabricated to build complementary transistors. By using a deep-depletion transistor in conjunction with one type of enhancement mode device, CMOS inverters can be designed which require only one type of epitaxial layer. As Boleky points out, deep-depletion technology requires stringent control over film thickness and doping. Nevertheless, these devices are currently utilized in existing SOS products.

In the few years immediately following the introduction of silicon on sapphire technology, several integrated circuit manufacturers and laboratories investigated the new technique at varying levels of effort. With the exception of two companies, the integrated circuit industry of that time came to the conclusion that silicon-on-insulating substrate technology had sufficient technical problems to not warrant further investigation. These problems were mainly process related difficulties associated with the degradation of electrical and mechanical properties of silicon films deposited on insulating substrates.

During the 1966-71 period a few companies maintained research and development programs in both bipolar and MOS devices. At the end of that period, SOS circuits were becoming available but not in sufficient quantities to develop serious mass market penetration. However, several other manufacturers became interested in producing SOS circuits. To improve the basic circuit performance, the integrated circuits manufacturers were developing CMOS circuits using two epitaxial silicon layers or implanted p-regions instead of one epitaxial layer with deep-depletion devices. One company in addition was working on bipolar circuits, with efforts to increase minority carrier lifetime. In the period since 1971, the list of manufacturers actively pursuing SOS technology has changed somewhat. At the present time, there are more manufacturers committed to and more products being developed in SOS than ever before. 4000 series CMOS circuits, 256 and 1024 bit RAM's, and other circuits are currently available. It remains to be seen if this manufacturing effort will be maintained and significant mass market penetration achieved in the

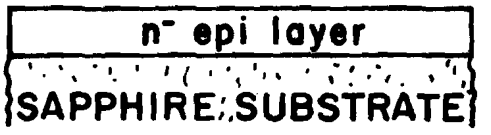
face of current poor market conditions. One signal that SOS production problems may still exist was the recent withdrawal of a small SOS memory system from the market due to lack of SOS parts in sufficient quantities.

B. Characteristics of SOS Circuits

The basic advantage of SOS technology is derived from the elimination of virtually all parasitic effects from the integrated circuits. This can be readily appreciated by studying the simplified sequence of processing steps shown in Fig. I-1.

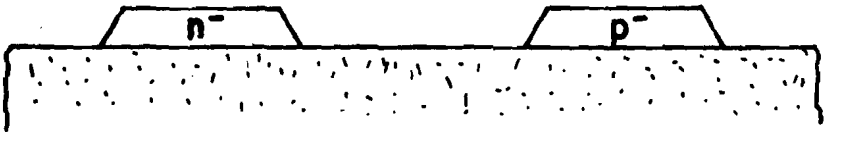
The figure illustrates the so-called double-epi method of fabrication. The first step is the growth of a 1 μm n-type silicon film with doping density $5 \times 10^{14}/cm^3$. This is a crucial step and requires good processing control. Then a standard oxide growth, photomasking, etc. defines the regions which will be used for p channel enhancement transistors, bonding pad pedestals, etc. Next the p-type epitaxial layer is grown and etched into islands. n^+ and p^+ diffusions are done simultaneously using appropriately defined phosphorus and boron doped glass for the diffusion sources. Part (e) of the figure shows a completed device with gate oxide (1000 Å) and metalization (10000 Å).

The advantages of such a structure are readily apparent. Instead of using various isolation schemes such as reverse-biased junctions, guard rings, etc., SOS circuits are constructed such that all of the thin silicon layer not incorporated in active or passive devices is removed. Ideally, the SOS circuit can be conceived as a set of perfectly isolated discrete devices in integrated form. All silicon in the field regions of the circuit is removed except for crossovers. Thus, field threshold and leakage problems with metalization are eliminated. Also, the metal interconnection lines have drastically less parasitic capacitance associated with them, being determined solely by the number of crossovers. From a reliability standpoint, the elimination of the field silicon eliminates possibilities of leaky junction isolation, low field threshold, incomplete action of guard rings, shorts due to field oxide pinholes, etc. From a silicon surface area utilization standpoint, area reductions can be made since guard rings in CMOS circuits and wide laterally diffused isolation junctions in bipolar

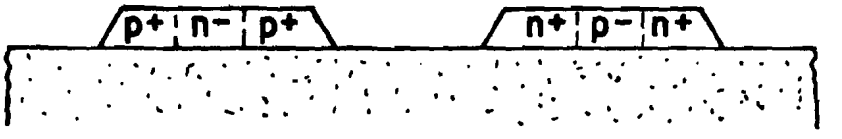


(a) growth of epi layer

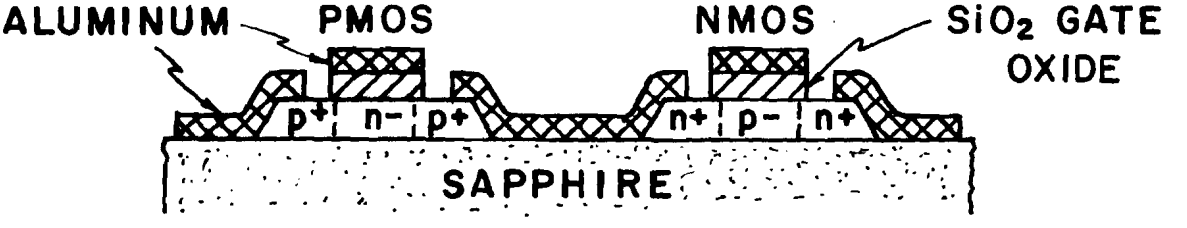
(b) n-island definition



(c) p-island definition



(d) source and drain diffusion



(e) completed device

Figure I-1. Processing steps for silicon on sapphire device fabrication.

circuits can be eliminated. Since silicon film thickness is only about 1 μm there are no deeply diffused junctions to waste surface area by lateral diffusion. Thus, circuit design rules can be tightened and simplified.

The circuit and device performance of SOS circuits are greatly enhanced as compared to their bulk counterparts. Due to the removal of parasitic silicon material, the isolation capacitance and leakage of bipolar transistors, diodes, and diffused resistors are virtually eliminated and the drain and source-to-channel region capacitance and leakage is reduced to just the sidewall component of a 1 μm diffusion. Since the parasitic capacitance associated with metal interconnection lines is eliminated, the response time of these lines is decreased appreciably, particularly on large LSI chips where metal conductors frequently run the length of the chip.

Another distinct advantage of SOS circuits is the increased hardness to transient radiation effects (16). The reduced junction area results in a dramatic reduction of the volume for photocurrent generation.

The advantages of SOS circuits compared to ordinary bulk CMOS can be summarized as follows:

1. higher speed
2. higher packing density
3. reduced power
4. increased transient radiation hardness
5. faster processing.

At the same time the technology has certain disadvantages. Among these are:

1. high cost of substrates
2. stringent requirements on certain processing steps
3. poorer electrical characteristics of films and devices
4. low production volume.

C. Scope of This Report

The purpose of this report is to assess the current state-of-the-art in SOS technology in terms of performance, cost, and availability for system design. The authors have attempted to bring together sufficient information to give a clear indication of where the technology stands at the present time and what the future trends will be.

The report is organized as follows. First the characteristics of sapphire substrates and their preparation will be described. Then the details of silicon deposition and circuit fabrication will be discussed in detail. Theoretical and experimental characteristics of MOS devices will be developed and illustrated, and specifications of several currently available circuits will be described. Finally, an overall evaluation will be made.

Recently a commercial report (56) on silicon on sapphire technology has become available. Although much of the information is the same as contained herein, it also has a comprehensive coverage of markets, applications, and semiconductor company activities. Readers interested in these aspects of the technology will find it a useful complement to the present report.

SECTION II

SUBSTRATES

A. Sapphire vs. Spinel

There are several factors to be considered in the choice of a substrate material for use in depositing a thin film of single crystal silicon. The lattice mismatch between the substrate and silicon must not be too great, and the thermal coefficients must be approximately equal. It must be possible to clean and otherwise prepare the substrate for epitaxial growth. The substrate surface must remain chemically inert during the deposition, and there cannot be any compounds formed at the interface which will cause inversion layers or leakage currents in the completed devices. The electrical properties of the silicon film must be similar to those of bulk silicon, and remain stable during subsequent processing steps and projected use life of the circuits. The silicon crystalline structure must be free of imperfections that can cause processing fall-out (such as diffusion spikes), and latent failures in the finished circuits.

More than ten single crystal materials have been considered as substrates for the growth of a thin epitaxial layer of silicon. Descriptions of these can be found in the literature (17-19). At the present time, however, only sapphire and spinel are being used. A comparison of the important physical properties of silicon, sapphire, and spinel is given in Table II-1. It would appear from an examination of the properties that spinel would be a superior substrate material because of its crystal class, better lattice fit, better thermal coefficient match, and lower hardness. This is not the case, however, as will be pointed out shortly.

Spinel substrate material has been successfully produced by the flame fusion, flux, and Czochralski methods. The composition of flame fusion spinel varies between $MgO \cdot Al_2O_3$ and $MgO \cdot 3.3Al_2O_3$, while flux grown spinel is stoichiometric (9). Czochralski grown spinel may vary from the stoichiometric composition by a small amount depending on growth conditions.

Spinel substrates grown by all three methods are presently available. Flux grown spinel is not widely used because it commonly contains contamination from the flux used in growth (20). Methods of removing the flux have been reported, but little work has been done with flux grown

Table II-1. Physical Properties of Silicon, Sapphire, and Spinel (8).

Parameter	Silicon	Sapphire α-Al_2O_3	Spinel $MgAl_2O_3$
Crystal structure system	face-centered cubic	rhombohedral	face-centered cubic
Unit cell dimension (Å)	a = 5.4301	a = 4.758, c = 12.991	a = 8.083
Density (g/cm^3)	2.33	3.98	3.58
Hardness (mohs)	7	9	8
Melting point (°C)	1412	2040	2130
Dielectric constant	11.7 (500 Hz-30 MHz)	9.4 (90° to c axis) (100 Hz-100 kHz)	8.4 (100 Hz-100 kHz)
Refractive index	3.4975 (at 1.357 μm)	1.7707 (at 5461 Å)	1.7202 at 5461 Å)
Optical transmission	transparent in infrared	80% min (0.24-6.0 μm, 0.0175 in.)	80% min (0.31-5.1 μm, 0.0175 in.)
Thermal conductivity (cal/cm·sec ·°C) at 25°C	0.30	0.0065 (60° to c axis)	0.035
Thermal expansion coefficient (1/°C) at 25-800°C	3.59×10^{-6}	8.40×10^{-6} (60° to c axis)	7.45×10^{-6}

material for this reason. Although flame fusion material is satisfactory, the trend has been to use the Czochralski method because of the preference for large diameter wafers (21).

The stability of flame fusion spinel in integrated circuit processing is inversely proportional to the alumina content within the compositional range $MgO \cdot Al_2O_3$ to $MgO \cdot 3.5\ Al_2O_3$ (8). Czochralski grown spinel becomes alumina rich during the hydrogen prefiring stage, and either the processing stability degrades considerably or the surface develops micro-cracks. In addition, large differences in electrical properties have been observed in silicon films deposited on spinels that differed only by small changes in lattice constants, and were otherwise nearly identical (21). Because of this result, it is evident that lattice matching considerations are secondary to the chemical reactivity between substrate, silicon, and carrier gas H_2.

It has been shown that the major imperfections in silicon on (111) spinel are stacking faults (22). Deformation in these films is anistropic and exhibits maximum values in one direction and minimum values in the perpendicular direction. Silicon on (100) and (110) spinel substrates shows similar results, although smaller deformations are observed. In non-stoichiometric material this deformation is accompanied by cracking of the substrate and exsolution of Al_2O_3 from the spinel. The maximum deformation is approximately five times the amount found in silicon on sapphire. The silicon is under compressive stress, $1\text{-}4 \times 10^{10}$ dynes/cm^2 (22).

In the past, spinel was thought to be the better substrate material because of its better crystallographic match, but this has been overridden by practical considerations. Higher quality silicon films can be deposited on sapphire because of its higher degree of chemical inertness and lower stacking fault density. The mechanical strength of sapphire allows wafer processing without special handling precautions, whereas spinel is more difficult to work with because of its brittle nature. Spinel substrates tend to shatter during thermal cycling unless temperature changes are less than 10°C/sec. When all considerations are taken into account, sapphire emerges as the substrate of choice. Spinel is still used in research

programs, but primarily for comparison with sapphire in order to better understand silicon on insulating substrates.

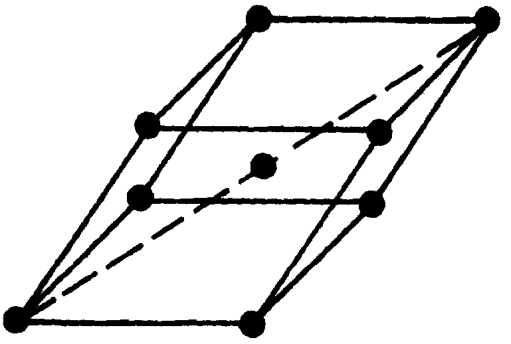
B. Physical Properties of Synthetic Sapphire

Substrates for SOS integrated circuits are of single crystal synthetic sapphire, chemical composition Al_2O_3 (aluminum oxide), sometimes referred to as corundum. It is the crystalline form of the ceramic commercially known as alumina, widely used in the electrical industry because of its outstanding electrical, mechanical, and chemical characteristics. Alumina is commonly used for the substrates of hybrid integrated circuits, cores for discrete resistors, electrical insulators, etc. It is not surprising that the two seemingly unrelated materials have similar physical properties.

The usefulness of sapphire is due to its outstanding physical characteristics, some of which include: low electrical conductivity, high thermal conductivity, low dielectric loss, chemical inertness, good mechanical strength, high abrasion and wear resistance, extremely low porosity, and extremely wide temperature range of usefulness. Its main drawback is its difficulty in machining.

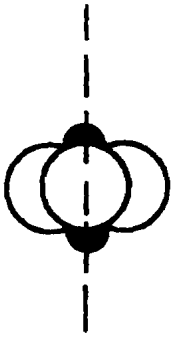
The unique properties of sapphire make it invaluable in a wide range of applications in addition to SOS integrated circuits. Some of these are: infrared windows and lenses, laser reflectors, solar cell covers, photomultiplier tubes, pressure cell windows, spectrometer windows, optical filter substrates, cryogenic heat sinks, jewel bearings, insulating washers and spacings, photomasks and plates, radiation light pipes, etc.

The structure of synthetic sapphire has been analyzed by Bragg and Bragg (23), by Pauling and Hendricks (24), and by Newnham and de Haan (25). The crystal forms are sometimes referred to rhombohedral axes and sometimes referred to hexagonal axes. The rhombohedral unit cell, molecular group, and hexagonal cell are shown in Fig. II-1. The diagonal of the unit cell is the c axis of the hexagonal cell. The molecular group is oriented at each lattice point of the unit cell such that a line drawn through the centers of the aluminum atoms is parallel to the c axis. The structure can also be viewed as approximately a hexagonal close-packed array of oxygen

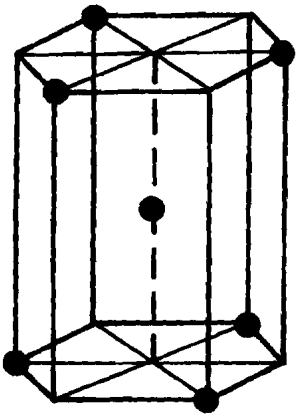


a = 5.12 Å
α = 55° 17'

rhombohedral cell

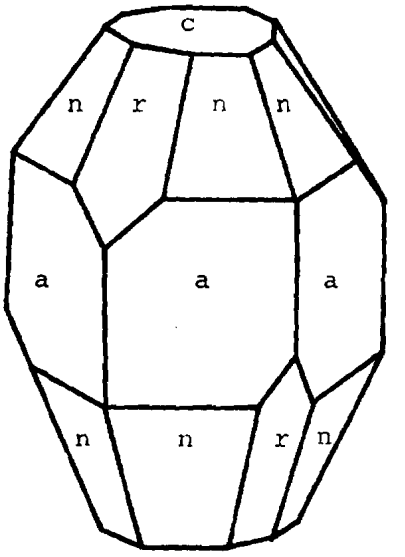


arrangement of 2 aluminum and 3 oxygen atoms at each lattice point



a = 4.76 Å
c = 12.99 Å

hexagonal cell



sapphire crystal faces

Figure II-1. Structure of synthetic sapphire. The dotted lines are parallel to the c axis. The "r" cut is used for silicon heteroepitaxy.

atoms with aluminum atoms in two-thirds of the octahedral interstices. If a large number of impurities are present colloidal particles may form, giving a smoky blue appearance to the crystal. If some of the aluminum atoms are replaced by chromium atoms, a pink (0.04% replacement) or deep red (0.5% replacement) ruby crystal results.

Sapphire absorbs little water, and is insoluble in water, common acids and common alkalis to 1000°C. It is not attacked by HF below 300°C. Physical properties of synthetic sapphire are summarized in Table II-2.

C. Growth and Machining of Sapphire Wafers

Single crystal sapphire can be grown by a variety of methods. The oldest method is the Verneuil or flame-fusion technique. It is still used, especially in Europe, but is a declining source of high quality sapphire in this country and is no longer used for SOS substrates. The first SOS films in 1963 were fabricated on 1/4 inch Verneuil sapphire, and flame fusion is used to prepare the crackle or polycrystalline material from which sapphire crystals are grown.

SOS quality sapphire is currently available from three manufacturers in this country, each of which uses a different method. Manufacturer A uses the well-known Czochralski technique with an iridium crucible to grow crystals up to 24 inches in length and up to 5 inches in diameter. Dislocation density is in the range 10^3-$10^4/cm^2$. This company maintains complete facilities for growing crystals, sawing, grinding, lapping, and polishing in both production and custom modes. In fact, wafers with epitaxial silicon can be supplied to order. Tolerances on sapphire optical parts include: surfaces to 250 Å absolute; optical finishes to 20-10; flatness, 1/10 wave for 1 inch diameter, 1/4 wave for 2 inch diameter, and 1 wave for 3 inch diameter; parallelism to 2 seconds or arc; thickness down to 1/2 mil. Prices for the commonly used 2 inch, 14 mil thick wafer are $35.25 each in small quantities, down to $23.25 in quantities of 4000 or more. Still lower prices, down to $14.00, are quoted on large volume, long-term orders. Vendor A is the largest supplier of SOS wafers at the present time.

Vendor B has been marketing sapphire substrates for about 2 years. A heat exchanger method is used to grow single crystals up to 10 inches in diameter and 5 inches long. Figure II-5 shows a diagram of a crystal growing furnace.

Table II-2. Physical Properties of Synthetic Sapphire*

Chemical Formula	Al_2O_3
Molecular Weight	101.94
Specific Gravity	3.98
Melting Point	2,040°C
Thermal Conductivity	0.46 watts/cm/°K at 300°K, 60 degrees from the hexagonal axis. Affected only slightly by crystallographic orientation. See Fig. II-2 for temperature variation.
Specific Heat	0.778 joules/gm/°K at 300°K. See Fig. II-3 for temperature variation.

Thermal Expansion Coefficient

Temperature range	Parallel to c axis	60° to c axis	90° to c axis
20-50°C	6.66×10^{-6}/°C	5.8×10^{-6}/°C	5.0×10^{-6}/°C
20-500	8.33	7.7	7.7
20-1000	9.03	8.4	8.3
20-1500	-	9.0	-

Dielectric Constant

Temperature	E at 90° to c axis	E along c axis
20°C	9.35	11.53
100	9.43	11.66
300	9.66	12.07
500	9.92	12.54
700	10.26	13.18

Dielectric Strength 1200 volts/mil at 60 Hz

Resistivity

0°C	10^{16} ohm cm
500	10^{11}
1,000	10^{9}
1,500	10^{5}
2,000	10^{3}

Hardness Moh 9 (quartz 7, diamond 10), Knoop 1525-2000

Bulk Modulus	300,000 psi	
Young's Modulus	at 30° to c axis	5.5×10^7 psi
	45°	5.1
	60°	5.0
	75°	5.6
Modulus of Rigidity	2.5×10^7 psi minimum	
Modulus of Rupture	at 30° to c axis	100,000 psi
	45°	78,000
	60°	65,000
	75°	94,000
Index of Refraction	(ordinary ray at 24°C)	
	λ = 0.2652 μm	n = 1.8336
	0.5461	1.7708
	1.0140	1.7555
	5.577	1.5864
Optical Transmission	85% at 0.55 μm. Useful transmission from 0.15 to 6 μm. See Fig. II-4.	
Radiation Hardness	No visible color after exposure to gamma radiation of 10^8 rads/hr for 60 minutes at -195°C. No change in transmission above 2.5 μm after exposure to 10^7 rads. No change in broadband transmission after 2.7×10^{15} electrons/cm^2 at 1.2 MeV.	
Electron Range	2.9 mm for 2.0 MeV electrons.	

*Values in this table are taken from various brochures supplied by manufacturers.

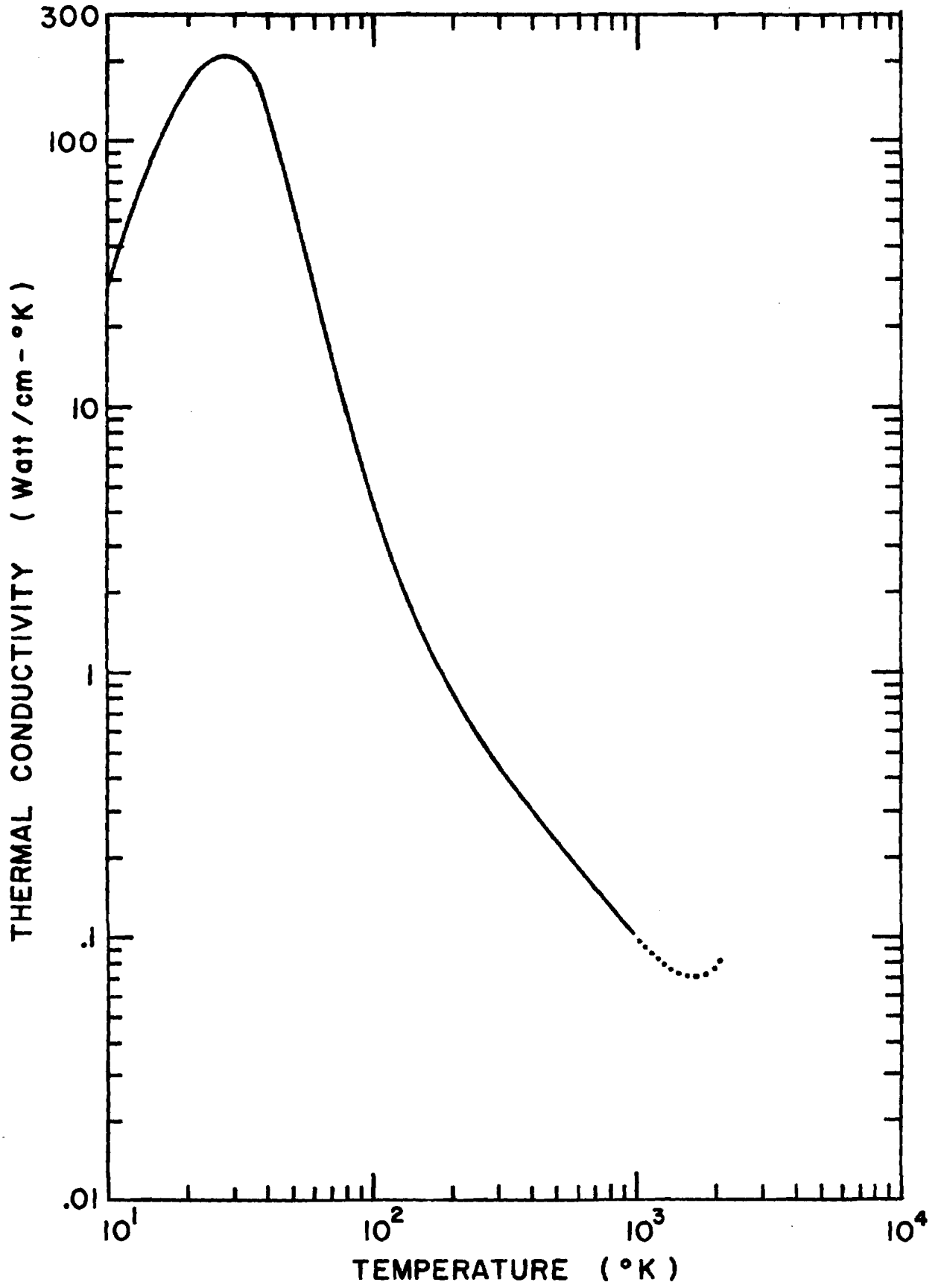


Figure II-2. Temperature variation of the thermal conductivity of sapphire. (Reference (57).

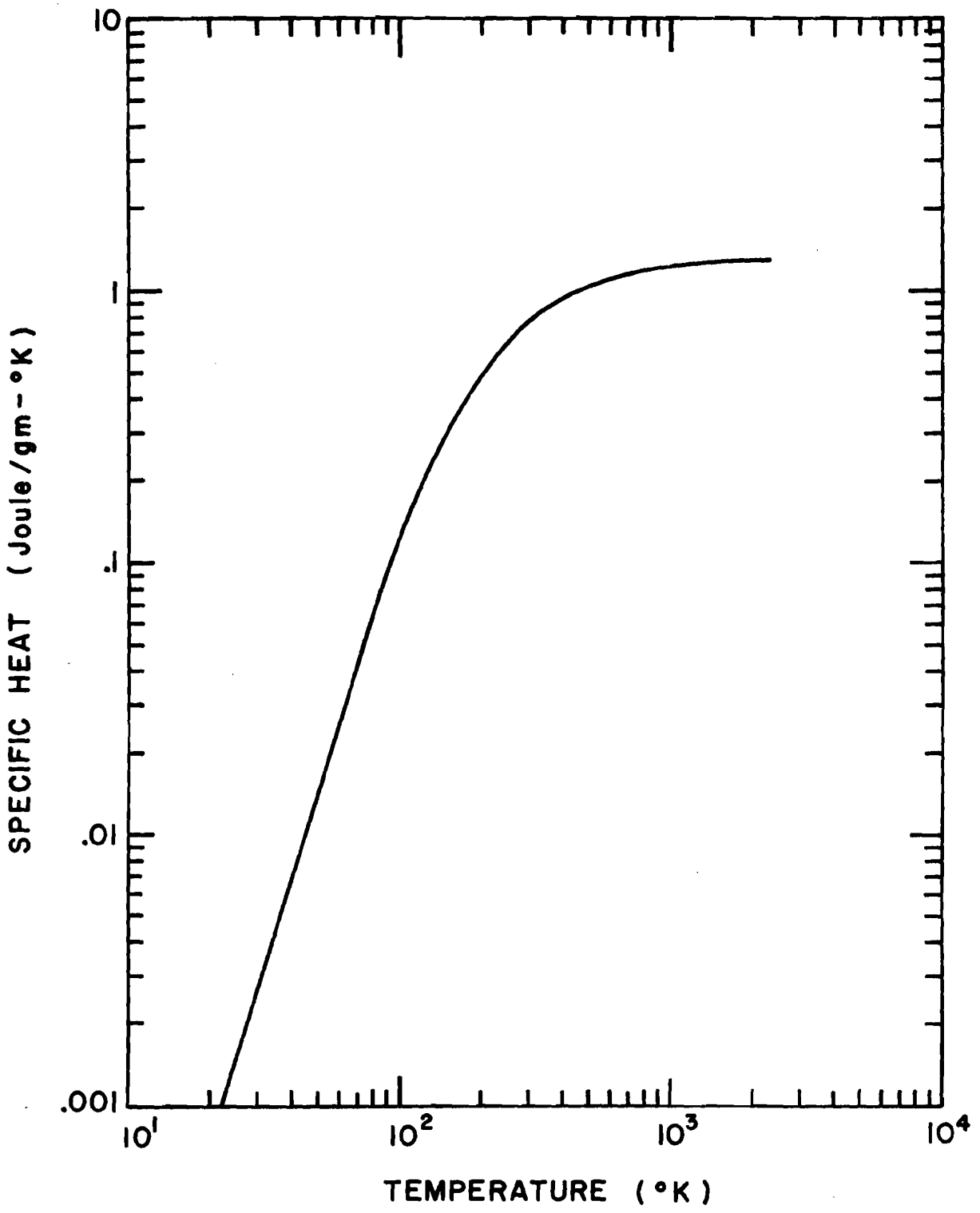


Figure II-3. Variation of the specific heat of sapphire. Reference (58).

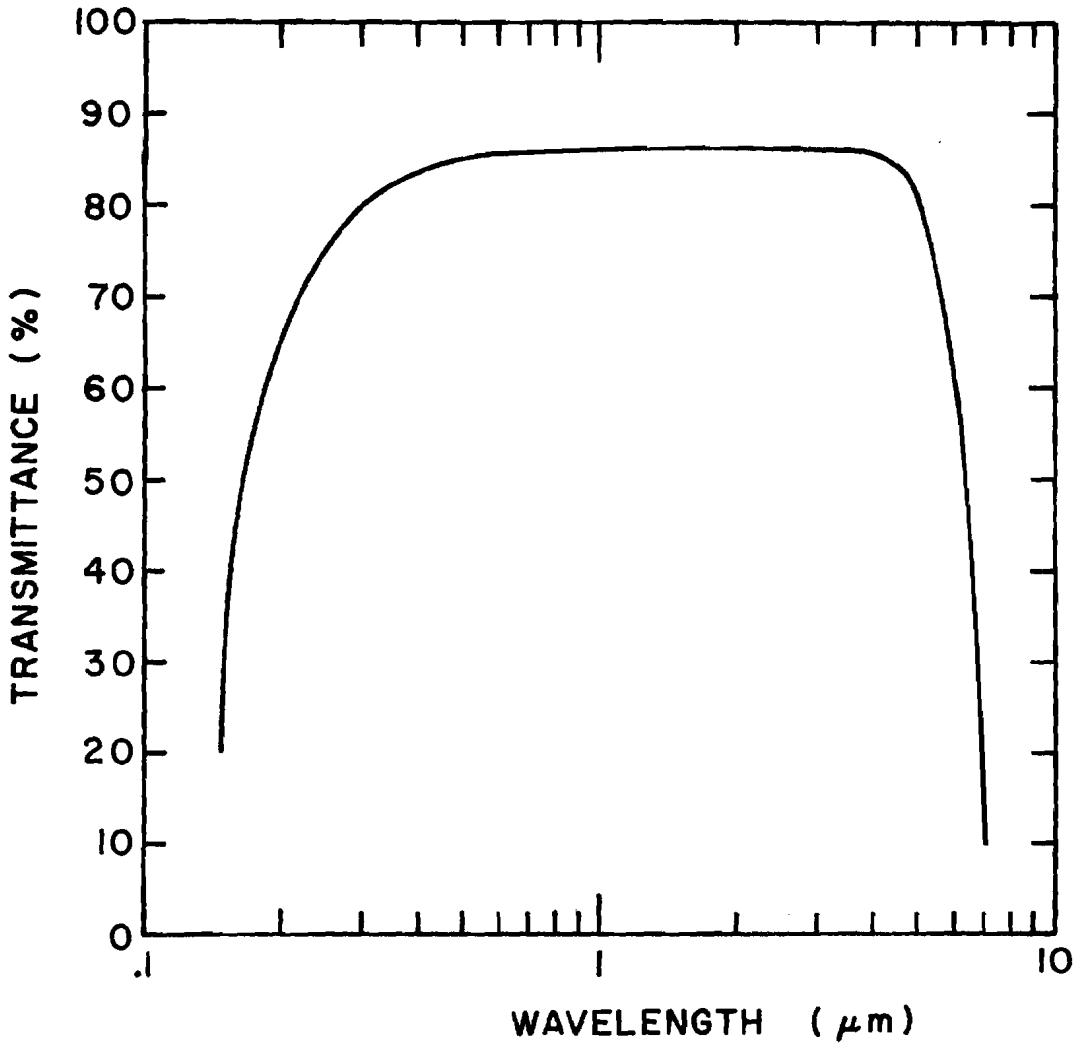


Figure II-4. Transmittance of sapphire as a function of wavelength.

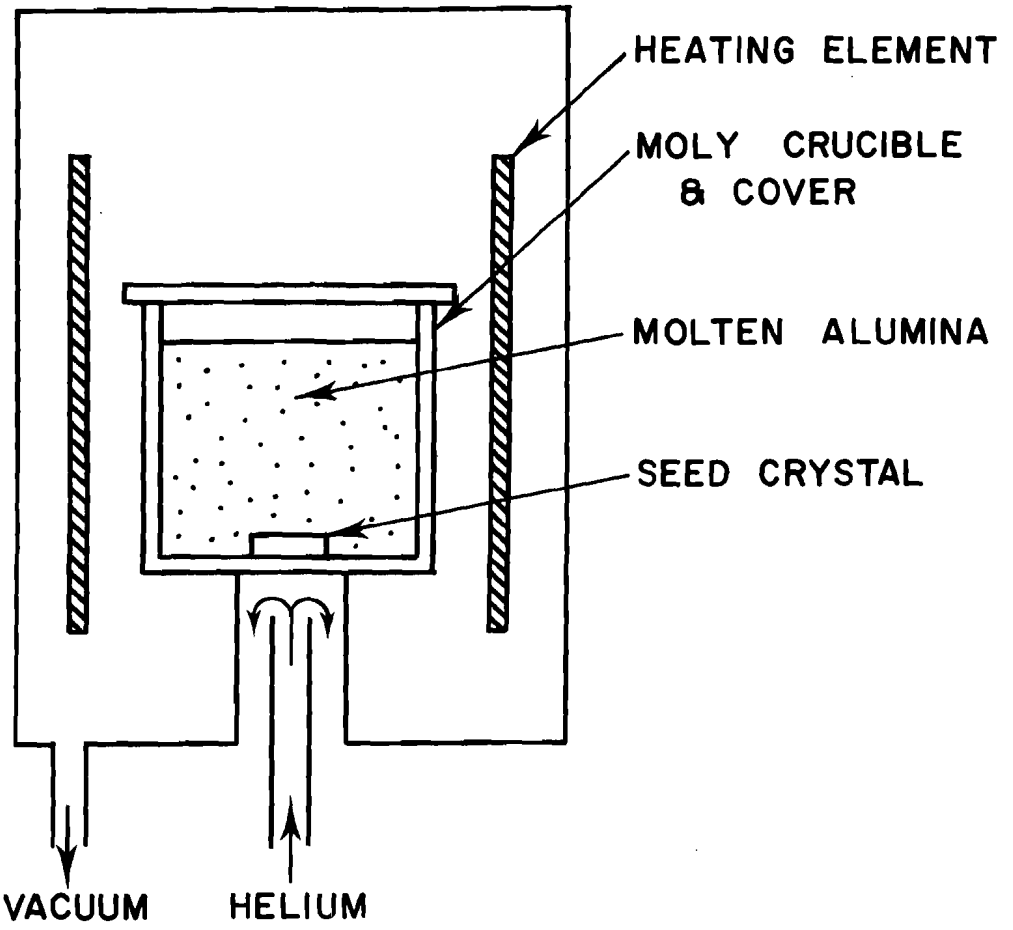


Figure II-5. Heat exchanger method of growing single crystals of sapphire. The helium flow is adjusted to control the growth.

A molybdenum crucible containing a sapphire seed is loaded with sapphire chips and placed inside a vacuum furnace. A graphite resistance heating element raises the temperature of the furnace to 2050°C. The temperature of the melt is controlled by the rate of flow of helium gas through the hollow tungsten crucible support. By reducing the flow of helium the crucible temperature is increased to the point where the charge melts and wets the seed crystal. Then the flow of gas is gradually increased, reducing the temperature and allowing the liquid to solidify on the seed.

There are several advantages to this system. First of all, there are no moving parts. The crucible, seed, heater, etc., are all stationary with respect to each other. Secondly, temperature gradients in the solid and liquid phases can be carefully and independently controlled. Finally, there is no mechanical turbulence in the melt since there is no relative motion, and temperature gradients are in a direction to avoid convection turbulence. This helps to eliminate temperature and concentration fluctuations at the interface, promoting uniform growth and crystal perfection. Dislocation density is less than $10^3/cm^2$.

Company B is new, having been formed only 3 years ago specifically to manufacture sapphire products. Consequently, it does not perform in-house all of the operations required to produce sapphire substrates. Sapphire substrates available from this company have the following specifications: orientation $(1\bar{1}02)$ $\pm$ 1°, reference flat on $(2\bar{2}01)$ $\pm$ 1°, 1/2 inch long; diameter - as ordered, with a tolerance of $\pm$ 5 mils; thickness - 14 $\pm$ 1 mil, less than 1 microinch absolute finish. Prices for 2 inch wafers range from $33.00 in small quantities to $25.00 each in quantities greater than 1000.

The third vendor uses a modification of the Czochralski technique to grow flat ribbons of sapphire. A molybdenum crucible is used, placed inside a carbon susceptor, with the whole apparatus inside a double-jacketed water-cooled quartz chamber. RF energy at 450 kHz is the heat source. Argon is used as a cooling gas. The apparatus is shown in Fig. II-6.

Also shown in the figure is the molybdenum die which defines the shape of the pulled crystal. The method is a refinement of an earlier

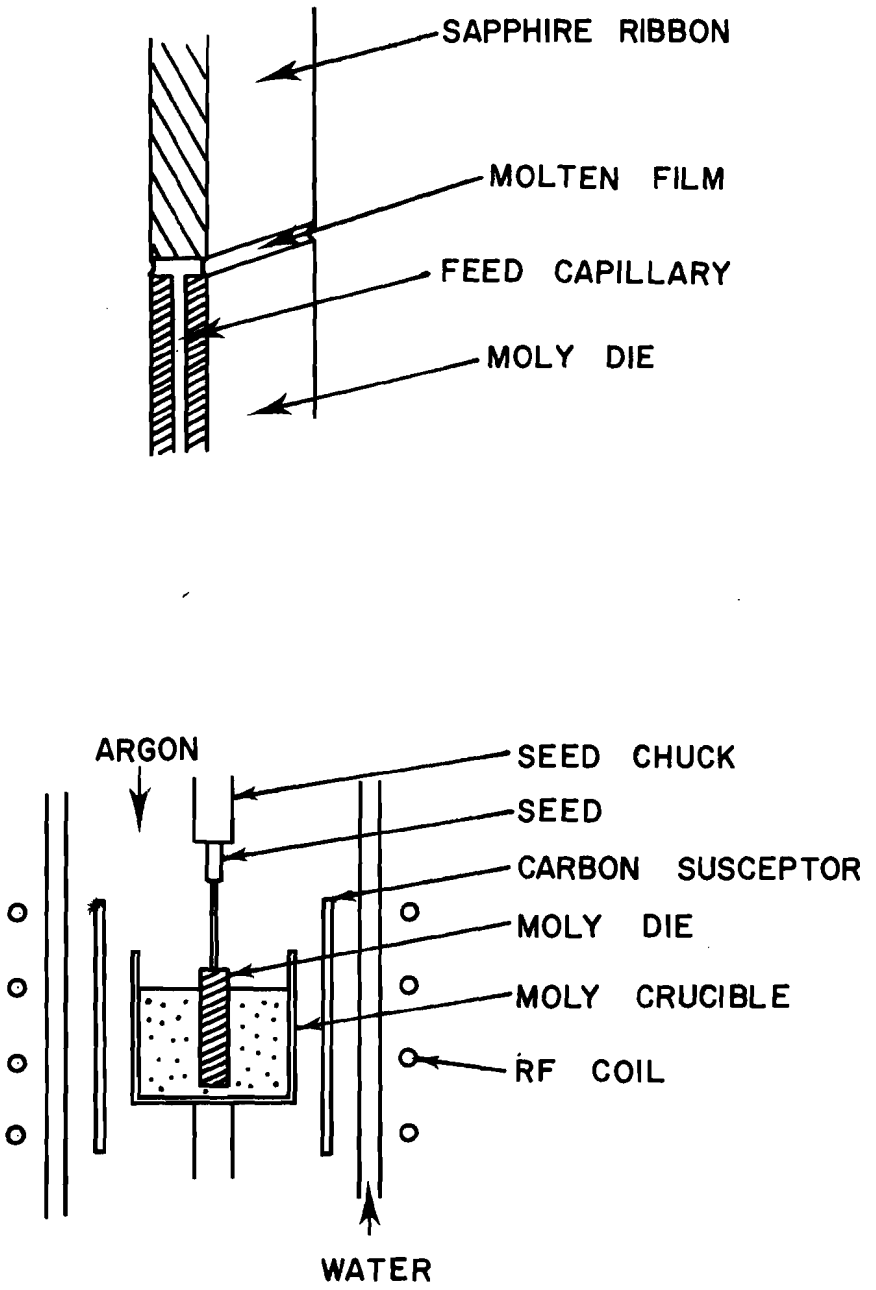


Figure II-6. Modified Czochralski method for growing sapphire ribbons. The molybdenum die defines the shape of the crystal.

technique which used a floating die with an orifice of the appropriate shape. One drawback of the earlier method is that the die is lowered as material is withdrawn, hence the solid-liquid interface continuously moves to a new area of the furnace where temperatures and temperature gradients are different. The feature of the capillary method is that fresh material is brought to the top of the structure by capillary action and is independent of the level of the charge in the crucible. Once the liquid reaches the top of the capillary it flows out over the top surface of the die. The size and shape of the plane surface of the die determine the size and shape of the finished ribbon. Crystals of complex cross-sectional shape can be readily fabricated by this technique.

Stable growth of high quality crystals by the ribbon method requires careful control of conditions at the interface. The main requirement is that there be a monotonic temperature gradient across the solid-liquid interface. This means that there must be a continuous flow of heat from the crucible to the solid crystal. Since the heat is removed rather slowly by conduction through the crystal and by radiation, the crystal pull rate is somewhat slower than could be otherwise achieved, a few cm/minute. If the proper temperature gradients are not maintained, growth becomes unstable, with dendritic growth and bubble formation.

This process has been licensed to four other companies to produce synthetic sapphire, including one company in the United States which is a producer of SOS circuits. Flat ribbon growth has interesting implications in regard to SOS substrates. In the first place the flat ribbons require little grinding or lapping. Secondly, the natural wafer shape is square or rectangular. Unfortunately, most semiconductor houses would need to do some retooling to utilize this new shape, but it has distinct advantages. For one thing, there would be little variation in structure across the wafer since it is grown in a long, continuous ribbon. In the normal Czochralski process there are radial variations which can affect the IC yield, particularly near the edge of the wafer. Also, with a square wafer there is little or no waste at the edges of the wafer. All of these features add up to a possible cost advantage for the ribbon growth process.

Vendor C specifications are as follows for 2 inch wafers: length and width (or diameter), 2 inches $\pm$ 10 mils; orientation, $(1\bar{1}02) \pm 2°$; orientation flat (on round wafers) parallel to $(0\bar{2}21) \pm 2°$; surface, side A scratch free at 70X, B fine grind to 10 microinch or better; bow, 3 mils maximum; taper, 2 mils maximum. Prices range from $35.25 in small quantities to $23.25 in quantities over 5,000.

Even though the crystal quality is somewhat different for sapphire. grown by the various methods there is evidence that the parameters of SOS devices are not affected provided adequate wafer polishing is performed (59). Even Verneuil crystals are satisfactory in this regard.

Three factors are considered to be important to subsequent SOS fabrication: misorientation, strain, and inclusions (26). Misorientations can cause line defects in epitaxial films and act as diffusion channels in downstream processing. Strain can cause excessive breakage during temperature cycling in later fabrication steps. Inclusions affect substrate finish and hence epitaxial film quality.

Testing procedures range from relatively simple ones requiring no special equipment, to elaborate, expensive techniques. Inclusions and bubbles can often be revealed by inspection under good lighting conditions. More sophisticated optical methods using crossed polarizers for birefringence, Schlieren photography, or Twyman-Green double-pass interferometry can detect misorientations and strain. Chemical dislocation etches such as KOH at 300°C are often used to determine dislocation densities. X-ray diffraction techniques such a Laue back-reflection patterns, Schulz-Wei patterns, and lattice parameter measurements can also be used to check for misorientation and strain.

Machining operations to fabricate wafers from solid stock are a difficult part of the processing since sapphire is unusually hard. The normal yield in the slicing operation is about 15 wafers per linear inch of material. Recently, yields of 20 wafers per inch have been announced by using an improved ID diamond saw which has kerf losses of only 25 mils instead of 45 mils with a standard process. In addition, wafer surfaces are much smoother, eliminating the need for a lapping operation. When

perfected, this could result in lower costs for SOS integrated circuits since the machining represents a considerable portion of the total wafer costs.

During the slicing operation deep damage, bow, taper, and cracking are serious problems. The grinding operation can induce surface damage, including microcracks. A two-step polishing process normally follows. One of the peculiarities of wafer fabrication is that surface scratches tend to be filled in by polycrystalline material during these last steps, making the scratches all but invisible. This is decidedly undesirable since the scratches are propagated into the epitaxial film. The same holds true for other surface defects. The object of the final polish is to ensure a flat, damage-free surface. Chips or cracks at the edges of the wafer are also deleterious since they can act as stress risers and cause excessive breakage during circuit processing. Wafer specifications should include a maximum number of nicks along the edge of the wafer.

Correct orientation of the final wafers can be checked by X-ray diffraction. Chemical etches can be used to detect and/or remove surface damage. Optical examination and Berg-Barret topography are also useful. Final inspection is primarily by visual methods, including Nomarski interference contrast. Interestingly enough, the most sensitive indicator of substrate surface quality is the epitaxial process itself.

Final cleaning takes place in several stages. First the wafers are degreased in trichloroethylene, chloroform, methyl alcohol, a hydrogen peroxide solution, or by ultrasonic cleaning in hot detergent. Further chemical cleaning, as by hot sulphuric acid, is sometimes used. After thorough rinsing in deionized water the wafers are spun or blown dry. Substrates are difficult to keep clean, since a high static charge builds up on the highly polished surface and attracts particulates.

Sapphire wafers for SOS substrates have a $(1\bar{1}02)$ plane (r-plane) orientation. Tentative parameters for sapphire substrates for silicon deposition were adopted by the Semiconductor Equipment and Materials Institute, Incorporated, in September, 1975. These are as follows:

Diameter Range	2" ± 10 mils Round	2" ± 10 mils Square on Side
	(51 ± .254 mm)	(51 ± .254 mm)
Flat Location	$(0\bar{2}21)$ ± 2°	Parallel to an Edge
Flat Length	5/8" ± 1/8"	-
	(16 ± 3.0 mm)	
Surface Orientation	On-orientation	$(1\bar{1}02)$ ± 2°
Thickness	13 ± 2 mils	(330 ± 51 μm)
Bow	2 mils max	(51 μm)
Taper	2 mils max	(51 μm)

D. Trends in Substrate Costs

Raw material for the manufacture of single crystal sapphire is somewhat cheaper than raw material for the production of silicon. Fabrication costs for finished wafers are considerably higher, due to the higher melting temperature and the difficulty of machining such a hard material. Present costs are five to ten times as high as silicon, and have not been declining as rapidly as early forecasts predicted. At the present time the standard 2-inch sapphire wafer sells for $15-$35, depending on quantity, versus about $3.50 for 2-inch silicon. 3-inch wafers cost from $30-$80, versus $4-$5 for silicon. 2-inch sapphire with epitaxial silicon is available for $25.00.

Several factors tend to reduce the cost differential in finished circuits. In the first place, SOS devices can be 30% smaller than their CMOS counterparts. Not only does this mean more dice per wafer, but a higher yield as well. Less breakage should be encountered with sapphire wafers, which are thicker and less fragile than silicon. Finally, there are no scrap wafers since rejects can be reclaimed at any stage in the processing simply by stripping all surface materials and regrowing the epi layer. A further advantage of sapphire is that there is no sagging of substrates during processing since the melting point is much higher than the processing temperatures. Silicon wafers greater than three inches in diameter must be carefully supported to prevent warping during diffusion and oxidation.

With the present market conditions, the pricing situation is not likely to change much in the immediate future. Capacity is adequate to meet present demands. Most of today's sapphire is Czochralski grown, a

process which requires large capital outlays. Substantial price reduction will come about only through a larger volume of sales, more aggressive marketing by suppliers, or new innovations in sapphire growing and wafer fabrication.

Vendor C's ribbon growing process could possibly provide dramatic breakthroughs in pricing in the next few years. This company has embarked on a large scale joint venture with another company to develop a multiple ribbon process for growing silicon. If it could be successfully applied to sapphire growth, prices would be lowered significantly.

In the first place, the ribbon process provides sapphire sheets with essentially parallel faces. This is not the case when sapphire must be cut from solid stock. Saw chatter and wander introduces wafer taper and bowing - a problem which becomes more severe with increasing diameter. The presence of bubbles in the ribbon is not a serious problem unless they intersect or come close to the surface; otherwise they have no effect on the epitaxial growth of silicon films.

The high cost of sapphire substrates is cited as one of the main drawbacks to the use of SOS. Although it is true that the cost of substrates (whether silicon or sapphire) is a significant portion of wafer fabrication costs, the picture is complicated by changes in device size and yield. An analysis of this factor will be undertaken in a future report.

In any event, substrate costs will be lowered only by high production volume. Some industry sources estimate that present production rates of 3" wafers (10-30,000/year) must reach 1,000,000/year to bring the price down to $12 - still double the price of silicon. Other predictions, based on improved production techniques have been as low as $0.50 per square inch ($3.50 for a 3" wafer).

FABRICATION

There are two companies that are commercial suppliers of SOS circuits at the present time.* At least half a dozen others are engaged in contractual work or laboratory programs at different levels of effort. The purpose of this section is to describe various aspects of the manufacturing processes utilized by these companies as described in the scientific and trade literature, company publications, or by private communication. Some of the details are proprietary and are not included here. Furthermore, the descriptions which follow are intended to be more or less inclusive, and it must be recognized that diverse combinations of the various processes could exist, even within a single company.

This section is divided into descriptions of the following phases of fabrication: substrate preparation, silicon epitaxial deposition, and circuit fabrication methods.

A. Substrate Preparation

In principle, incoming wafers should be inspected for scratches, edge chips, crystallographic misorientation, etc. This is not worthwhile, however, and companies are willing to accept a 5% or so loss in wafers after epitaxial deposition. It is generally acknowledged that the growth of high-quality epitaxial layers of silicon does require careful substrate preparation. The as-received wafers contain a work-damaged surface layer which, if not removed, results in poor silicon films with high dislocation densities and poor electrical characteristics. In particular, the low lifetime of minority carriers is a major concern. It has been shown that work hardening of sapphire occurs by plastic flow during machining operations (27). In addition to work hardening, work damage can extend into the material to a depth two to three times the grit diameter of the final polishing agent. Surface scratches are especially deleterious. They will cause preferential silicon

*In the fourth quarter of 1974 there were four. One ceased commercial operations due to technical difficulties, and the other due to financial difficulties.

growth with the result that scratches propagate to the surface of the silicon film. Polycrystalline growth may also occur, or diffusion spikes in later processing.

Substrates are first degreased and cleaned with standard agents. Caro's acid can be used (5 parts H_2SO_4, 1 part 30% H_2O_2, 1 part H_2O). Further chemical conditioning by hot phosphoric acid or borax has been tried, with good results (5,28,29). A commercial chemical etch-polish, consisting of colloidal silica suspended in an ammonium hydroxide solution, is widely used. Some of the other etches and conditioners which have been tried are: potassium hydroxide, berylium oxide, vanadium pentoxide, lead fluoride, hydrochloric acid, potassium persulphate, sulphur hexafluoride, and chlorotrifluoromethane. Most of these have not been developed to the point of producing high-quality surfaces for production purposes.

One of the earliest methods of preparing substrates is still the most widely used - the technique of hydrogen etch and anneal (13). Prolonged heating in pure hydrogen at 1100-1500°C has been found to result in high quality epitaxial films. This method is compatible with other steps in the epitaxy process, and hydrogen is not an electrically active species. Sapphire is removed via the chemical reaction

$$Al_2O_3 + 2H_2 \rightarrow Al_2O + 2H_2O$$

This method is not without its drawbacks. Work-damaged sapphire can be rapidly removed, with the formation of microscratches. The reaction is relatively slow, etching away about 2 μm/hour. Some manufacturers feel that the procedure increases the density of fast states at the interface and increases autodoping of the silicon films by aluminum from the substrate. Nevertheless, the method has found wide acceptance in the past decade.

Advances in surface polishing have made substrate preparation less important in recent years. A 10-minute hydrogen prefire at 1150°C is considered to be entirely sufficient by one manufacturer, while another uses only a high temperature air anneal. Still another manufacturer has developed a process which eliminates high temperature hydrogen prefiring altogether, yet produces epitaxial silicon with excellent mechanical and

electrical characteristics.

B. Epitaxial Growth

The orientation of the silicon films depends on the orientation of the underlying substrate. Some of the relationships reported in the literature are listed in Table III-1. The density of silicon - silicon dioxide surfaces states using silicon substrates is in the ratio 3:2:1 for (111), (110), and (100) silicon. Therefore (100) surface silicon is the most desirable for MOS devices and (100) silicon on $(1\bar{1}02)$ sapphire is used exclusively.

Table III-1. Silicon-Sapphire Orientations

Silicon	Sapphire
(100)	$(1\bar{1}02)$ (r-plane)
(111)	(0001), $(11\bar{2}3)$, $(11\bar{2}4)$, $(11\bar{2}0)$

It is not intuitively obvious how a single crystal face-centered cubic material can be grown on a hexagonal lattice at all, which is one reason why spinel was utilized so extensively in early work. A great deal of attention has been focused on such things as lattice mismatch, lattice strain, etc. Nevertheless, good silicon films of the same orientation can be grown by either chemical vapor deposition or physical vapor deposition (vacuum deposition) under a wide range of growth conditions. The main requirement is that the sapphire surface be oriented to $(1\bar{1}02)$ within 4° for (100) growth (more stringent for other orientations). This in itself indicates that the atomic structure of the host lattice is important. The answer seems to be that the lattice structure of the substrate crystal is modified at the surface boundary in some manner not yet understood. Growth does not proceed by the buildup of monolayers of silicon, however, but by the growth of individual islands which eventually coalesce into a single crystal layer of one crystallographic orientation.

Silicon films have been deposited on sapphire through the hydrogen reduction of silicon tetrachloride and the chlorosilanes, and by the pyrolysis of silane using hydrogen and helium as the carrier gases. Both

vertical and horizontal reactors have been successfully used. It has been reported that silicon deposited on sapphire at 1200°C using a hydrogen/silane source is of one orientation and that the nuclei prior to coalescence were not faceted (30). On the other hand, silicon deposited at the same temperature and rate using hydrogen reduction of trichlorosilane was of more than one orientation and faceted. The nucleation was three dimensional in both cases with appreciable reaction with the substrate when using trichlorosilane due to the formation of hydrogen chloride which, with the hydrogen carrier gas, caused etching.

Many experiments have been performed to find the optimum epitaxial deposition conditions. Temperatures in the 1100 - 1200°C range coupled with relatively slow deposition rates of 100 Å/min have been found to produce silicon films with good crystallinity, while in the low temperature range of 850 - 950°C the crystal quality becomes poor. There is also a tendency for the silicon to become (110) rather than (100) (31). As deposition temperatures are increased, film quality increases up to a point where chemical reduction of the substrate occurs and autodoping becomes a serious drawback. Polycrystalline silicon is deposited when substrate reduction occurs because of the additional surfaces that are exposed during the reduction of the substrate. At still higher temperatures (1250°C) silicon leaves the sapphire surface in gaseous form.

Early stages of film growth are complicated by the formation of oxides of silicon and aluminum. Two reactions which can occur are

$$2Si + Al_2O_3 \rightarrow Al_2O + 2SiO$$

and

$$2H_2 + Al_2O_3 \rightarrow Al_2O + 2H_2O$$

Oxygen and aluminum complexes are deleterious to good film growth and it is important to eliminate even trace amounts of oxygen from the reactor and source gases. Below about 1000°C SiO is stable and hinders good film growth. Above 1100°C both SiO and Al_2O leave the surface in gaseous form. Al_2O from the back surface as well as the front surface can lead to autodoping problems

in the film. Once the surface has been sealed off by coalescence of the growing nuclei, growth conditions are much more ideal. It is in the first several hundred Angstroms of film growth that poor crystallinity and reaction products lead to the poor electrical characteristics of the film. The best growth method would tend to minimize this problem.

The pyrolysis of silane with hydrogen as the carrier gas is used almost exclusively at present. The useful range of temperature and deposition rate has been determined to lie in the range 950 - 1250°C and 0.1 - 10 μm/min. Surface preparation has developed to the point where low deposition temperatures of 950 - 1000°C at deposition rates of 4-5 μm/min are capable of routinely producing high quality films. One technique which is used is a rapid initial growth to seal off the surface, followed by a slower growth rate to produce high quality silicon.

One group of researchers has developed a technique whereby a heavily doped layer is initially grown to act as a barrier to outdiffusion of aluminum and other contaminants (10). Others have experimented with helium as the carrier gas and report that lower deposition temperatures and films with less contamination can be produced (28,31). A vacuum deposition technique for sputtering silicon has also been described (32,33). Further investigation of these methods is warranted, since the epitaxial film growth still remains a critical step in SOS circuit fabrication.

With the deposition rates currently being used thickness variations in the silicon film can be a significant problem. Reactor geometry, temperature variations, and gas flow rates are all extremely important for uniform deposition. With a carefully designed reactor, film thickness can be controlled to within 10%. A high degree of uniformity is required in later processing steps such as island etching and metalization.

A great deal of effort has been expended on determining the electrical properties of the silicon films. Hall effect, resistivity, and van der Pauw measurements are commonly used. Carrier concentrations, mobilities, and lifetimes are the parameters which determine device performance and hence are of major concern. These parameters are dependent on the substrate quality and preparation, the silicon thickness, doping density, deposition temperature, and deposition rate. Electrical properties can also change

during subsequent processing steps; the oxidation cycle is particularly important in this respect.

Figures III-1 through III-4 show the effects of film thickness, carrier concentration, and oxidation treatments on the carrier mobilities in epitaxial (100) silicon films on $(1\bar{1}02)$ sapphire substrates. Generally speaking, Hall mobilities between 25% and 50% of those found in bulk-MOS transistors are typically obtained in silicon on sapphire films.

Film properties are determined by three sublayers: the substrate-film interface region, the film-SiO_2 interface region, and the region between these two layers. Early growth conditions affect the quality of the substrate-silicon layer. Chemical reactions between the substrate, silicon, doping gases, etc. cause aluminum and reaction compounds to be incorporated into this first layer. Aluminum autodoping is not a problem at deposition temperature below 1050 °C, and fast initial growth of the film quickly seals off the substrate and halts the accumulation of oxygen complexes in the interface layer. Complete coverage of the substrate occurs after a few hundred Angstroms of film growth. This first grown layer is heavily disordered, with a high concentration of impurities and neutral complexes, and is thought to be responsible for the low lifetime of carriers.

Figures III-3 and III-4 show that the properties of the silicon film change after the film has been oxidized. The change in carrier concentration (increase in electrons, decrease in holes) is caused by the absorption of aluminum into the thermal oxide (35). It has also been proposed that oxide charges can deplete the silicon - silicon dioxide interface layer, resulting in a decrease in average carrier mobility in lightly doped films (36).

It is not yet understood why the electrical properties of hetero-epitaxial films differ from ordinary bulk silicon properties. One would expect that as the film is made thicker the effect of the two interface layers would become less important. Furthermore, one would expect the silicon between the interface layers to be similar to that of homoepitaxial films. Instead, one finds that the average mobility in thick films still does not approach that of bulk material. Because low carrier lifetimes and reduced mobilities affect the performance of SOS devices, this whole area merits further investigation.

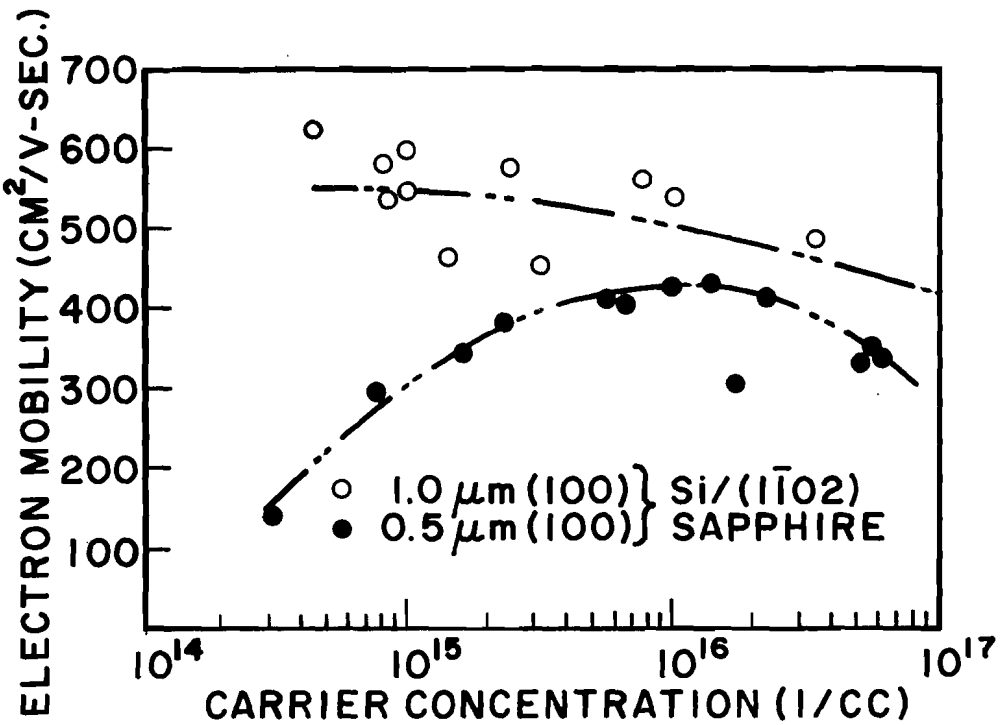


Figure III-1. Hall mobility of electrons in silicon on sapphire. (After Cullen and Corby, reference 9.)

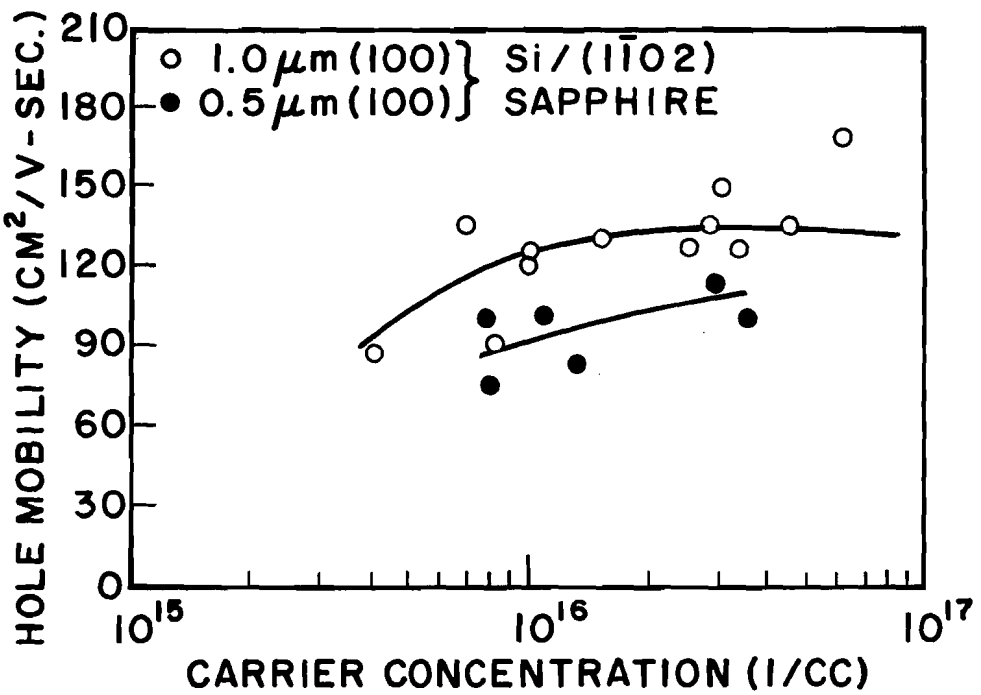


Figure III-2. Hall mobility of holes in silicon on sapphire films. (After Cullen and Corby, reference 9.)

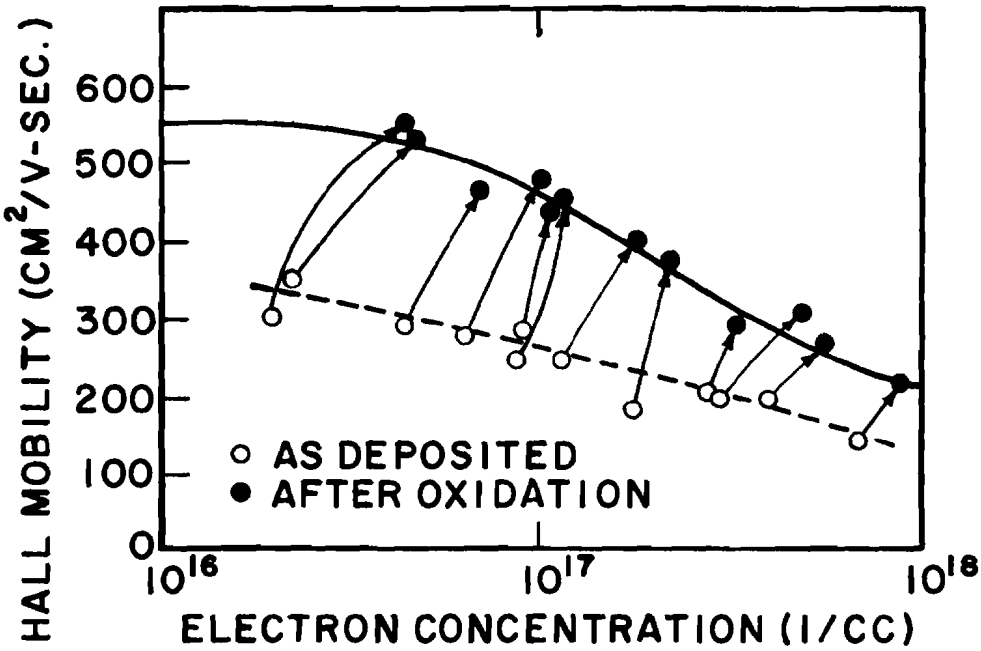


Figure III-3. Change in electron mobility and concentration due to oxidation of film. (After Dumin and Robinson, reference 34.)

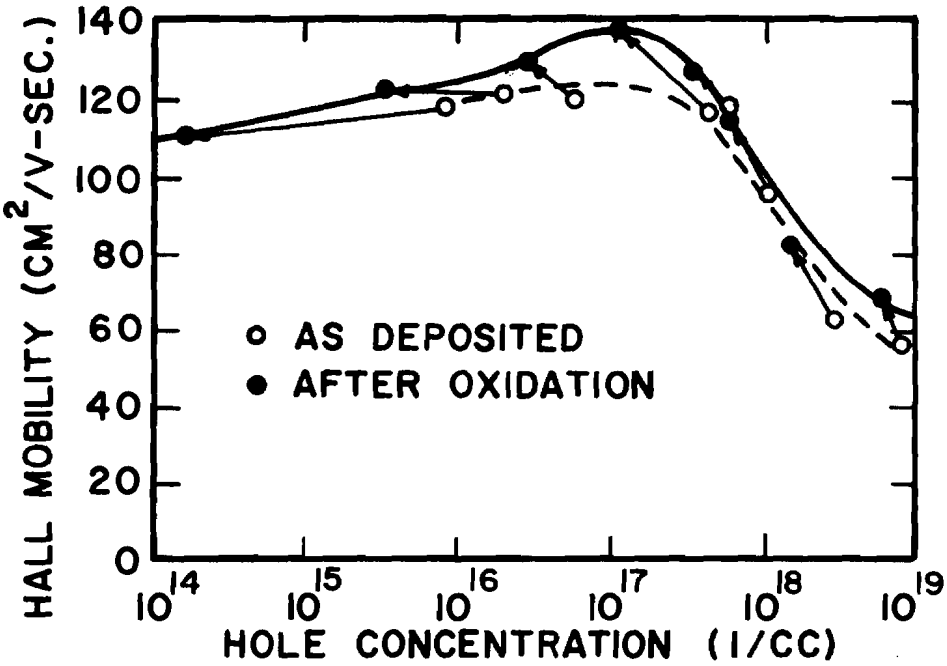


Figure III-4. Change in hole mobility and concentration due to oxidation of film. (After Dumin and Robinson, reference 34.)

C. Fabrication Methods

Epitaxial film deposition is a crucial step in the fabrication process. Many manufacturers have encountered difficulties in growing their own films and purchase epitaxial wafers directly from suppliers, especially during early SOS development programs. In the subsequent steps of circuit fabrication a variety of methods are used, sometimes more than one at the same time by the same manufacturer. A description of the various techniques is given below.

1. Device Structures Available

Figure III-5 shows the various devices that can be made with silicon on sapphire technology. The fabrication of any one or a combination of any of the MOS devices is easily accomplished with standard SOS processing techniques. The bipolar structure is relatively unimportant, and will be discussed only briefly. It is fabricated with four epitaxial layers to eliminate diffusion control requirements that are normal in ordinary bipolar transistors. Enhanced diffusion along stacking faults and other defects creates nonplanar junctions, emitter-collector shorts, etc. The extremely low minority carrier lifetime of 1 ns or so makes bipolar devices impractical at the present time.

All forms of the various MOS structures are variations of the complementary pair, and detailed fabrication processes will be described only for complementary metal gate and silicon gate devices. Deep-depletion MOS transistors (14) can be obtained merely by changing the starting material and/or a dopant source.

2. Double Epitaxial Metal Gate Process

High quality n and p-type films can be sequentially deposited on the same sapphire substrate. In fact, depositing and removing a silicon layer, then redepositing a second layer improves the quality of the second layer because of the impurity gettering action of silicon.

Referring to Fig. III-6, the complementary structure can

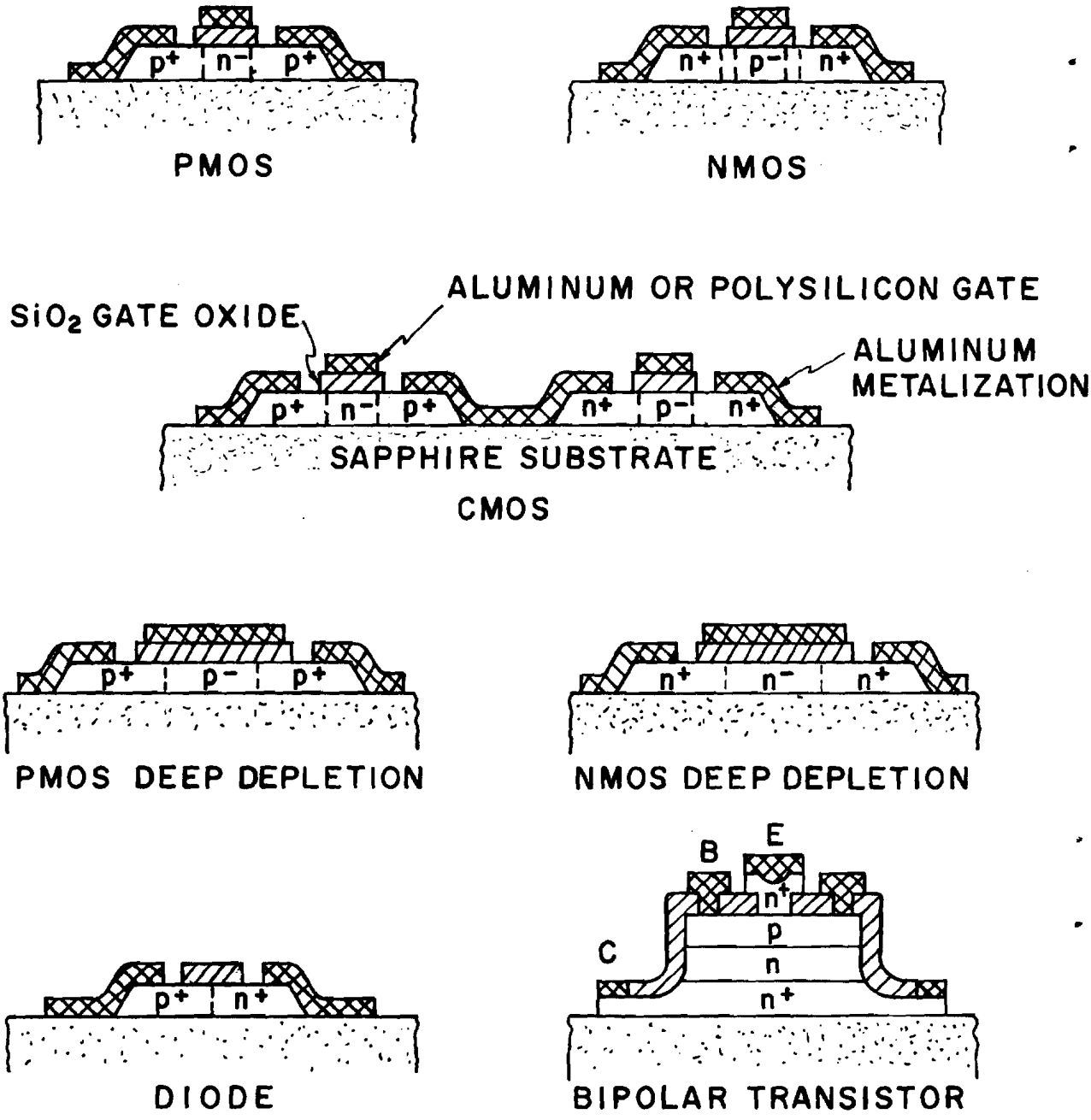


Figure III-5. Devices which can be fabricated with SOS technology.

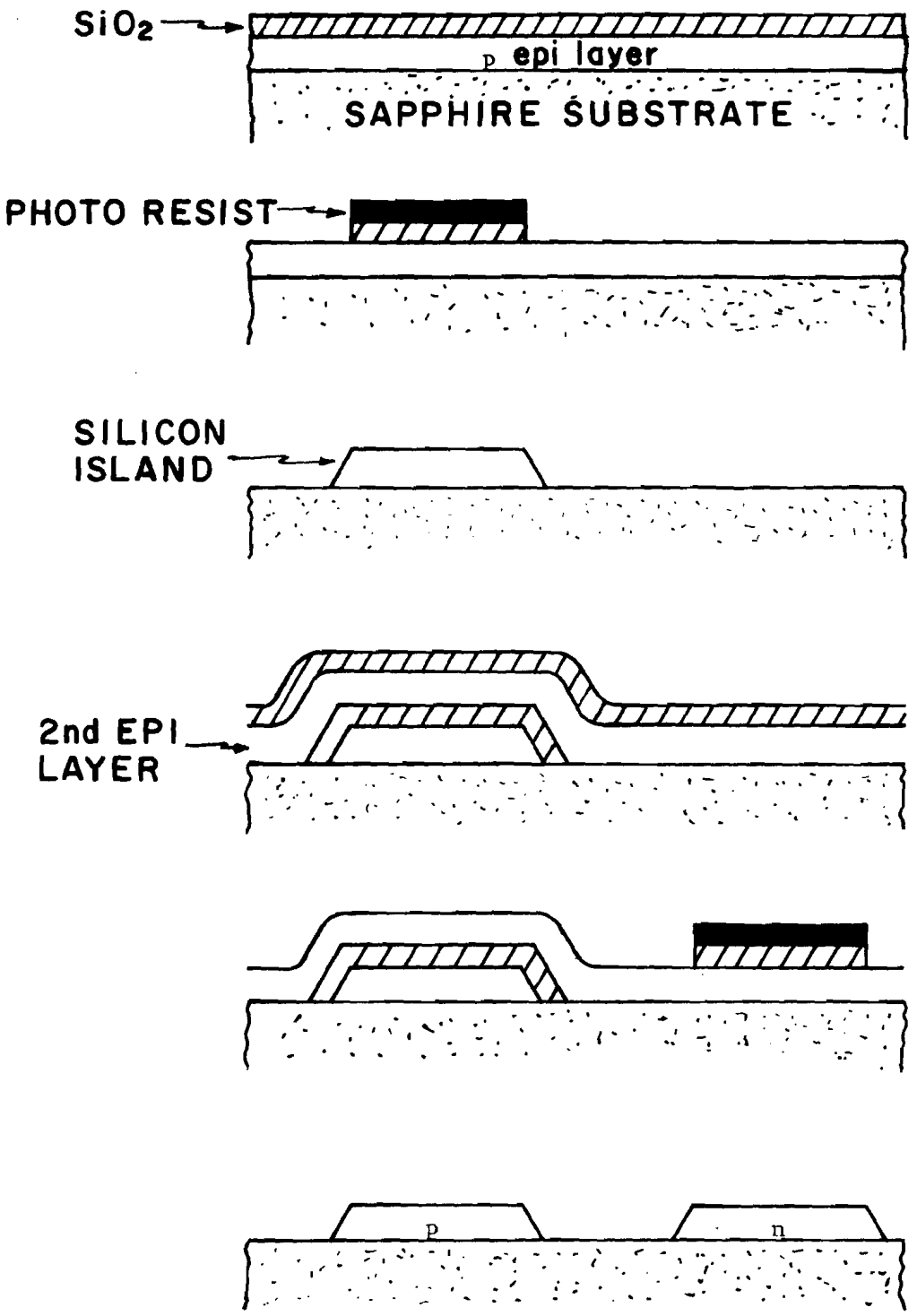


Figure III-6. Fabrication sequence for the double epi metal gate process.

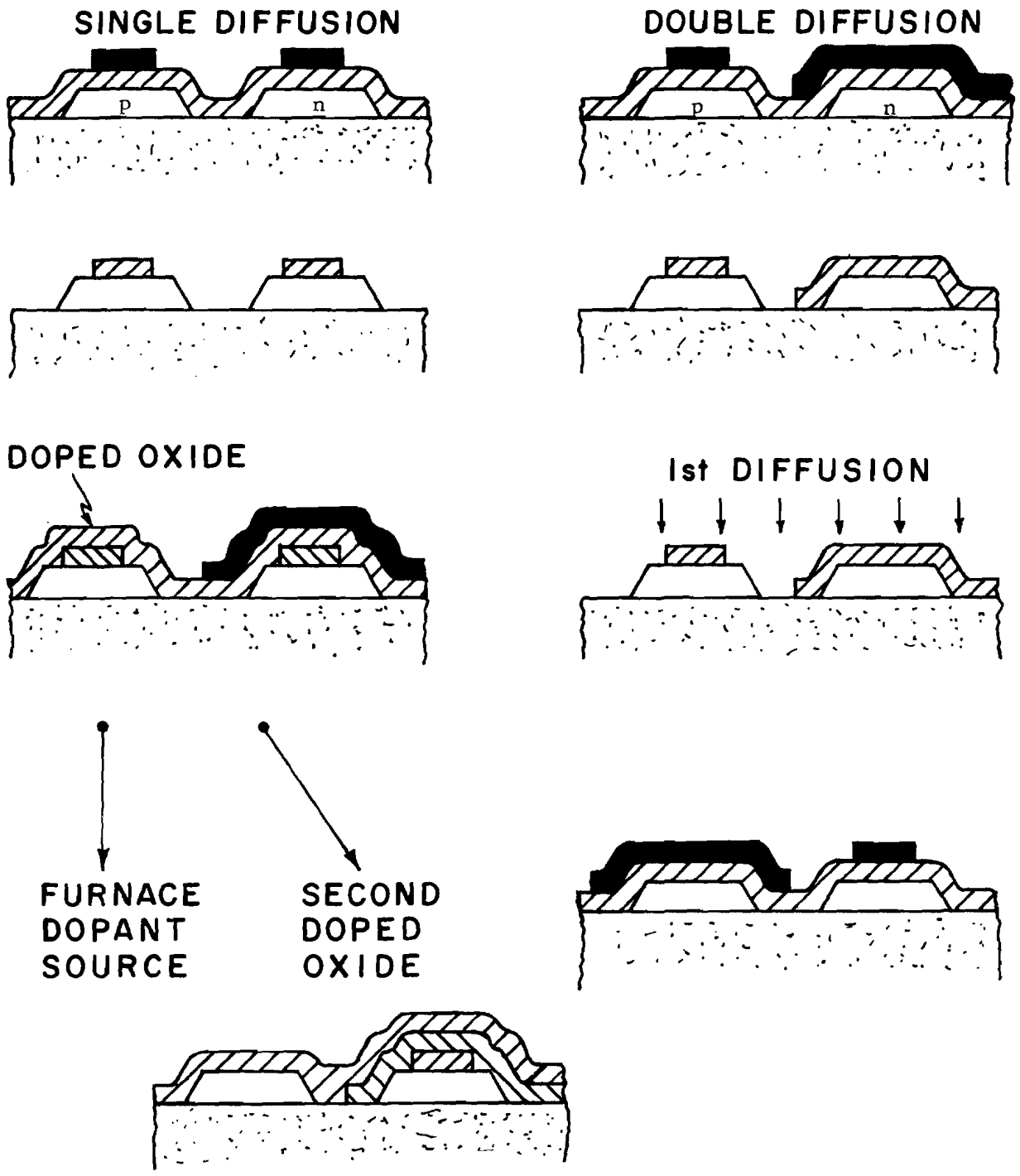


Figure III-6. Continued

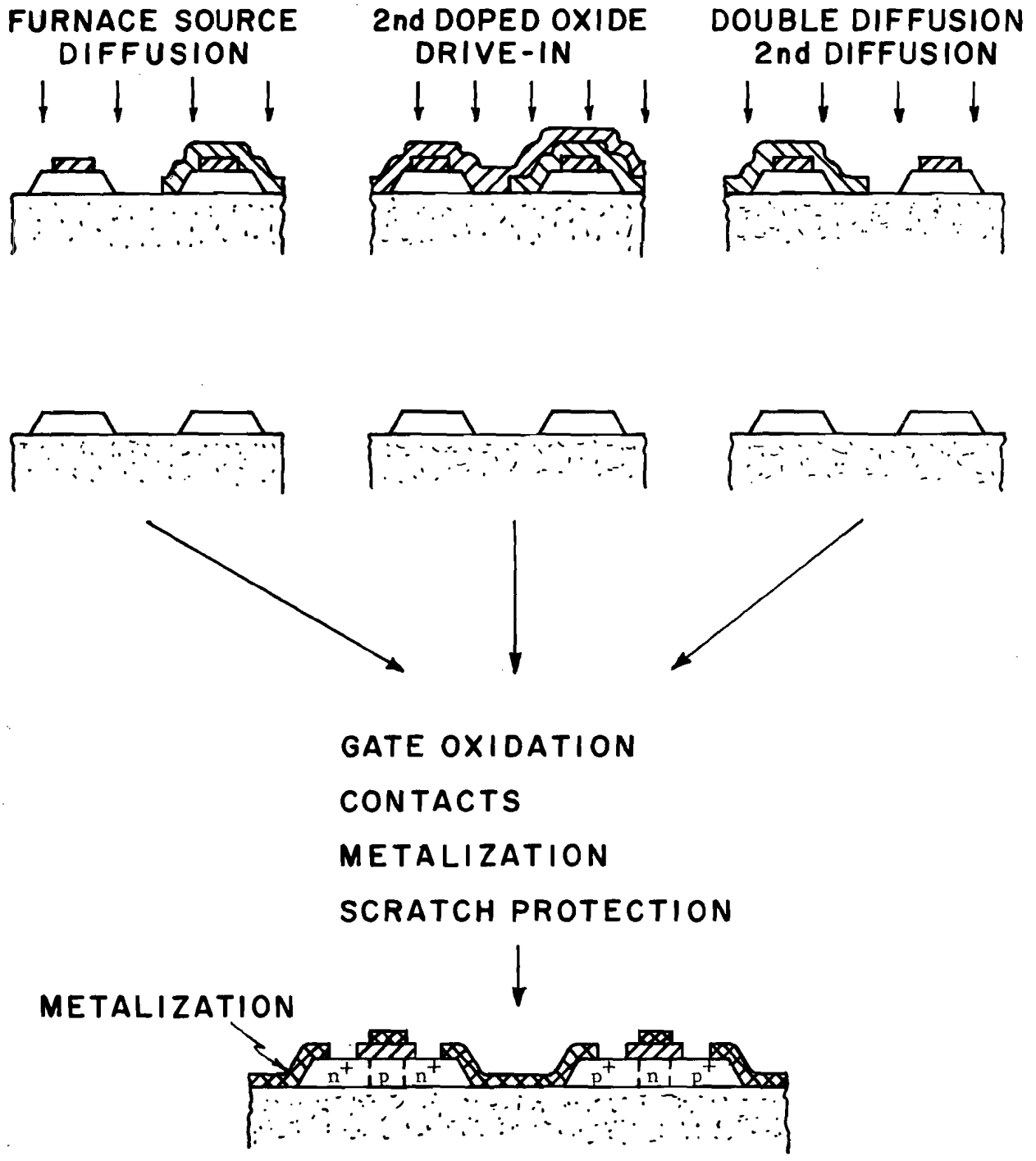


Figure III-6. Continued

be achieved through the following process steps.

a) Epitaxial deposition of either n or p-type silicon
b) Thermal oxidation
c) Photolithographic process to define silicon islands
d) Oxide etch to form silicon etch mask
e) Silicon etch
f) Removal of oxide mask
g) Thermal oxidation
h) Epitaxial deposition of opposite conductivity type
i) Repetition of steps c) through f) to define second type of silicon islands

Isolated n and p-type complementary islands now exist. The process continues from this point with either of two diffusion techniques, the single or double diffusion.

Single Diffusion

j) Pyrolytic deposition of oxide
k) Photolithographic delineation of transistor channel areas
l) Oxide etch to form diffusion mask over channel areas
m) Pyrolytic deposition of either phosphorous or boron doped oxide
n) Photolithographic delineation to leave doped oxide only in selected source and drain areas
o) Doped oxide etch
p) At this point, a choice for the opposite dopant source can be made.
   i) Furnace solid or gas dopant source
   ii) Second layer of doped oxide
q) Drive-in diffusion to form source and drain areas
r) Removal of all oxides

The wafers are now ready for the gate oxidation cycle.

Double Diffusion

j) Pyrolytic deposition of oxide

k) Photolithographic delineation of one transistor type gate area

l) Oxide etch to form diffusion mask over channel areas

m) Furnace diffusion operation to form one conductivity type source and drain areas

n) Removal of all oxides

o) Repetition of steps j) through n) to form source and drain areas of complementary transistors

The product is now ready for gate oxidation, which will be discussed in a later section. Following gate oxidation, the pre-ohmic etch, metalization, and scratch protection steps are standard to the industry and will not be discussed.

The total number of masks required is 7, including scratch protection.

3. Single Epitaxial Metal Gate Process with Ion Implantation

Conversion of the original epitaxial layer into the opposite impurity type can be accomplished with an ion implantation step, as opposed to the double epi method of growing a second epitaxial film. Much better control of the doping in the channel region can be achieved, with correspondingly better control of threshold voltage in the completed device. Furthermore, distinct photolithographic advantages are realized because a totally planar geometry is maintained up to the silicon island etch, which is done late in the process.

Referring to Fig. III-7, the complementary structure is fabricated in the following steps.

a) Epitaxial deposition of either n or p-type silicon

b) Thermal oxidation (optional)

c) Photolithographic operation to define well areas

d) Ion implantation (with either oxide or photoresist as mask)

e) Thermal oxidation

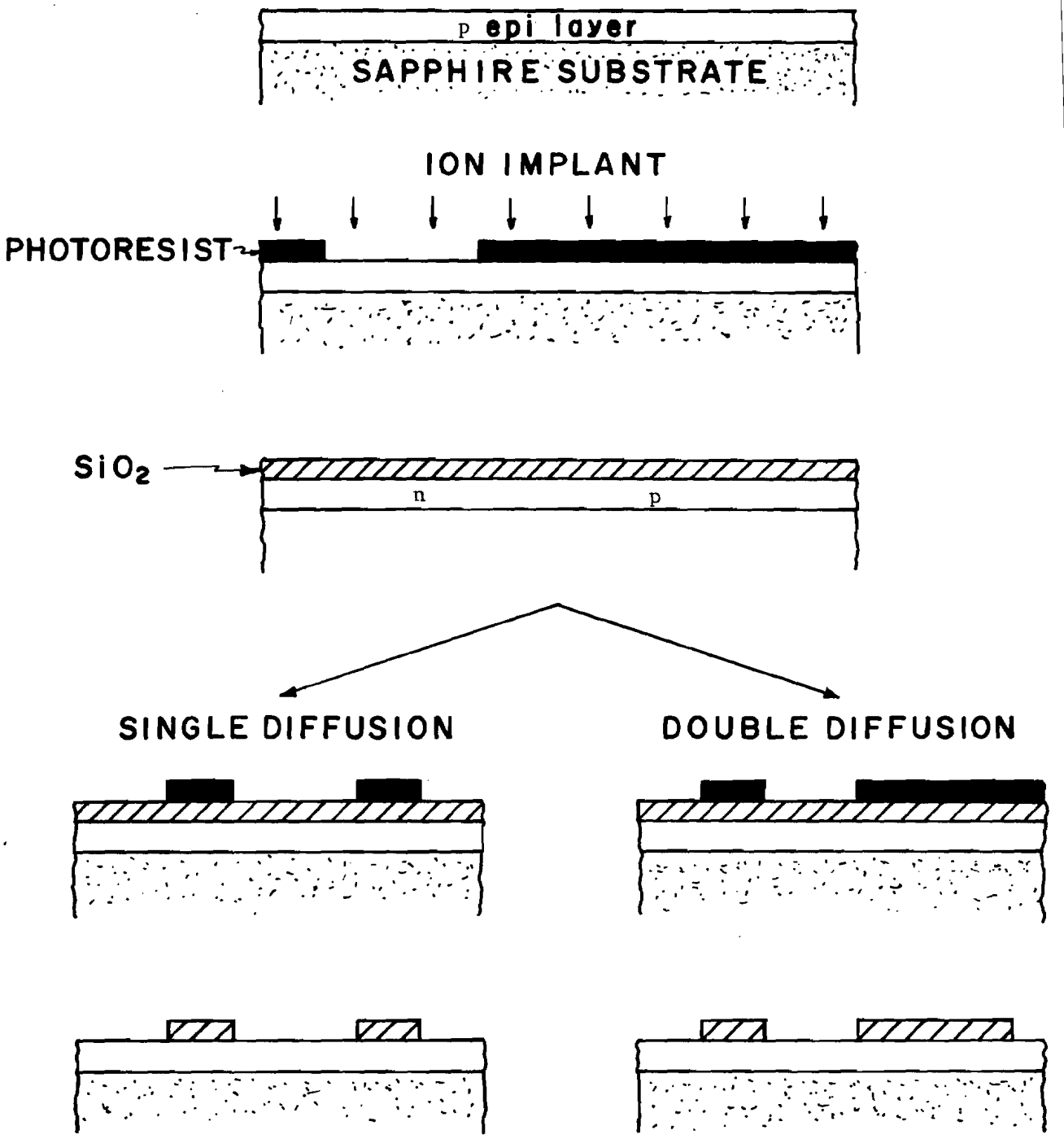


Figure III-7. Fabrication sequence for the single epitaxial metal gate process with ion implantation.

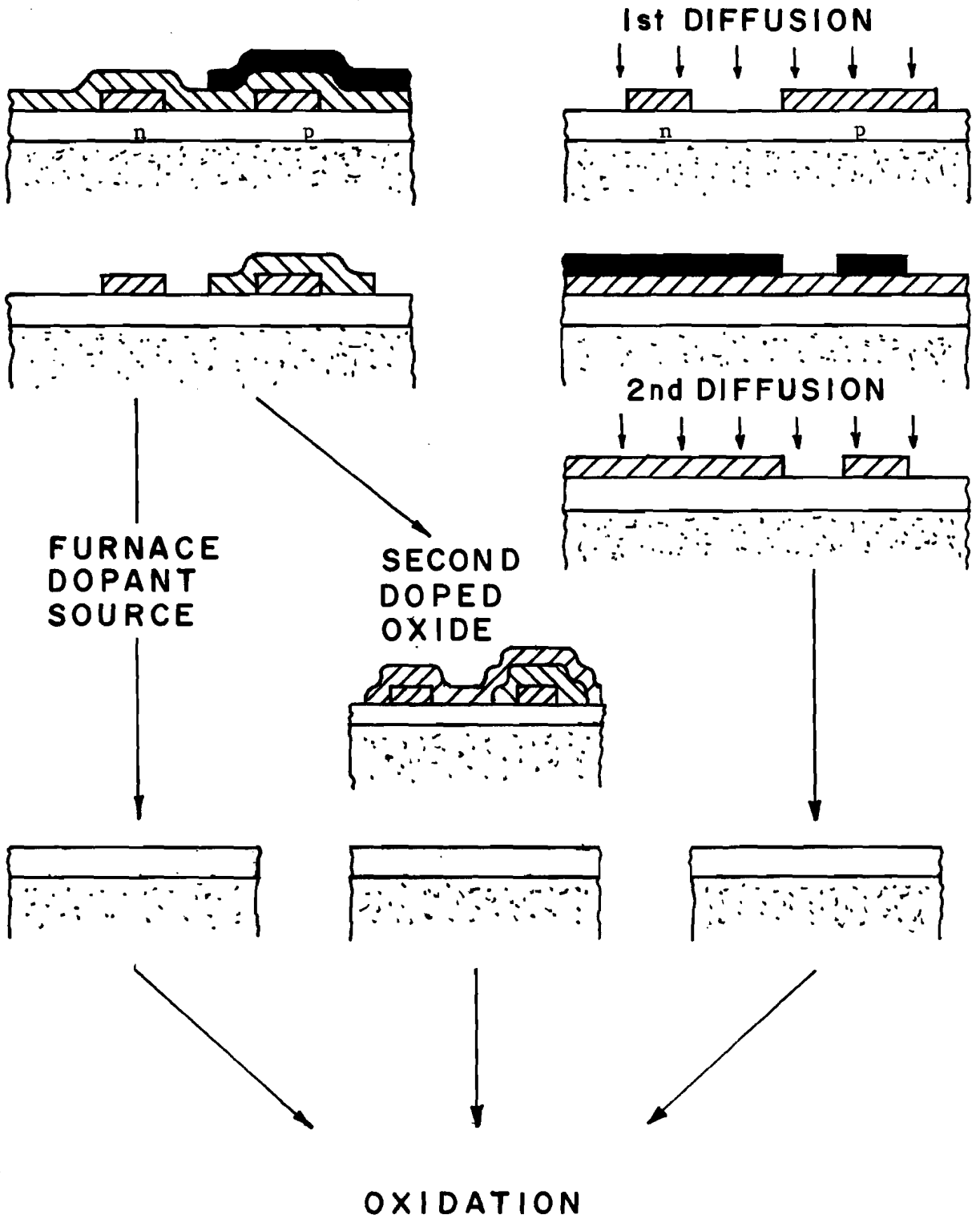


Figure III-7. Continued

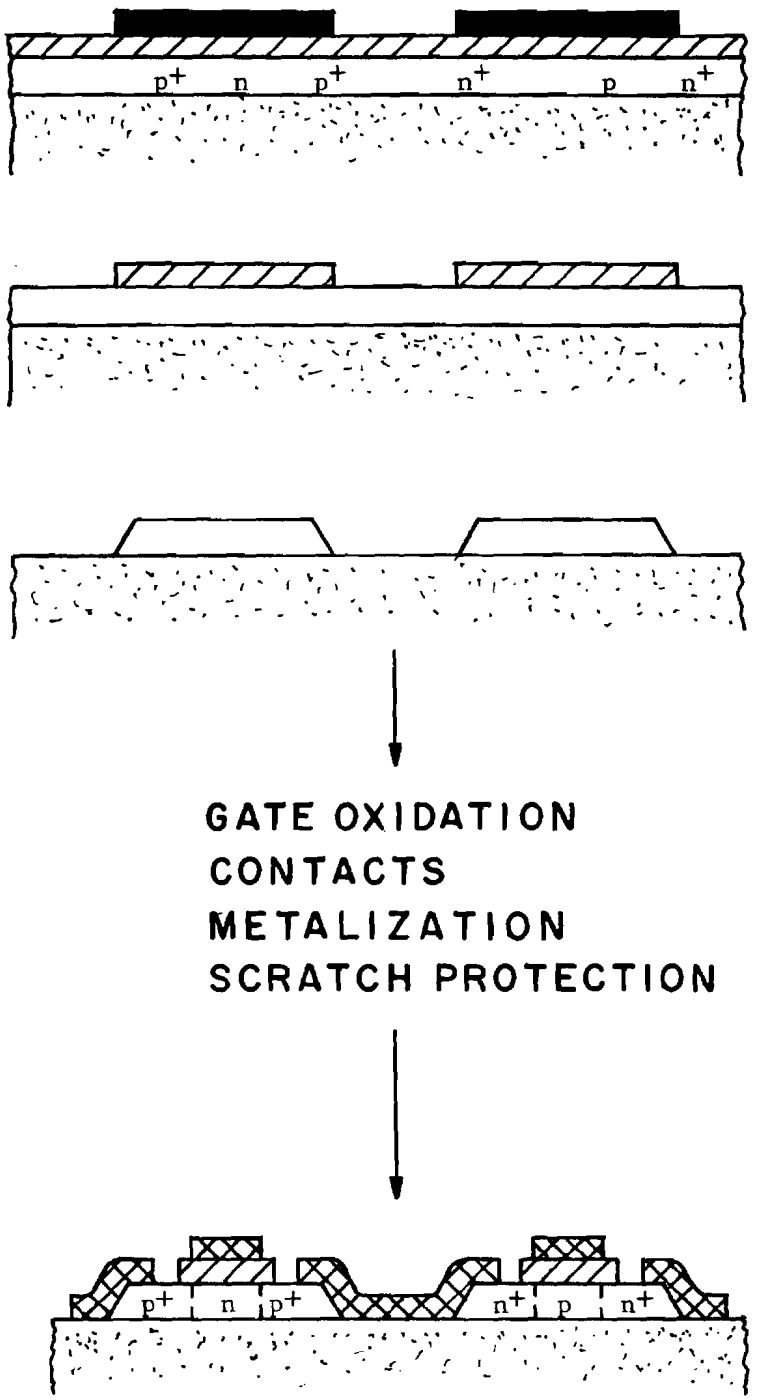


Figure III-7. Continued.

Single Diffusion

f) Repetition of steps k) through r) of single diffusion, double epi process.

Double Diffusion

f) Repetition of steps k) through o) of double diffusion, double epi process.

The following steps are common to either the single or double diffused variations.

g) Thermal oxidation

h) Photolithographic operation to define silicon islands

i) Oxide etch

j) Silicon etch

k) Removal of all oxides

The wafers are now ready for gate oxidation. The total number of required masks is 7.

Ion implantation offers several advantages, as explained earlier, and given the opportunity, most manufacturers would select it rather than a second epitaxial growth or an additional diffusion. Certain advantages can also be realized by implanting the source and drain regions in addition to the channel. The major advantage is that a photoresist layer can be used as a mask. This simplifies the process by eliminating one or two oxidation steps required to define the transistor channel areas. In addition, high temperature predeposition steps are eliminated. Finally, and most important, the technique makes possible a self-aligned gate process with its distinct advantages in terms of lower capacitance, small geometries, less registration problems, etc.

Although ion implantation is now used to a limited extent for source and drain doping, it requires a long time to achieve the heavy doping with presently available beam currents. Crystalline damage and a need for subsequent annealing must also be considered.

4. Planar Metal Gate Process

Distinct photolithographic advantages would be realized if a planar structure could be maintained throughout the total process. A planar surface can in fact be achieved by incorporating a silicon nitride layer and oxidation step for the silicon oxide etch in the double epitaxial process.

The following process steps would be employed, as shown in Fig. III-8.

a) Epitaxial deposition of either n or p-type silicon
b) Deposition of Si_3N_4 (thin layer of SiO_2 grown first)
c) Deposition of an oxide layer, or oxidation of the Si_3N_4
d) Photolithographic step to define silicon islands
e) Etch oxide and Si_3N_4
f) Oxidation (10-12 hours)
g) Removal of oxide and Si_3N_4 used to define island areas
h) Photolithographic step to define channel areas
i) Ion implant of channel areas

Dielectrically isolated n and p-type complementary islands now exist. The process would continue along the same lines as used in the double epitaxial metal gate process beginning with step j), and depicted in Fig. III-6. Care must be taken in the succeeding steps to preserve the field oxide so that a planar surface will be maintained throughout the final operations.

Total number of masks required is 7.

5. Silicon Gate Process

A silicon gate process differs only slightly from those discussed so far. One reason for utilizing a silicon gate is to lower the threshold voltages for low voltage applications. In addition, the polysilicon layer can be utilized for a second layer of interconnects. The most important advantage of the silicon gate technique described below is the self-aligned gate feature. A typical process would consist of the following steps (Fig. III-9).

a) Epitaxial deposition of either n or p-type silicon
b) Photolithographic definition of well areas

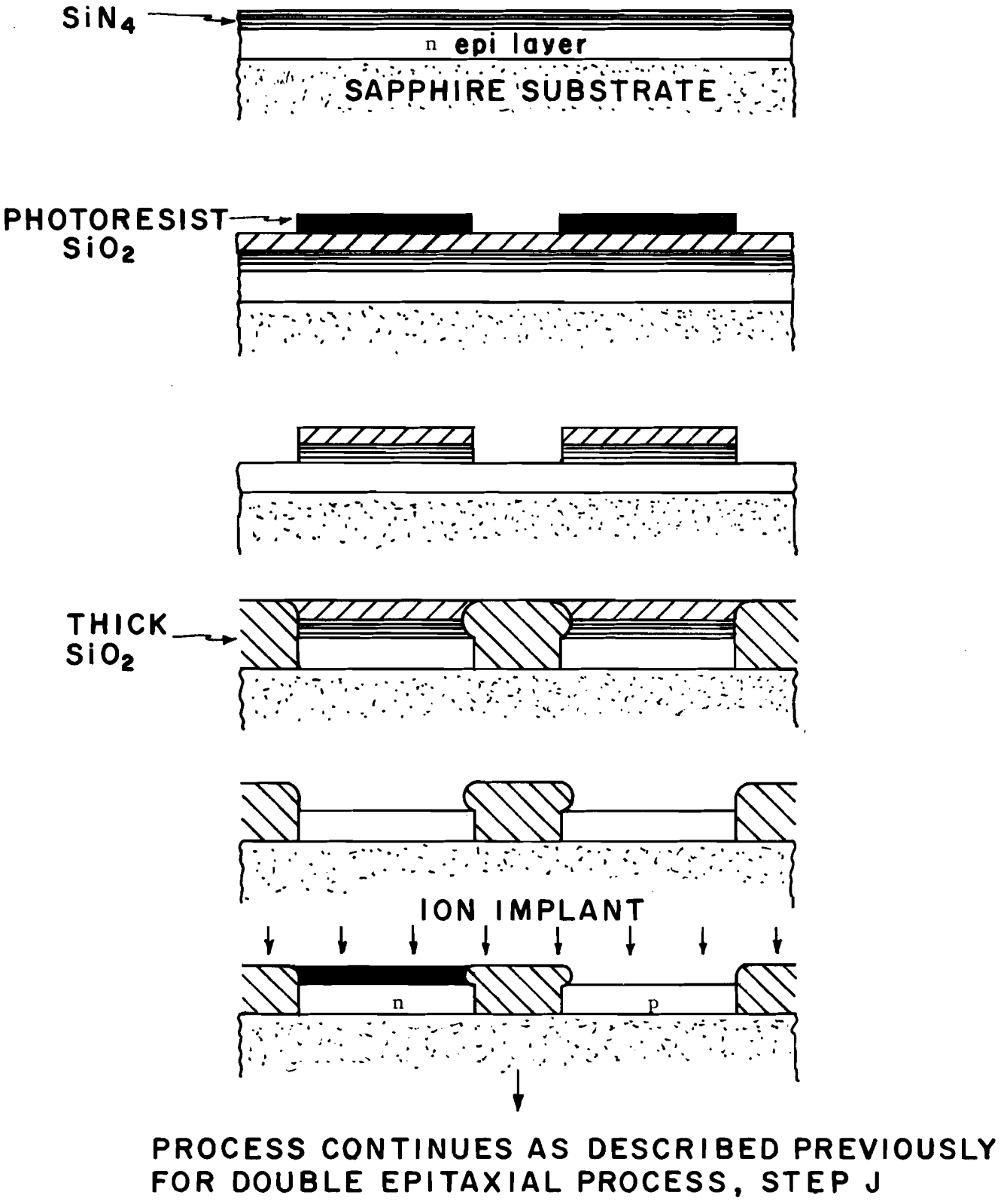


Figure III-8. Planar metal gate process.

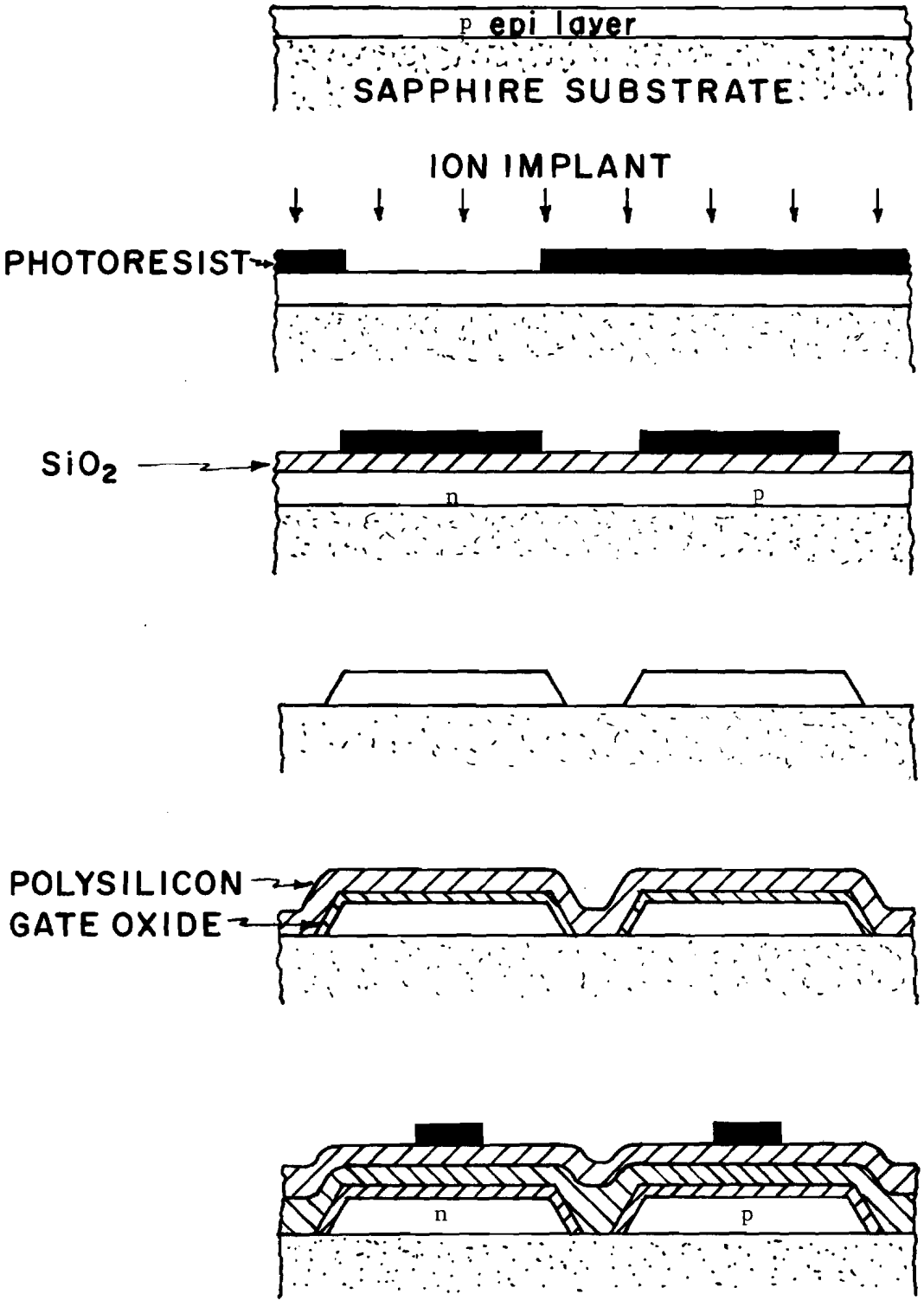
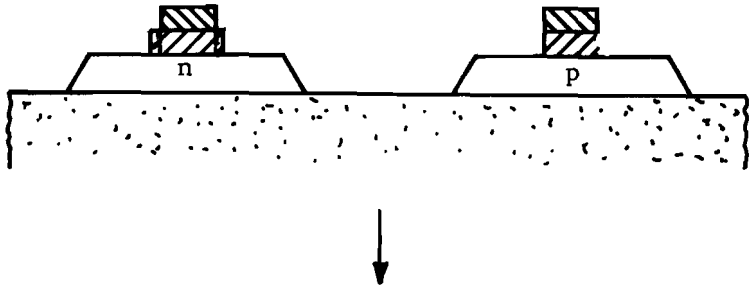


Figure III-9. Circuit processing by the silicon gate, ion implantation method.



PROCESS CONTINUES AS DESCRIBED PREVIOUSLY FOR DOUBLE EPITAXIAL PROCESS, STEP M

Figure III-9. Continued.

c) Ion implantation of well areas

d) Thermal oxidation

e) Photolithographic step to define island areas

f) Oxide and silicon etch

g) Oxide removal

h) Gate oxidation (described in detail later in this section)

i) Deposition of polycrystalline silicon

j) Oxidation

k) Photolithographic operation to define gate regions

l) Oxide and silicon etch

m) Gate oxide etch to open source and drain regions

At this point the processing steps would be the same as for the double epitaxial process, starting with step m), except that gate oxidation has already been accomplished. Seven photomasks are required.

It must be pointed out that the silicon gate process described here is a basic one, and there are as many variations as there are companies. For example, the polycrystalline silicon could be intrinsic or doped n^+ or p^+ at the deposition step, or it could be doped later with either a gas, solid, or doped glass source. If doped oxides are used for diffusion sources, they may be left on or removed, etc. Source and drain doping can be performed prior to ion implantation. The variations are endless.

6. Gate Oxidation

Gate oxides are thermally grown to a thickness of about 1000 Å. Thinner oxides would produce better transistor characteristics, but would increase the number of gate to channel shorts due to oxide pinholes.

The current literature indicates that the gate oxide can be doped by introducing HCl gas during the oxidation process (37,38,39). Source to drain leakage currents are found to depend on the gate oxidation conditions. Leakage can be controlled by proper choice of a gate

oxidation and anneal cycle with HCl. For example one manufacturer uses a low oxidation temperature to minimize leakage in p-type transistors, and a high oxidation temperature to reduce leakage in n-type transistors.

The gate oxidation operation is probably the most critical step in silicon on sapphire processing. It involves the highest temperature used in the process, and is of primary importance in adjusting transistor electrical characteristics.

7. Thick Oxide

Prior to gate oxidation, an optional thick crossover oxide may be used. This involves the formation of a thick oxide on the silicon underpasses across which aluminum metalization runs will lie. An additional photomasking step is required to remove the thick oxide from transistor gate areas so that the proper gate insulator thickness can be formed. With thick oxide, the parasitic capacitance between interconnections can be reduced.

8. Silicon Etch

A great deal of work has been done on techniques of taper etching silicon to promote step coverage of deposited oxides and metalization. Experiments have included preferential etches such as KOH:isopropyl alcohol, controlled roll-back of photoresist, and step etching. Although a great deal of progress has been made, metalization open circuits at the silicon step is still a problem of concern.

N-type silicon layers ($10^{15}/cm^3$) used in MOS processing and p-type layers ($10^{16}/cm^3$) both etch at about the same rate, while polycrystalline silicon requires approximately 50% longer to etch in the commonly used KOH:IPA solution. The difference in etch rate reverses with HF/HNO_3 etches. This different etch rate coupled with silicon thickness variations can cause problems in dimensional control when single crystal and polycrystalline silicon are both present, such as occurs during double epitaxial processing.

The most severe silicon etching problems occur when

conductivity types ranging from p^+ through p^-, n^-, and n^+ are all being etched at the same time. KOH:IPA solutions etch n^+ material rapidly and p^+ areas slowly. HF/HNO_3 can be used when buffered properly, but undercutting has an adverse effect on dimensional control.

The recent introduction of plasma silicon etching has solved many of the problems associated with variations in etch rate and silicon thickness. Plasma etching with silicon dioxide as the etch mask allows good dimensional control while defining heavily doped ($10^{19}/cm^3$) silicon regions of opposite conductivity type.

9. Process Reproducibility

One of the more important steps in wafer fabrication is the epitaxial deposition. This step affects both the electrical characteristics of finished devices and dimensional control problems. Regarding the latter, absolute control of film thickness is not as important as variation in film thickness. In fact, rejected wafers can be salvaged at intermediate stages. When reprocessed, these wafers undergo a greater number of oxidations, hence the silicon film thickness is less than normal.

Because of advances in surface preparation and epitaxy, it is possible now to complete this stage with a high degree of reproducibility. Stringent quality control procedures are all that is required to maintain control and uniformity in the deposition process.

In later fabrication steps silicon etching and gate oxidation operations are the most critical and therefore require more careful control than other processing steps. In all fabrication techniques except the planar process it is necessary to deposit oxides, polycrystalline silicon, and metalization over silicon steps. As explained earlier, if the shoulder is too steep, good coverage will be difficult to maintain.

Device leakage currents continue to be a problem in SOS circuits. The gate oxidation cycle is critical, and exceptionally clean environments and handling techniques must be maintained.

10. Dicing

Because of the hardness of sapphire the dicing operation is more difficult than with silicon. The ability to separate dice depends on the sapphire thickness and die size. For small circuits a back lapping operation to thin the sapphire to 5 mils or so is sometimes performed. The scribing operation is accomplished with a standard diamond scribe, with single or multiple passes in each scribe lane. The circuits are then broken apart against a sharp edge.

Laser scribing techniques are also used. Because of the transparency of sapphire in the visible region, CO_2 lasers at 10.7 μm are used, since this wavelength seems to couple well with sapphire. Laser scribing can be done either from the back surface, or from the front surface with suction to remove debris.

11. Mounting and Packaging

SOS circuits are mounted with a non-conducting epoxy cement. The transparency of the sapphire is an advantage in examining bond integrity, particularly in high reliability applications.

SOS circuits can be mounted in all the commercially available ceramic and plastic packages. Glass encapsulated CERDIP packages can also be used with reasonable yields. This process involves heating in air to about 500°C. Bulk MOS devices have been found to experience a considerable loss in yield under these conditions whereas SOS devices do not.

12. Summary

The silicon gate, single epitaxial layer, ion implant process is generally acknowledged to be the superior one, and will probably be universally accepted in SOS work. The single epitaxial layer is a distinct advantage, indeed a necessity for companies in initial stages of SOS development who buy only epitaxial wafers. Ion implantation, apart from its better control over channel doping and threshold voltages, makes it possible to build CMOS circuits from a single epi layer, as opposed to single channel

plus deep depletion devices.

The silicon gate makes possible a self-aligning feature. With lower gate to source and gate to drain capacitances higher speeds can be obtained. Also, more stringent design rules can be used, leading to smaller devices, smaller chip area, and higher yield. The availability of a second layer of interconnections is also an important advantage.

In the silicon gate process the gate oxide is grown early in the fabrication sequence, then immediately covered by polysilicon. Hence the gate is protected from contamination during later processing steps. This could have important ramifications regarding reliability, since the gate oxidation phase is so crucial.

All the processes which were described here require seven levels of masking. Standard bulk CMOS circuits likewise require seven masks: p-well diffusion, p^+ diffusion, n^+ diffusion, channel oxide etch, contact windows, metalization, and overcoat protection. There are three diffusions required, as opposed to only one or possibly two shallow, non-critical diffusions in SOS processing. There is no question but that SOS circuits can be processed at a faster rate than bulk CMOS, and this may well prove to be the most significant advantage of silicon on sapphire technology.

SECTION IV

ELECTRICAL PROPERTIES OF SOS DEVICES AND CIRCUITS

A. Bipolar Diodes and Transistors

Initially, p-n junction diodes and npn bipolar transistors were fabricated with structures similar to those of usual bulk silicon types. However, it quickly became apparent that new approaches were required to overcome several inherent problems. One of these was the difficulty of fabricating good planar junctions in material with 10^8-10^{10} dislocations/cm^2. Enhanced diffusion along defects produced nonuniform junctions and junction shorts. To reduce this problem, epitaxial layers 10-40 μm thick were grown, and all devices were fabricated close to the surface where the best single crystal material existed.

p-n diodes are now fabricated in a lateral configuration. This eliminates many diffusion problems since the diffusions extend completely through the thin epitaxial layer. The junction is a vertical junction, extending all the way to the silicon-sapphire interface. As in any SOS device, aluminum autodoping and crystal imperfection can still be problem areas. Since the initial silicon layers are of poor crystallinity and contain aluminum from the substrate, it is not easy to maintain the n-region of a diode. Enhanced diffusion constants of the initial layers can cause diodes to have different electrical characteristics from the desired p^+-n-n^+ or p^+-p-n^+ diodes. A problem of some concern is the high reverse leakage current, one or two orders of magnitude greater than similar bulk devices. This leakage current is due to bulk leakage caused by high dislocation density and is not entirely due to the substrate interface layer, as deduced from its variation with silicon film thickness (40). A related problem is the extremely low minority carrier lifetime, 0.1-1.0 ns, which can be increased an order of magnitude by HCl gettering techniques (38). As with all lateral devices, the series resistance of the n-region becomes significant at higher current levels and limits the maximum current rating of the diode.

It is concluded that diodes in SOS circuits have rather poor electrical characteristics and must be limited to applications where the electrical limitations of the diode do not significantly limit the circuit performance.

In the investigation of bipolar transistors the bulk device structure was found to be undesirable for the same basic reasons. An attempt to utilize lateral geometry met with limited success. To eliminate the diffusion control problems, a four epitaxial layer structure was developed (11,12). Arsenic was used for the n-type dopant. By using HCl-Cl gettering, carrier lifetime was increased from the subnanosecond range to about 10 ns, with a resulting current gain increase from 5 to 65 at 5 mA. However, the low current beta dropped off rapidly. The frequency response of this structure has been improved and exhibits a measured f_t of almost 1.2 GHz and a projected f_t of 3 GHz.

More recent developments (10) demonstrate that the minority carrier lifetime can be increased significantly, which will improve the electrical characteristics of diodes and bipolar transistors. The technique of doping the first epitaxial layer with a phosphorus doped glass is used to obtain a heavily doped n^+ layer which traps aluminum and other lifetime killing impurities. The second epitaxial layer which is grown on the n^+ layer has achieved carrier lifetimes in the 1 μs range. This development has been mainly utilized in microwave varactors and pin diodes, but does have direct application to bipolar integrated circuits. It would appear that currents gains and f_t's should increase significantly, and leakage currents should decrease somewhat. However, a four epitaxial layer system (such as shown in Fig. III-5) is inherently costly and difficult to control. Thus it is concluded that SOS bipolar circuits that utilize the improved lifetimes will still be relegated to specialty applications.

B. Single Channel PMOS Or NMOS

Single channel circuits have been processed on sapphire substrates. The fabrication of either PMOS or NMOS SOS circuits is almost as complex as CMOS and there appears to be no great advantage to a single channel SOS technology. It is true that devices of one polarity can be optimized for the key leakage, threshold, and gain parameters without regard for the devices of opposite polarity. In particular, there are several techniques for reducing the leakage current of n-channel or p-channel devices alone. However, the optimized leakage would still be higher than comparable bulk MOS devices.

The speed of single channel SOS devices should be greater than that of bulk MOS devices of the same channel type since parasitic capacitances are less. However, larger devices would be needed to achieve the same current drive capability. If the increase in speed is essential it would be better to utilize an SOS CMOS circuit.

Since most single channel MOS circuits are ratio-type, the device geometries are more critical than in CMOS circuits, which are ratioless. The problem from this standpoint with SOS circuits is control of the channel width, which is determined solely by the width of the etched silicon islands. In bulk MOS the channel width is determined by an oxide etch and can be controlled to much better tolerances.

C. CMOS Devices

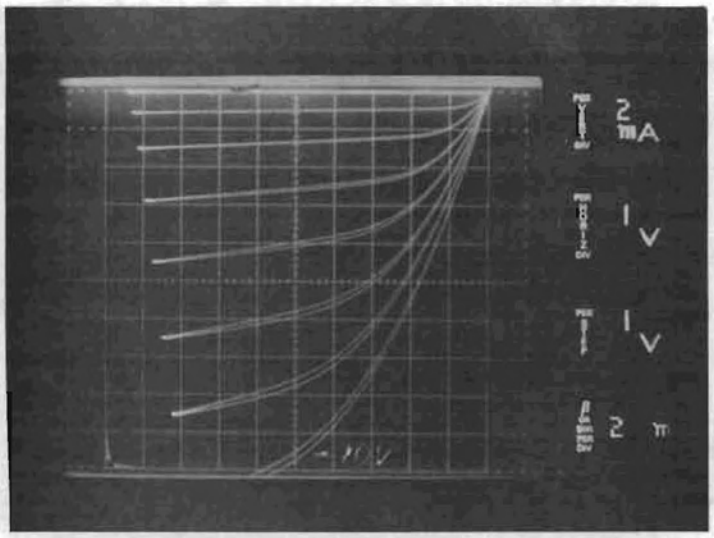
1. General Characteristics

In general, the characteristics of enhancement mode MOS devices processed on sapphire are similar to the characteristics of conventional bulk MOS devices. The same device equations apply in both technologies. SOS threshold voltage, device gain, and breakdown voltage are similar to those of comparable bulk MOS devices for both aluminum gate and silicon gate processes (41,42). The same transistor models are also used for both types of devices (43), although the transient performance is somewhat different due to the difference in some capacitance parameters. The major differences between SOS and bulk MOS devices are source to drain leakage and the phenomena which result from the "floating substrate" in SOS devices.* These effects will be described shortly.

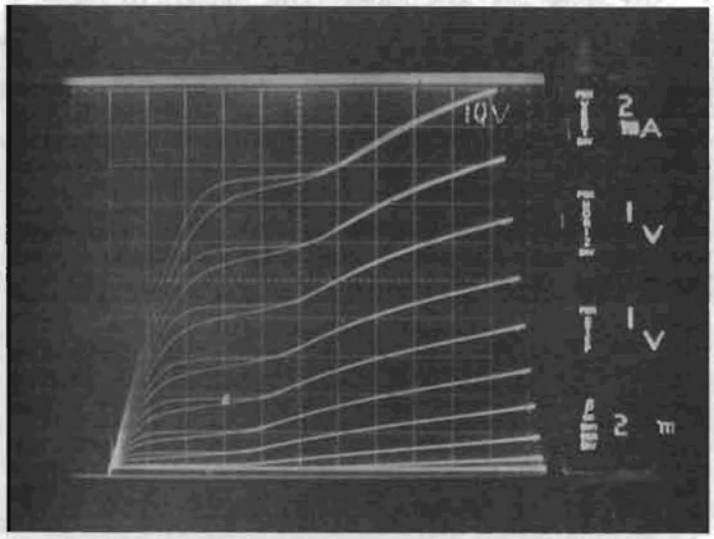
A drain characteristic curve for a typical p-channel aluminum gate SOS transistor is shown in Fig. IV-1. This transistor has a channel width of 20 mils, an effective channel length after diffusion of approximately 0.2 mils, and a 1 μA threshold voltage of 1.8 volts. Characteristics for an n-channel transistor with the same dimensions are also shown in the figure. In this case the 1 μA threshold voltage is 2.0 volts. The gate oxide thickness for both devices is 1100 Å.

---

*This unfortunate choice of terminology has been adopted from useage in bulk MOS technology. The term "floating substrate" does not refer to the sapphire substrate, which is an insulator. It means that the silicon in the channel region below the induced inversion (or depletion) layer is not electrically connected.



(a) p-channel



(b) n-channel

Figure IV-1. Drain characteristic curves for aluminum gate SOS transistors with 20 mil channels.

Assuming that the drain characteristics can be approximated reasonably well in the saturated region by the commonly used square law relationship, two important electrical parameters can be found. The pertinent equation is

$$I_D = \frac{\beta}{2} \frac{W}{L} (V_{GS} - V_T)^2 \qquad \text{(IV-1)}$$

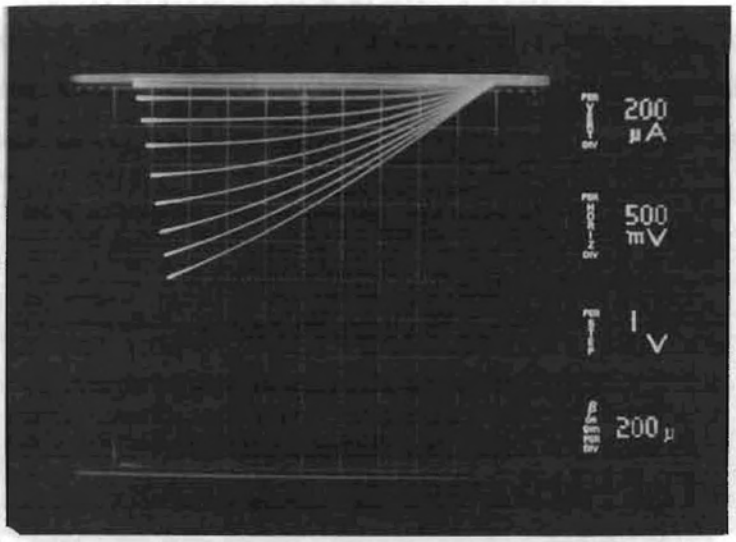
where

$$\beta = \frac{\mu \epsilon_{ox}}{t_{ox}} \qquad \text{(IV-2)}$$

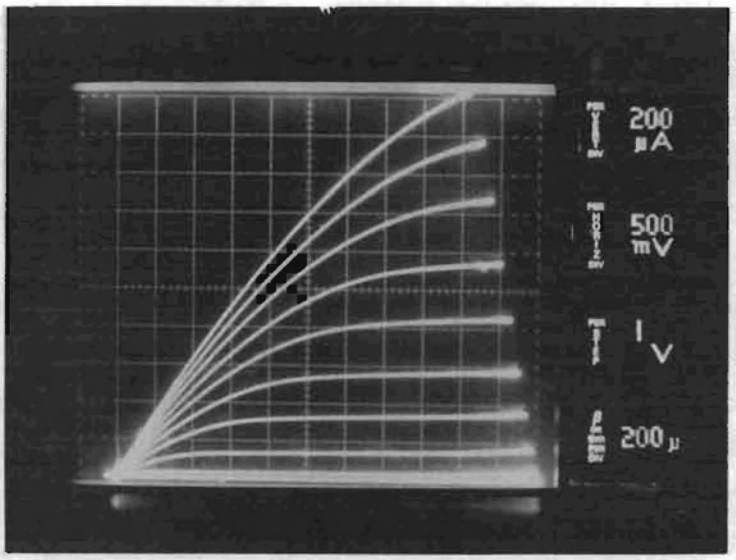
and W is the channel width, L is the effective length, V_{GS} is the gate voltage with respect to the source, V_T is the threshold voltage, μ is the carrier mobility, ϵ_{ox} is the permittivity of the gate oxide (4 ϵ_o, where $\epsilon_o = 8.85 \times 10^{-14}$ F/cm) and t_{ox} is the thickness of the gate oxide. The p-channel device β calculated from Eq. IV-1 at the operating point $V_{GS} = V_{DS} = 5$ volts in Fig. IV-1 is 5.3 $\mu A/V^2$. The p-channel mobility calculated from Eq. IV-2 is then 160 $cm^2/V\text{-}s$. Similar calculations for the n-channel device yield $\beta_N = 6.2\ \mu A/V^2$ and $\mu_N = 190\ cm^2/V\text{-}s$. The low value of n-channel mobility is typical of many aluminum gate SOS devices.

Drain characteristics of silicon gate SOS transistors are shown in Fig. IV-2. The change in characteristics from the aluminum gate device is mostly due to the smaller geometry. The silicon gate transistors both have a channel width of 2 mils, a channel length of 0.25 mils, a threshold voltage of 1.5 volts, and a gate oxide thickness of 1300 Å. From the characteristics at $V_{GS} = V_{DS} = 5$ volts, it is found that $\beta_p = 5.1\ \mu A/V^2$, $\mu_p = 190\ cm^2/V\text{-}s$, $\beta_N = 11\ \mu A/V^2$, and $\mu_N = 400\ cm^2/V\text{-}s$.

Ever since the deep-depletion MOS transistor was described in 1966 (44), it has been investigated as a potentially viable addition to the SOS family. The device can be either p-channel or n-channel as shown previously in Fig. III-5. In CMOS/SOS deep-depletion technology, one transistor in the complementary pair is a normal SOS enhancement mode device and the other is a deep-depletion device. The primary advantage



(a) p-channel



(b) n-channel

Figure IV-2. Drain characteristic curves for silicon gate SOS transistors with 2 mil channels.

is that only one transistor bulk deposit polarity is required. Thus, only a single epitaxial layer need be grown, and no counter doping by diffusion or other means is required.

For digital circuit operation the deep-depletion transistor must operate like an enhancement mode transistor. That is, it must turn off with zero gate to source voltage. This off condition is achieved by depleting the entire channel region down to the sapphire substrate. To achieve this there are two significant parameter boundaries in designing the deep-depletion device (14). First, the zero gate voltage depletion width must extend completely through the epitaxial layer. Second, this depletion width must be less than the maximum possible value for the given channel doping. Since the same epitaxial layer is used for both the deep-depletion device and the enhancement mode device, the epitaxial film doping concentration is no longer an independent parameter but must be chosen to simultaneously meet the deep-depletion device requirements. Unfortunately, these requirements are rather stringent.

Deep-depletion devices have low threshold voltages and are not suited to general purpose applications over wide voltage ranges. The device is now being used in low voltage specialized circuits such as timing chips for watches. Strict processing parameter control is required, particularly with regard to epitaxial layer thickness and resistivity.

2. Threshold Voltage

The threshold voltage equation for enhancement mode SOS transistors is identical to the bulk MOS threshold equation and is given by (45)

$$V_T = \phi_{MS} - q\frac{N_{SS}}{C_{ox}} - \frac{Q_B}{C_{ox}} + 2\,\phi_F \qquad \text{(IV-3)}$$

where ϕ_{MS} is the work function difference between gate material and semi-conductor channel material, q is the electronic charge, equal to $1.6\text{x}10^{-19}$ Coulombs, N_{SS} is the surface state density in states/cm^2, C_{ox} is the gate capacitance/area, Q_B is the bulk charge density in the depletion layer of the silicon channel, and ϕ_F is the Fermi level of the channel material. The bulk charge density is given by

$$Q_B = (2\ \epsilon_{Si}\ \epsilon_o\ q\ N_B\ |2\ \phi_F|)^{\frac{1}{2}} \qquad \text{(IV-4)}$$

where ϵ_{Si} is the relative permittivity of silicon, 12, and N_B is the impurity density in the channel material. The Fermi level can be calculated from

$$\phi_F = \frac{kT}{q}\ \ell n(\frac{N_B}{n_i}) \qquad \text{(IV-5)}$$

where kT/q is equal to 0.026 eV at room temperature and n_i is the intrinsic carrier concentration in silicon, $1.5 \times 10^{10}/cm^3$.

The quantities in Eq. IV-3 have the following algebraic signs for enhancement mode devices.

	p-channel	n-channel
V_T	-	+
N_{SS}	+	+
Q_B	+	-
ϕ_F	-	+

Typical values of ϕ_{MS} for background concentrations in the range 5×10^{14} to 2×10^{16} have been computed for aluminum gate, p-doped silicon gate (assuming 100 Ω/□ gate sheet resistance) and n-doped silicon gate (assuming 10 Ω/□ gate resistance) (45,46). For the n-doped silicon gate the assumed sheet resistance approaches the value for solid solubility limit and the work function is the same as that for aluminum. Values of work functions under these conditions are listed below.

Gate Conductor	Substrate Doping	Difference in Work Function
Aluminum or n-type silicon	n-type	-0.3 volts
p-type silicon	n-type	+0.6 volts
Aluminum or n-type silicon	p-type	-0.9 volts
p-type silicon	p-type	0 volts

Because N_{SS} for <111> oriented silicon is high, n-channel devices tend to be depletion type and thus cannot be turned off in digital applications. Consequently, <111> is not used for n-channel or CMOS circuits. <110> silicon has a surface orientation dependent N_{SS} and the thresholds would vary within a circuit layout, depending on the transistor orientation. Therefore <110> silicon is not used for MOS devices. <100> is the orientation chosen for n-channel and CMOS circuits. Well processed <100> silicon has an N_{SS} value of approximately $5 \times 10^{10}/cm^2$ which is a good value for matching n- and p-channel threshold voltages.

The threshold voltage is plotted in Figures IV-3 - IV-6 as a function of background concentration and gate oxide thickness for both aluminum gate and p-doped silicon gate p-channel and n-channel devices. The work function difference between the two types of gate is clearly evident.

Threshold matching of p and n-channel transistors can be readily achieved for different fabrication processes by a proper choice of the background concentration for each transistor. The following combinations have been selected from the curves as examples of threshold matching.

Gate	Transistor Type	Threshold Voltage	Background Concentration
p-doped silicon	p-channel	1.5 volts	$4.5 \times 10^{15}/cm^3$
p-doped silicon	n-channel	1.5 volts	$2.2 \times 10^{15}/cm^3$
Aluminum	p-channel	2 volts	$1.5 \times 10^{15}/cm^3$
Aluminum	n-channel	2 volts	$1.5 \times 10^{16}/cm^3$
p-doped silicon	p-channel	1.5 volts	$4.5 \times 10^{15}/cm^3$
n-doped silicon	n-channel	1.5 volts	$1 \times 10^{16}/cm^3$

3. Drain Current and Gain

The drain current in the triode region is given by

$$I_D = \frac{\mu \epsilon_{ox}}{t_{ox}} \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \qquad \text{(IV-6)}$$

and in the saturation region (from Equations IV-1 and IV-2)

$$I_D = \frac{\mu \epsilon_{ox}}{2 t_{ox}} \frac{W}{L} (V_{GS} - V_T)^2 \qquad \text{(IV-7)}$$

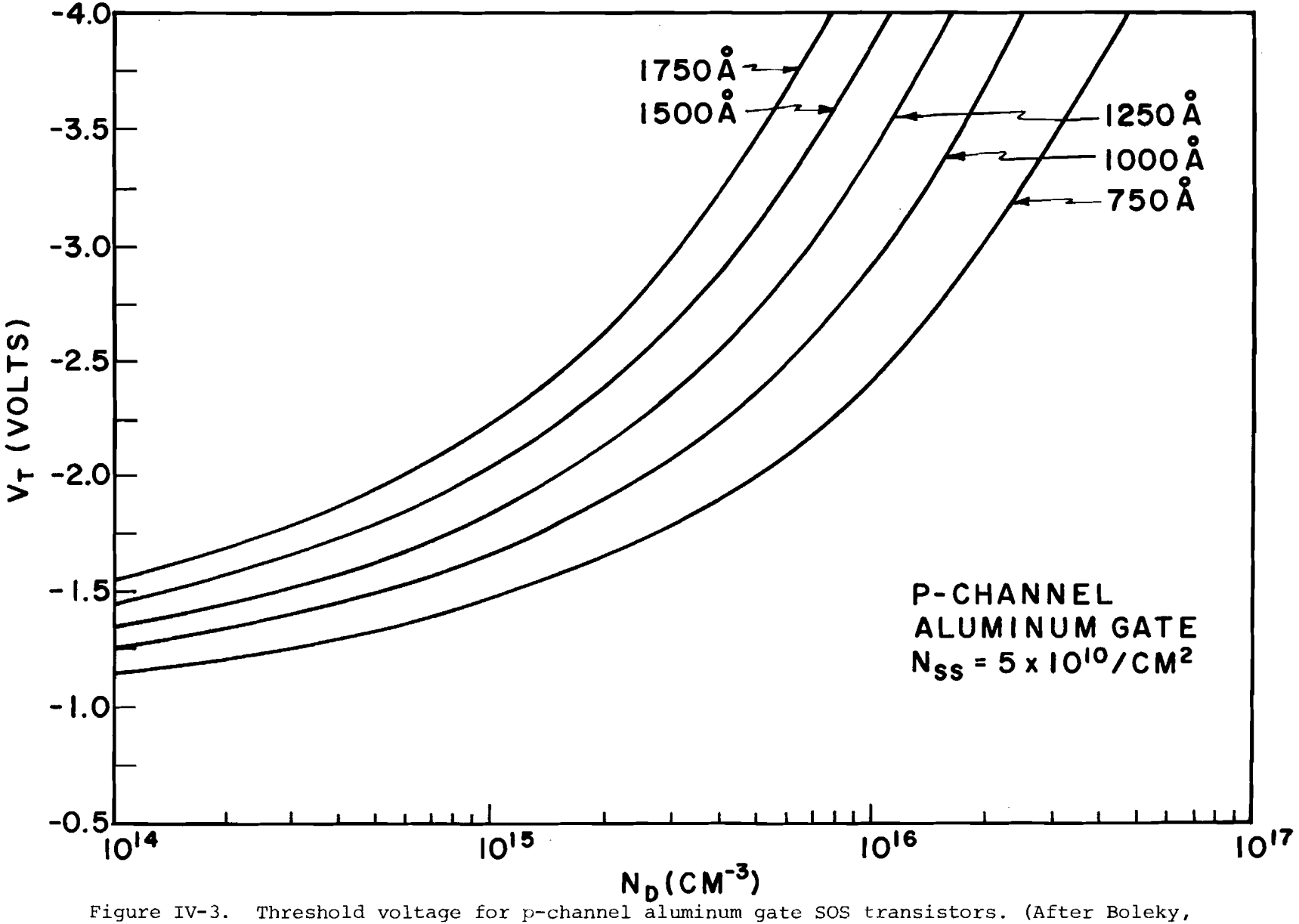


Figure IV-3. Threshold voltage for p-channel aluminum gate SOS transistors. (After Boleky, reference 14.)

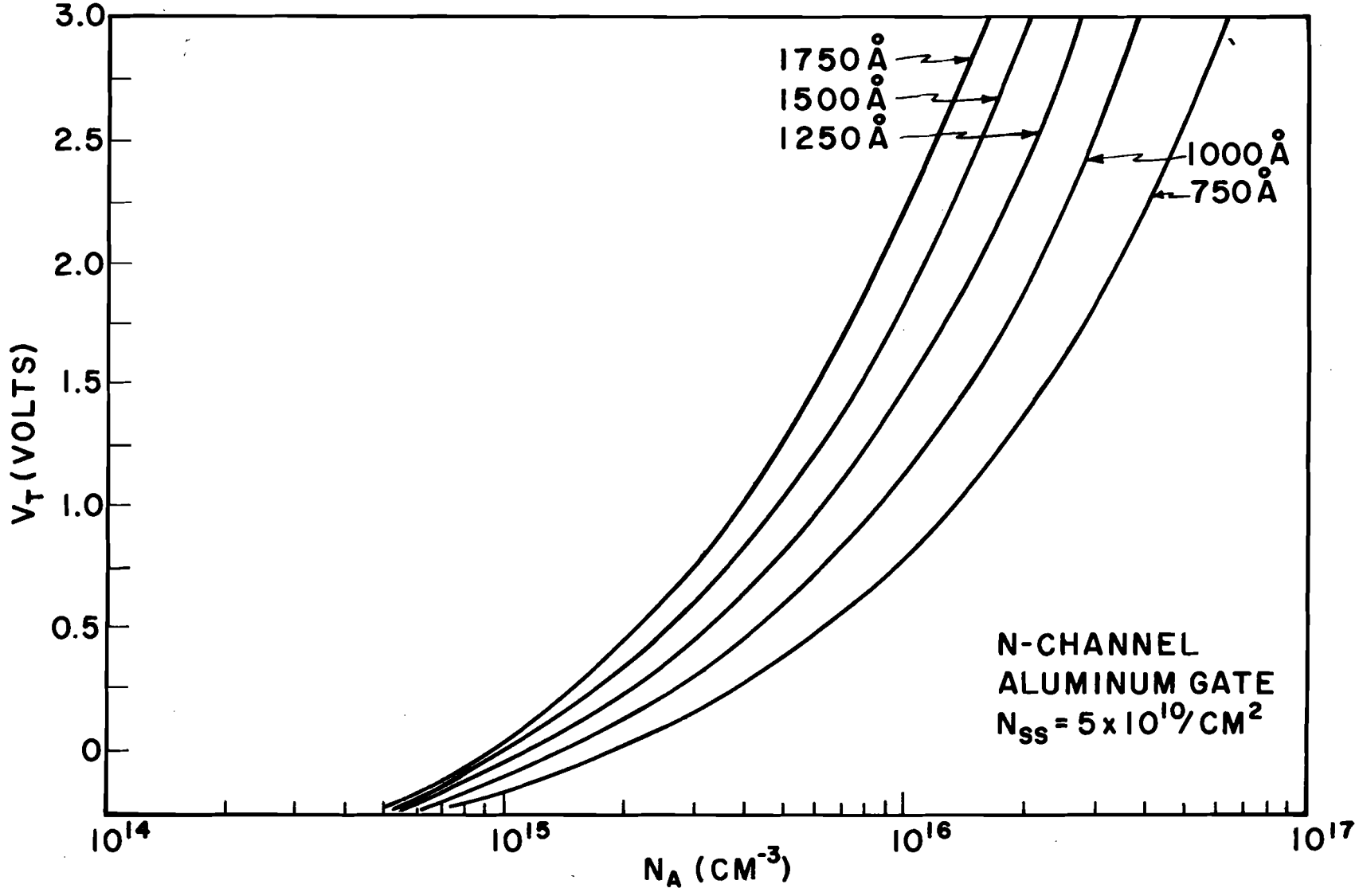


Figure IV-4. Threshold voltage for n-channel aluminum gate SOS transistors. (After Boleky, reference 14.)

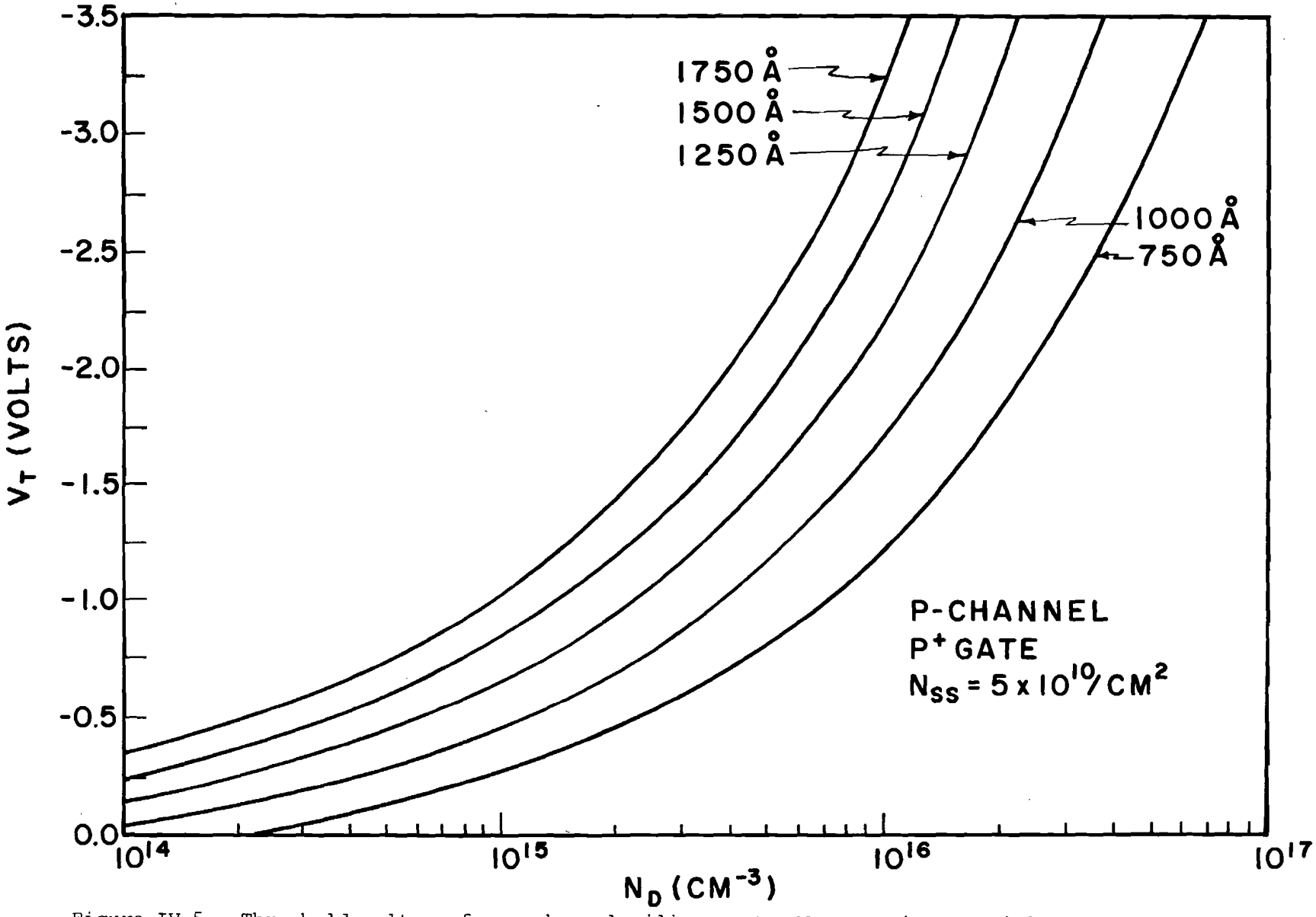


Figure IV-5. Threshold voltage for p-channel silicon gate SOS transistors. (After Boleky, reference 14.)

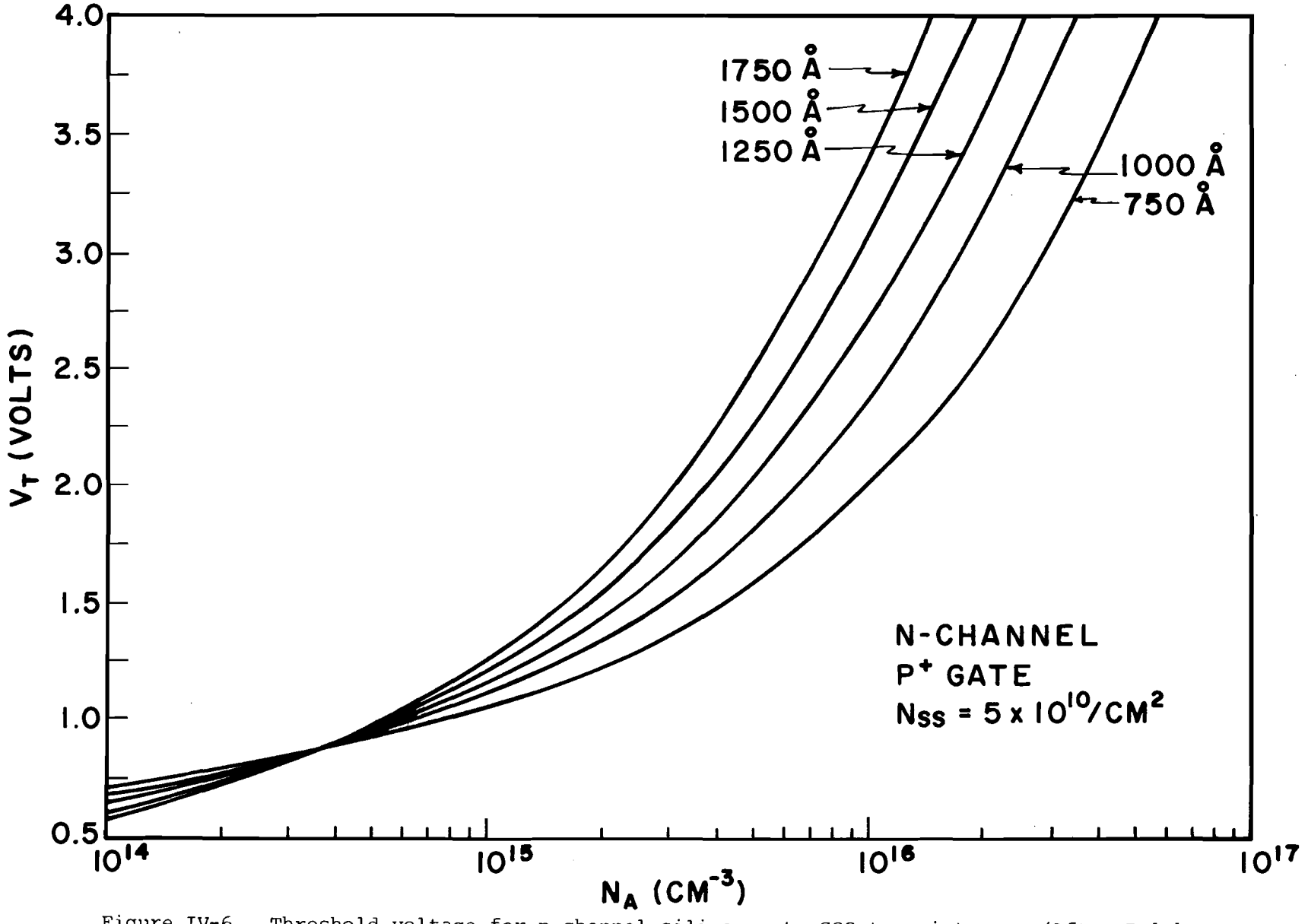


Figure IV-6. Threshold voltage for n-channel silicon gate SOS transistors. (After Boleky, reference 14.)

The small signal gain in the triode region is

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \frac{\mu \epsilon_{ox}}{t_{ox}} \frac{W}{L} V_{DS} \qquad \text{(IV-8)}$$

and in the saturation region

$$g_m = \frac{\mu \epsilon_{ox}}{t_{ox}} \frac{W}{L} (V_{GS} - V_T) \qquad \text{(IV-9)}$$

These equations are based on the gradual channel approximation and do not take into account the variation of effective channel length with drain voltage nor the saturated velocity of carriers, among other second order effects. Consequently, the equations are not necessarily accurate representations for all geometries. Nevertheless, they are widely used because of their simplicity.

It is apparent that drain current and gain are functions of μ/t_{ox}. In general, the goal is to maximize I_D and g_m with a given W/L ratio and operating voltage. Gate oxides are usually about 1000 Å. Thinner oxides are more susceptible to pinhole failures or oxide rupture and thus lead to potential reliability problems. As explained previously, SOS mobilities are somewhat less than bulk values for reasons not clearly understood. The current drive capacity of SOS devices is thereby reduced, necessitating larger output transistors. This is an area of continued concern to SOS manufacturers.

4. Source to Drain Leakage

Leakage current in SOS is a problem because it is an order of magnitude higher than in comparable bulk SOS circuits. The increased quiescent power dissipation in SOS CMOS is of little consequence since it is completely overshadowed by the dynamic power dissipation which is several orders of magnitude higher. In certain applications where the circuit operates in the standby mode for long periods of time the battery drain may be an important consideration. But the main reason for concern is that the increased leakage currents indicate that the

semiconductor industry has an incomplete understanding of the reasons for its existence and, as yet, an inability to design a process to control the problem. Any lack of knowledge or control indicates a potential reliability problem.

It has been shown (47) that the significant leakage currents in CMOS devices are due to the reverse junction current of the drain to channel diode of the transistor. Typical reverse current versus voltage measurements of SOS junctions indicate that both the p^+n and n^+p vertical junctions perform as step junctions in accordance with the Sah-Noyce-Shockley theory (48) of generation current in silicon diodes. The reverse current is proportional to $V^{\frac{1}{2}}$ over several orders of magnitude of the voltage. This is due to the dependence of the depletion region width, x_J, and hence depletion region volume on voltage.

The generation current per unit area of junction is

$$I = \frac{q n_i x_J}{2\tau} \qquad \text{(IV-10)}$$

where τ is the minority carrier lifetime and

$$x_J = \left[\frac{2\,\epsilon_{Si}\,\epsilon_o (2\,\phi_F + V_R)}{q\,N_B}\right]^{\frac{1}{2}} \qquad \text{(IV-11)}$$

where V_R is the applied reverse voltage.

With typical background concentrations of $10^{15}/cm^3$ for p-channel transistors and $10^{16}/cm^3$ for n-channel devices, x_J is 3.7×10^{-4} cm and 1.18×10^{-4} cm, respectively, for a reverse bias of 10 volts. With a minority carrier lifetime of 0.5 ns and 1 μm silicon thickness, a leakage current of 225 pA/mil of width for a p-channel device and 72 pA/mil for an n-channel device is predicted by the above equations. (This compares with measured values of 150 and 80 pA/mil).

5. Kink Effect

The n-channel drain characteristic in Fig. IV-1 indicates an anomalous increase in drain current at a drain voltage of about 5 volts. This current increase is the so-called "kink effect." An explanation for this effect which correlates well with measurements is as follows (49). With a sufficiently high drain to source voltage (usually less

than 10 volts) and with the gate voltage greater than the threshold voltage so that the surface is inverted, a channel current exists. An additional component of current flows from the drain into the bulk (that is, into the channel region which has not been inverted) due to carrier avalanche multiplication in the depletion region at the drain end of the channel. The electrons generated in the depletion layer flow toward the drain whereas holes flow into the bulk. These holes tend to forward bias the bulk with respect to the source. The change in the bulk voltage has the effect of lowering the threshold voltage, thus increasing the drain current. This increase in current is the kink effect.

A second kink effect can also occur due to lateral bipolar transistor action (50). In this mode the source of the MOS transistors acts as an emitter, the channel region the base, and the drain the collector terminal. Electrons from the source are injected into the channel region to initiate the transistor action. At relatively high drain currents the effective base width is small due to the Early effect, and β of the parasitic bipolar transistor can be high. The collector current adds to the normal drain current to produce the second kink in the characteristics. Normally this transistor action is only significant at higher drain voltages (greater than 10 volts) and should not be a major factor in SOS design.

Although the kink effect is mentioned in conjunction with n-channel SOS transistors, it also occurs in p-channel devices but is much subdued because of the lower ionization coefficient for holes. Kink effects can be completely eliminated by grounding the substrate (that is, the channel region). In fact, it will be demonstrated in the next chapter that the kink effect can be seen even with bulk MOS transistors if the substrate is left floating.

The distortion in the curves produced by the kink effect does not seriously affect circuit operation. Figure IV-7 illustrates the path of operation during switching of a device which exhibits a kink effect. At point 1 the device is turned off and there is no kink effect since no current is flowing. Assuming that the load is capacitive, turning

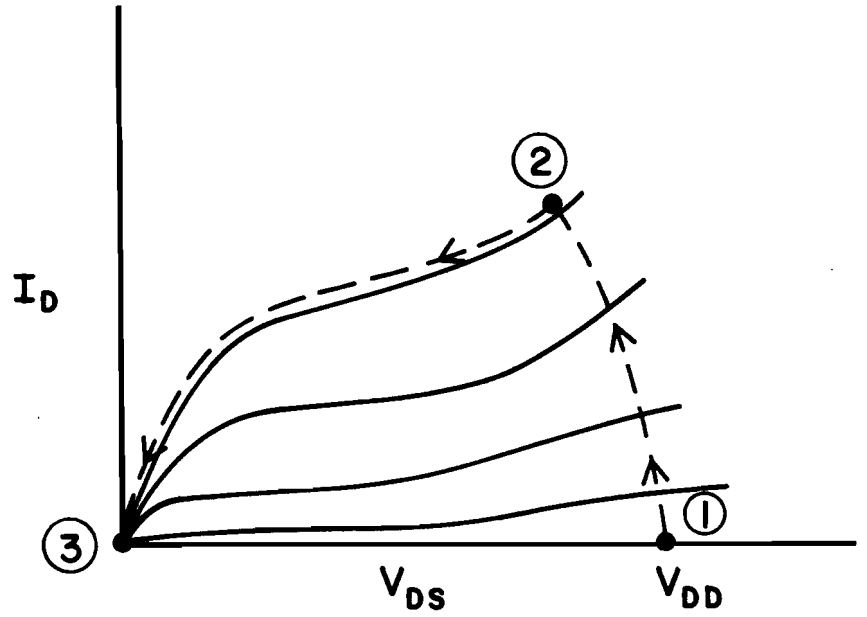


Figure IV-7. Switching path in a device which exhibits a kink effect.

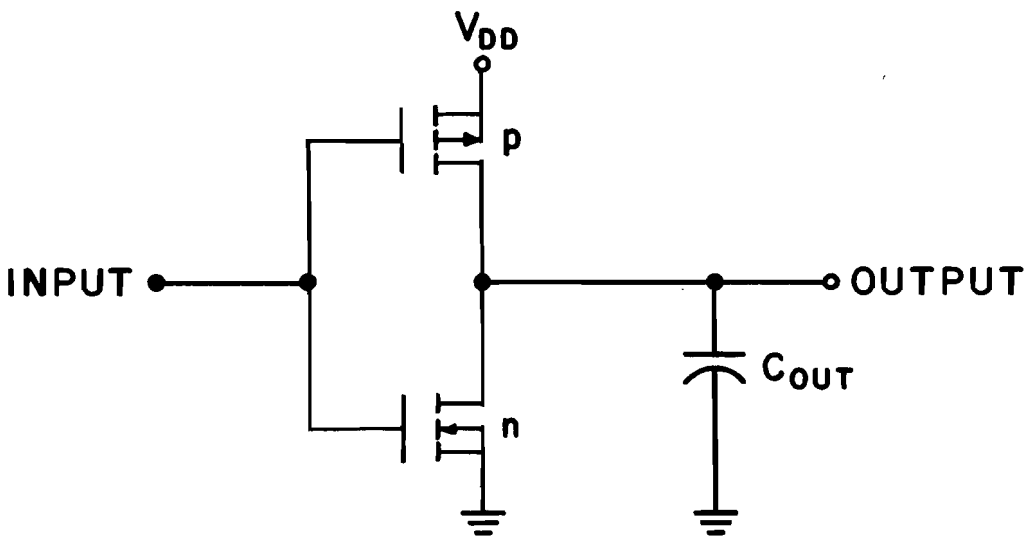


Figure IV-8. Basic CMOS inverter.

on the transistor will cause the operating point to shift momentarily to point 2, then to point 3. The enhanced current due to the kink effect at point 2 will actually cause faster switching. At point 3 the device is turned on but is in the deep triode region with very low drain to source voltage so that the depletion layer between drain and channel regions is very much reduced.

6. Edge Effects

The leakage current in an SOS transistor fabricated in a closed geometry, that is, in a pattern of concentric circles, is considerably less than that of an equivalent open geometry device, where the edges of the gate are exposed. Similarly, in comparing the electrical characteristics of an SOS transistor in which the gate overlaps the silicon island edge with one in which the gate is confined to only the top face of the island, it is found that an additional current flows through a surface channel at the edge of the island. At first the effects were thought to be due to the difference in crystallographic orientation at the top surface and edge of the island, with a corresponding change in N_{SS} and threshold voltage. This is not adequate to explain the effect in both p and n-channel devices however.

More recently (51) the edge effect has been explained in terms of two charge transfer mechanisms which produce interface charge at the island edge and a shift in threshold voltage. The first mechanism is charge transfer by tunneling. Its effects can be seen on both n-and p-channel devices from a variety of manufacturers. The second mechanism involves charge transfer through the gate dielectric itself. It is observed only at elevated temperatures. Furthermore, it is strongly affected by the type of gate which exists, being reduced to a negligible amount if a p^+ silicon gate is used, or if an appropriate dielectric exists between the gate and gate oxide.

In the continuing trend to smaller and smaller geometries, the current carried by the channels at the island edges becomes a larger and larger fraction of the total device current. Unfortunately, the only methods for eliminating the edge effect are to build closed geometries or to preferentially increase the doping density at the island edges.

D. CMOS Logic Circuits

All CMOS logic circuits are extensions of the basic CMOS inverter in Fig. IV-8 (45,52). The circuit is normally operated so that both transistors are not turned on simultaneously, except during input transients. In the static condition there is no connection between V_{DD} and ground since one of the transistors is always blocking current. The static power dissipation, due only to leakage current, is typically less than 10 nanowatts per inverter.

Although MOS circuits are notoriously slow speed, CMOS begins to approach the speed of slower bipolar circuits (10 to 20 MHz) due to the fact that the transistors provide a relatively low impedance drive to the load capacitance for both rising and falling output transitions.

CMOS circuitry achieves good noise immunity (typically 40% of the power supply voltage) because of the high circuit gain in the transition region and because of symmetry in the characteristics of the n and p-channel transistors. The high gain results from the use of an active load and the symmetry is obtained by adjusting the channel widths to compensate for the difference in mobilities.

Typical dc transfer characteristics for a CMOS inverter are shown in Fig. IV-9 for three different power supply voltages. It is clear that the transfer characteristic is symmetrical over a wide range of supply voltages and that the transition region is centered at $V_{DD}/2$. The gain is also high in the transition region.

The cross-hatched bands are drawn between the 40% and 60% V_{DD} voltages. The points of intersection with the transfer characteristic are circled and define the output voltage when noise of 40% of the power supply voltage is applied to the input as either positive transients with respect to 0 volts or negative transients referenced to V_{DD}. In all cases the output perturbation is less than 2 volts and is not sufficiently large to switch a similar CMOS circuit connected to the output. Therefore, the noise immunity in this example is at least 40% of the power supply voltage. In practice, variations in processing parameters and more complex logic circuits with several inputs result in non-ideal transfer curves. However, in almost all cases noise immunity of at least 30% of the power supply voltage is achieved.

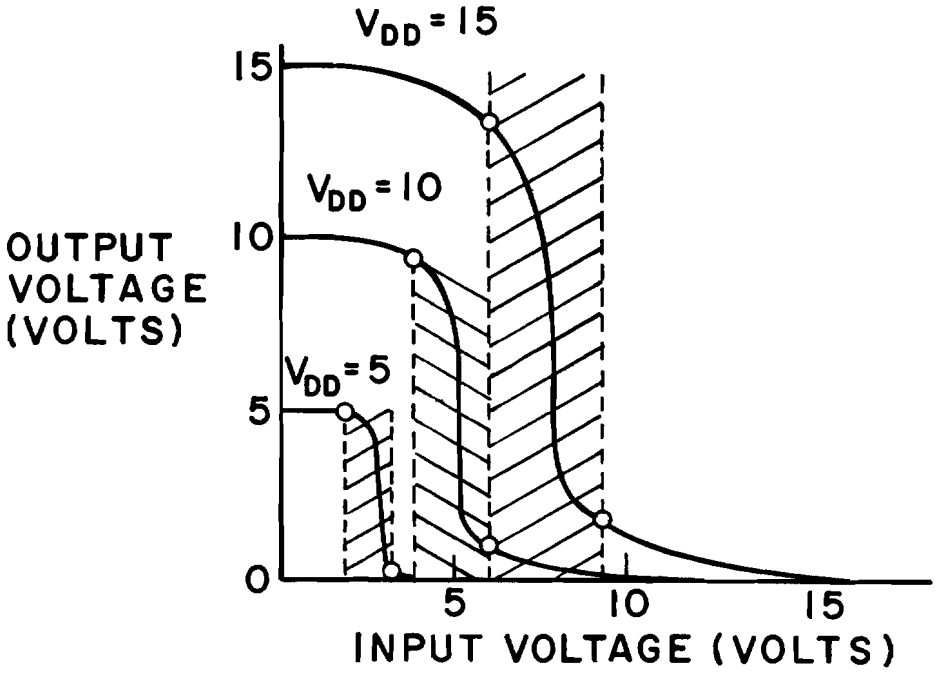


Figure IV-9. Transfer characteristics for a typical CMOS inverter.

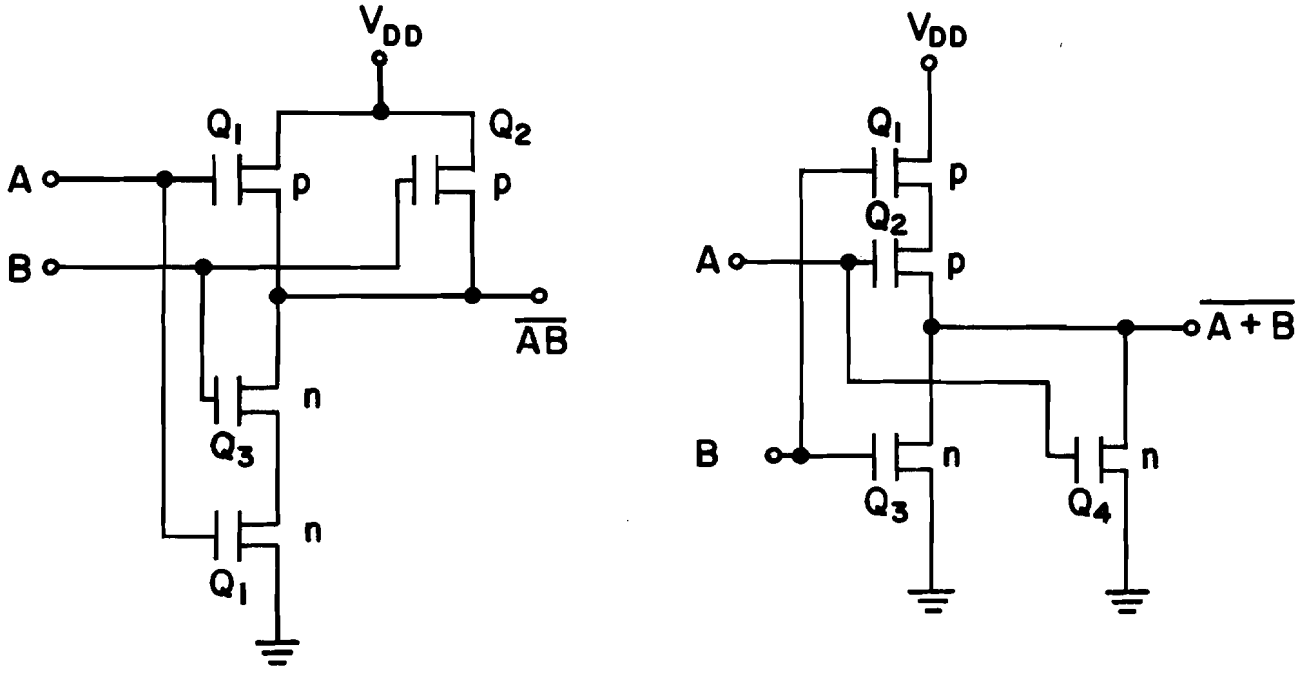


Figure IV-10. Basic CMOS NAND gate and NOR gate.

A NAND logic gate can be constructed by connecting two p-channel transistors in parallel and two n-channel transistors in series as shown in Fig. IV-10. This circuit configuration is typical of all CMOS technologies. The NOR gate configuration is the dual of the NAND configuration, with the n-channel transistors in parallel.

With bulk CMOS logic the substrates of all n-channel transistors are connected to ground. As a result, some transistors will exhibit a threshold body effect since the source and substrate are not at the same potential.* This can be compensated for by making these transistors slightly larger to restore the gain. As an alternative measure the substrate could be connected to the source, but this increases the layout area since the transistors would have to be individually isolated. In SOS CMOS the situation is entirely different since the transistors are individually isolated and the substrates are left floating anyhow.

E. Transient Performance

The switching speed of fast bipolar digital circuits is typically 5 to 50 times faster than comparable bulk MOS circuits (47). The primary reason is the lower device gain per unit area for MOS circuits compared to bipolar circuits. It can be shown (47) that the transit time for carriers is roughly the same for both MOS and bipolar devices. With lower transistor gain for MOS transistors it is imperative that circuit capacitance be reduced to a minimum to achieve high speed operation.

An equation for the propagation delay t_{pd} of a CMOS inverter that has been widely used and which agrees reasonably well with measured values is (52)

$$t_{pd} = \frac{1.8 t_{ox} L_n C_{out}}{\varepsilon_{ox} \mu_n W_n V_{DD}} \left[\frac{1}{\left(1 - \frac{V_{Tn}}{V_{DD}}\right)^2} + \frac{1}{\left(1 - \frac{V_{Tp}}{V_{DD}}\right)^2 B} \right] \qquad \text{(IV-10)}$$

where

$$B = \frac{\mu_p W_p L_n}{\mu_n W_n L_p} \qquad \text{(IV-11)}$$

*The substrate, in effect, acts as a second gate and causes a shift in threshold voltage.

t_{ox} and ϵ_{ox} are the gate thickness and permittivity, respectively. C_{out} is the total capacitance at the output node. L, W, μ, and V_T are the channel length, width, mobility, and threshold voltage magnitude, respectively, with the n and p subscripts designating the transistor polarity. V_{DD} is the power supply voltage.

This equation will be used to calculate the delay time of an inverter driving another inverter as shown in Fig. IV-11. The following parameters will be used.

$$\mu_p = 200 \text{ cm}^2/\text{Vs}$$

$$\mu_n = 400 \text{ cm}^2/\text{Vs}$$

$$t_{ox} = 1200 \text{ Å}$$

$$\epsilon_{ox} = 3.54\text{x}10^{-13} \text{ F/cm}$$

$$V_{Tn} = 2 \text{ V}$$

$$V_{Tp} = 2 \text{ V}$$

Minimum device geometries of 0.3 mil channel width and 0.2 mil effective channel length, and a power supply voltage of 10 volts will be used. Substituting into Eq. IV-10, the delay from point A to B is 48 ns/pF of capacitance at node B. This capacitance consists of the gate input capacitances of Q3 and Q4 and the drain-to-channel capacitances of Q1 and Q2.

The capacitance of the gate oxide is 0.19 pF/mil^2. Assuming an aluminum gate which is 0.5 mil long and 0.3 mil wide, the sum of the gate capacitances of Q3 and Q4 is 0.057 pF. The drain-to-channel capacitance exists only along the vertical junction between the drain of each transistor and the channel. The area of this junction is equal to the epitaxial layer thickness (taken as 0.04 mils) times the transistor width. The junction capacitance for typical channel doping levels and 10 volts applied is estimated at 0.1 pF/mil^2. Therefore, the total capacitance contributed by Q1 and Q2 is 0.0048 pF. This is negligible compared to the gate capacitance. The propagation delay based on this assumption is then 2.74 ns, in fairly good agreement with measured values.

If the power supply voltage is reduced to 5 volts the propagation

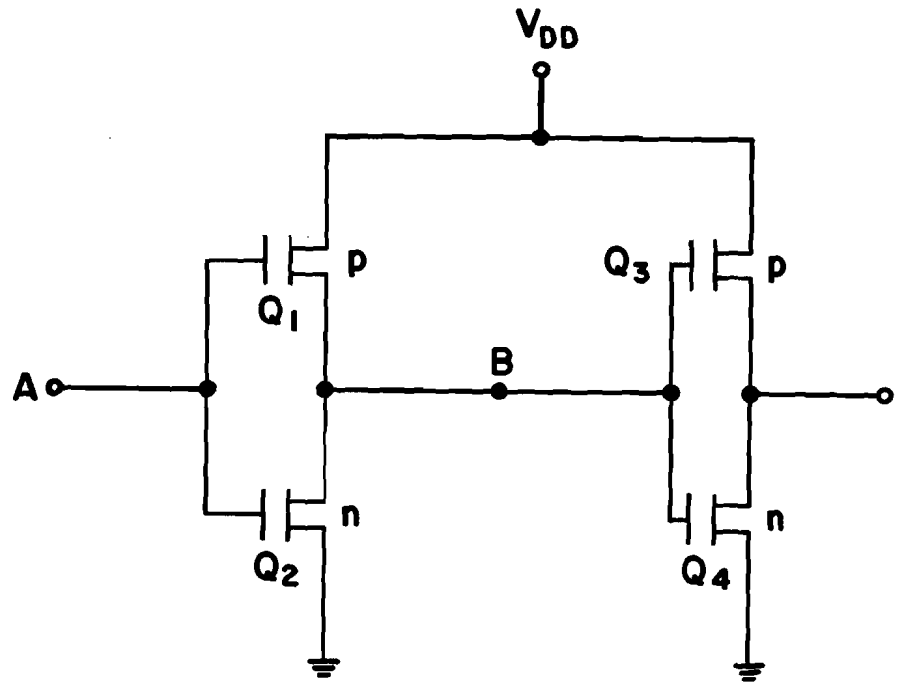


Figure IV-11. Basic configuration of one inverter driving another for purposes of calculating delay time.

delay is 85 ns/pF, or a total of 9.6 ns.

For CMOS logic circuits with 2 or more inputs the series p-channel or n-channel transistors are usually enlarged or scaled up in width to achieve the same approximate gain as a single parallel transistor. Referring again to Fig. IV-10, Q3 and Q4 of the NAND gate would each be doubled in width so that when both are turned on, the composite would be equivalent to one single transistor. Similarly, Q1 and Q2 of the NOR gate would be doubled in width. Circuits with more than 2 inputs would have the series transistors scaled up linearly with the fan-in. The input capacitance for each input of an N input gate would be

$$C_g = C_{gI}\left(\frac{N+1}{2}\right) \qquad \text{(IV-12)}$$

where C_{gI} is the input capacitance for an inverter. Thus the input capacitance for a 2 input gate would be 1.5 C_{gI}.

As a more realistic propagation delay example, consider a circuit driving 2 gates, each with 2 inputs. The total gate capacitance is then 3 C_{gI}. The propagation delay for V_{DD} = 10 volts would be 8.2 ns.

When a silicon gate structure is used the self-aligning feature decreases the gate conductor length and reduces the gate capacitance. For this reason silicon gate SOS circuits will switch faster than aluminum gate circuits. With a typical gate length of 0.3 mils for a silicon gate vs. 0.5 mils for an aluminum gate, all of the previous propagation delays would be multiplied by 0.6. Delays of about 500 ps per inverter stage have been reported (43) with self-aligned gate structures using channel lengths of 3 μm, and 1100 ps with channel lengths of 5 μm, both with 10 volt supply voltages.

The propagation delay for bulk CMOS is strongly affected by the drain-substrate capacitance since there is junction capacitance between the entire drain region and substrate. If the drain region measures 1x0.6 mils, the drain-substrate capacitance at 0.1 pF/mil^2 would be 0.06 pF, which is equal to the total SOS inverter capacitance. Thus the drain-bulk capacitance alone adds the equivalent of one logic load to the circuit, increasing the propagation delay accordingly.

The interconnect capacitance in bulk CMOS is also significant. Unlike SOS, all bulk CMOS conductors form a capacitance with the underlying substrate. If the field oxide is 10,000 Å thick this capacitance will be 0.023 pF/mil^2. Hence, every 2.5 mil^2 of interconnection will be equivalent to the input capacitance of one SOS inverter.

F. Power Dissipation

Power dissipation in CMOS circuits can be divided into quiescent power and dynamic power. The quiescent power is very low and is due to leakage currents in the devices. Dynamic power is supplied when the circuit is switching and is due to the power required to drive the internal circuit capacitance and to the power supply current spikes which occur when both p and n-channel transistors are simultaneously conducting during the switching transition.

It has been shown that the quiescent power dissipation of CMOS circuits is due to the drain current of the off transistor (47). If the transistor is well below threshold voltage, for example, $V_{GS} = 0$, the sub-threshold current will be small compared to the drain leakage current. Typical values of leakage current are 80 pA and 150 pA per mil of transistor width for n and p-channel SOS devices, respectively (47). A 500 node circuit with 0.3 mil wide transistors operated at 10 volts would have a quiescent power dissipation of about 200 nW on this basis.

The portion of the dynamic power dissipated in the transistors during the switching transient has been determined to be small compared to the power required to replenish the charge on the circuit capacitances (43,47). This is the classical CV^2f power, where C is the total capacitance being switched, V is the voltage swing (usually V_{DD}), and f is the switching frequency. The previously calculated node capacitance for an inverter driving 2 logic gates with 2 inputs each was 0.171 pF. For a typical circuit with 500 nodes and $V_{DD} = 10$ volts the dynamic power dissipation is $8.55 \times 10^{-9} f$ watts. At 1 MHz the dynamic power would be 8.55 mW.

G. Bulk CMOS Layout

To show the possibility of realizing greater circuit density with SOS, the layout of a CMOS inverter in both SOS and bulk MOS technologies will be illustrated. Figure IV-12 shows a typical aluminum gate bulk CMOS inverter

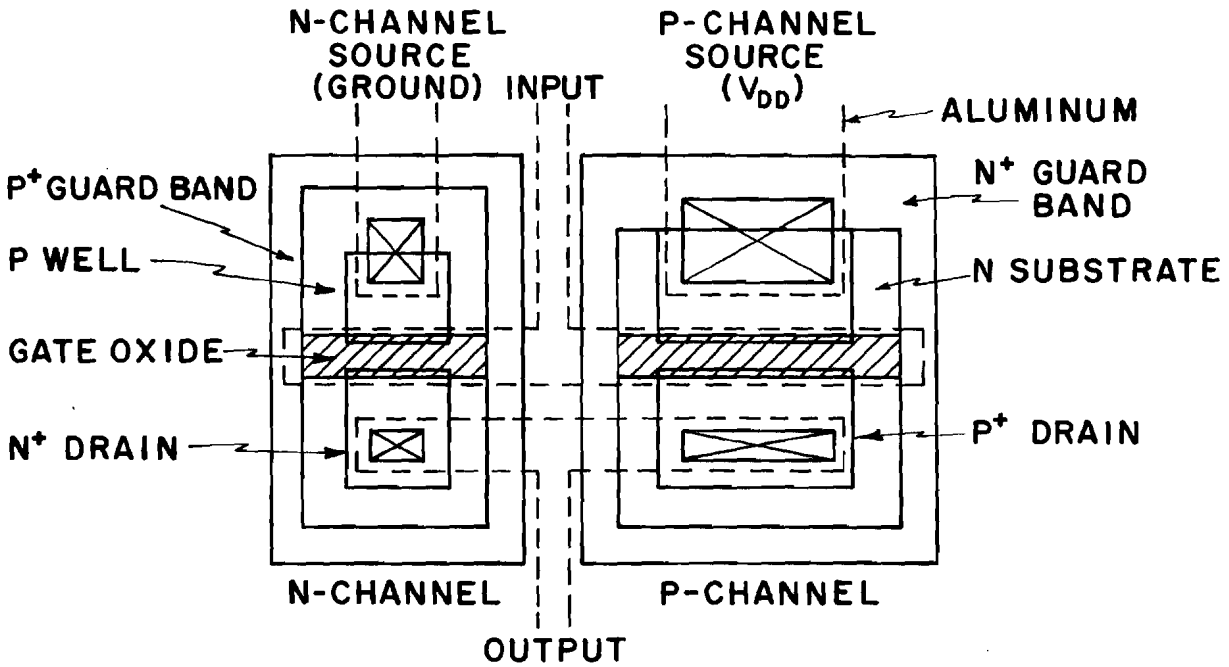


Figure IV-12. Bulk CMOS inverter layout.

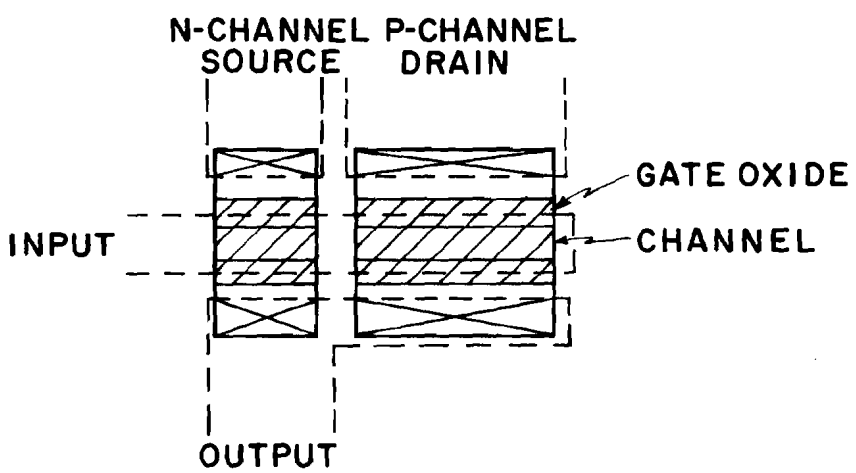


Figure IV-13. Aluminum gate SOS CMOS layout.

pair. The p-channel transistor has a channel width of 1.8 mils and the n-channel, 0.9 mils. The difference in widths is to balance the transconductance, since electron mobility is about twice the hole mobility. These are representative sizes for logic transistors which drive other gates on the same chip. Output transistors are usually much larger.

Note that each transistor is surrounded by a channel stop or guard ring to prevent field inversion current flow. The gate oxide extends from the side of the channel to the edge of the guard ring so that the silicon surface potential between the transistor and the guard ring will be completely controlled by the gate conductor at all times. The gate conductor overlaps the guard ring to complete the channel blocking effect. The guard rings are spaced 0.5 mils from the transistor drain regions to prevent reduced breakdown voltage. Not shown in the figure is the field oxide and the oxide which covers the remainder of the inverter.

The n-channel transistor is located entirely within a diffused p-well which serves as the substrate and channel region. The n-channel source is connected to the p-well by the overlapping contact shown in the figure. This grounds the p-well to prevent back-gate threshold effects and 4-layer diode action. The p-channel source is connected to the n^+ guard ring (and hence to the n substrate) for the same reasons.

The area of this inverter is 22 mil^2 measured from the outside edges of the guard rings. In a more complex layout some transistors of like polarity could share p-wells and guard rings. Further area reduction would occur with a silicon gate process, but probably not more than 40%.

Many low voltage bulk CMOS applications, such as watch and clock circuits, eliminate the guard rings since the applied voltages are lower than the field inversion threshold voltage. However, the maximum supply voltage which can be safely used in such specialized cases is 5 volts.

H. SOS CMOS Layout

The layout of an aluminum gate SOS CMOS inverter is shown in Fig. IV-13. The channel widths are 1.8 and 0.9 mils for the p-channel and n-channel transistors, respectively. Each transistor is an island of silicon, with silicon completely removed around each trasistor. Apart from the sapphire substrate, the only other materials are the thin gate oxide

and the aluminum gate conductor and interconnect system.

The most distinctive feature is the simplicity of the entire structure, with the complete lack of wells, guard rings, etc. Note also that where the gate oxide is etched away to allow for ohmic contact between the aluminum and the source and drain regions bare silicon is sometimes exposed. This is contrary to all other integrated circuits where the aluminum contact always overlaps the edges of the contact window. Also note that no provision is made for contacting the silicon in the channel region. As mentioned earlier, this allows the "substrate" to float and produces the kink effect.

The area of the inverter as shown is only 6.7 $mils^2$, measured from the edges of the silicon islands. In an actual circuit the average inverter area is considerably larger to allow for spacing between devices, aluminum interconnects, plus space for bonding pads, scribe lanes, alignment marks, etc. A silicon gate process allows some saving in chip area by allowing smaller devices and a second layer of interconnects.

I. Linear Circuits

At first glance there appears to be little reason to utilize SOS technology for linear circuits. The gain per unit area is less than that of bipolar circuits, with lower frequency response. The power dissipation is no lower than bipolar when operated in a linear mode. In addition the noise is considerably higher.

The kink effect results in low amplifier gain. This can be corrected by connecting the channel to the source, but at the expense of greater chip area. Also the leakage current is higher than bulk MOS (a disadvantage for balancing differential amplifiers).

In spite of these drawbacks one group (60) has pointed out certain advantages of linear SOS circuits. In contrast to bipolar integrated circuits they will operate at cryogenic temperatures, and are much more tolerant to transient radiation effects. Frequency response is better than ordinary CMOS. Also, circuit elements can be interconnected in any fashion since there is no common conducting substrate. This means that zeners, temperature compensating diodes, etc., can be readily employed.

As yet, no commercial products are planned for the forseeable future.

SECTION V

SOS CIRCUITS AND PERFORMANCE CHARACTERISTICS

Many silicon on sapphire circuits are available as pin for pin replacements for bulk CMOS circuits. This permits a direct comparison between bulk CMOS components and circuits and their SOS counterparts. Although the advantages of SOS are not necessarily realized in small scale integrated circuits, detailed comparisons can be more readily made at this level. Some large scale integrated circuits will also be discussed, together with such performance characteristics as are available at this time. Bear in mind that SOS technology is still in its infancy, with new designs and improvements constantly being made. Nevertheless the facts and figures presented here are useful in making a meaningful assessment of its potential.

A. Bulk CMOS Dual Complementary Pair Plus Inverter

A schematic of a bulk CMOS device is shown in Fig. V-1. It is a dual complementary pair plus inverter consisting of three pairs of n and p-channel transistors. One pair is connected as an inverter, while the other two pairs have common gates. Note that all three substrates of each type of transistor are connected together.

This circuit is available in a welded-seal dual-in-line ceramic package, a frit-seal dual-in-line ceramic package, a welded-seal ceramic flat pack, and a dual-in-line plastic package. Several of the welded-seal, DIP parts were ordered. One part, with a 408 date code was opened. A photomicrograph appears in Fig. V-2.

In this particular package the chip is epoxy bonded within a round well 105 mils in diameter. The chip is 7 mils thick, and measures 42 x 42 mils. Clearly visible in the photograph are the three p-channel enhancement mode transistors (the large interdigitated devices at right center), the three n-channel enhancement mode transistors (the three smaller devices at left center), an n-channel test transistor (between pads 12 and 13), a p-channel test device (between pads 7 and 8), three resistors forming part of the input gate protection circuits (connected to pads 3, 6, and 10), and four alignment keys. The main alignment key is located between pads 6 and 7. A second alignment key between pads 9 and 10 facilitates contact alignment to

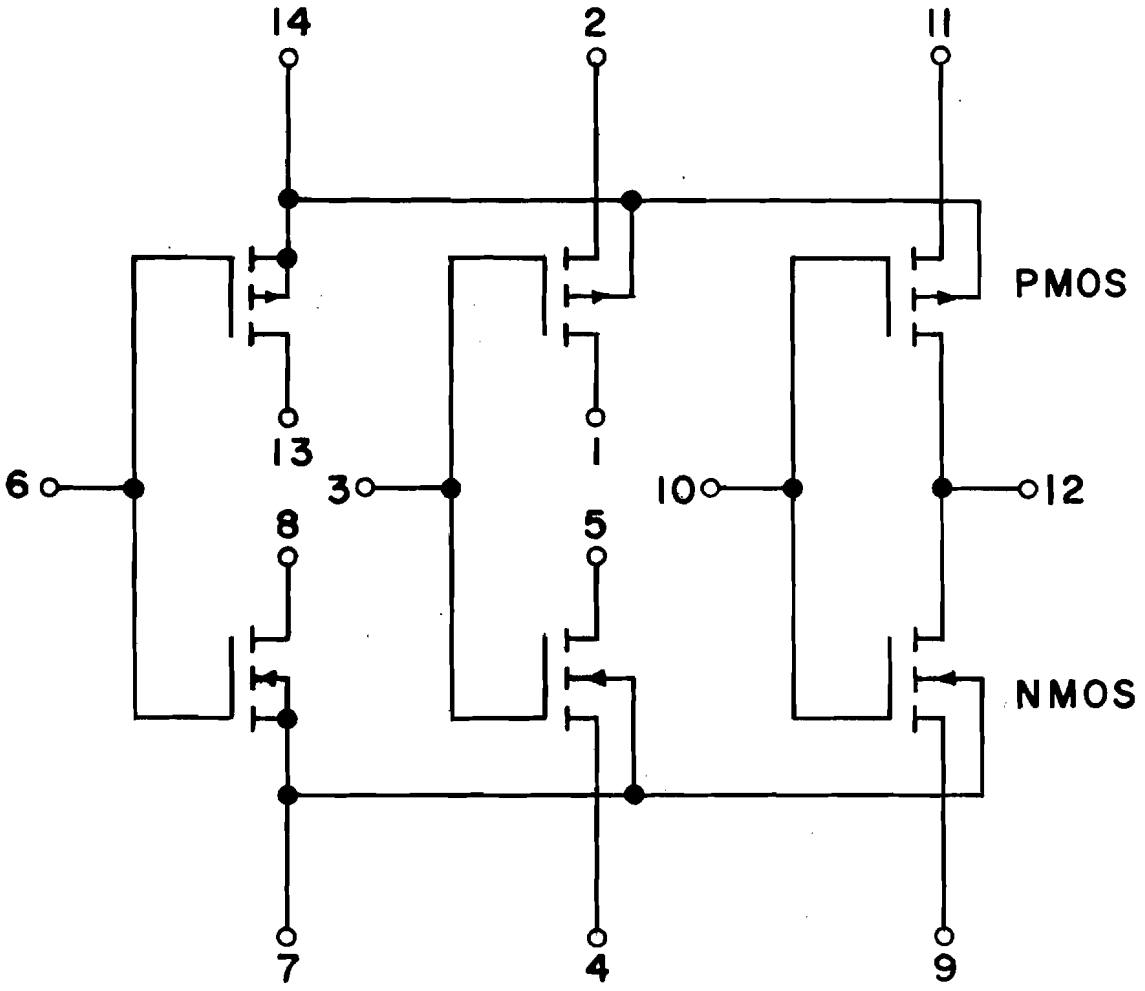


Figure V-1. Schematic of a bulk CMOS dual complementary pair plus inverter.

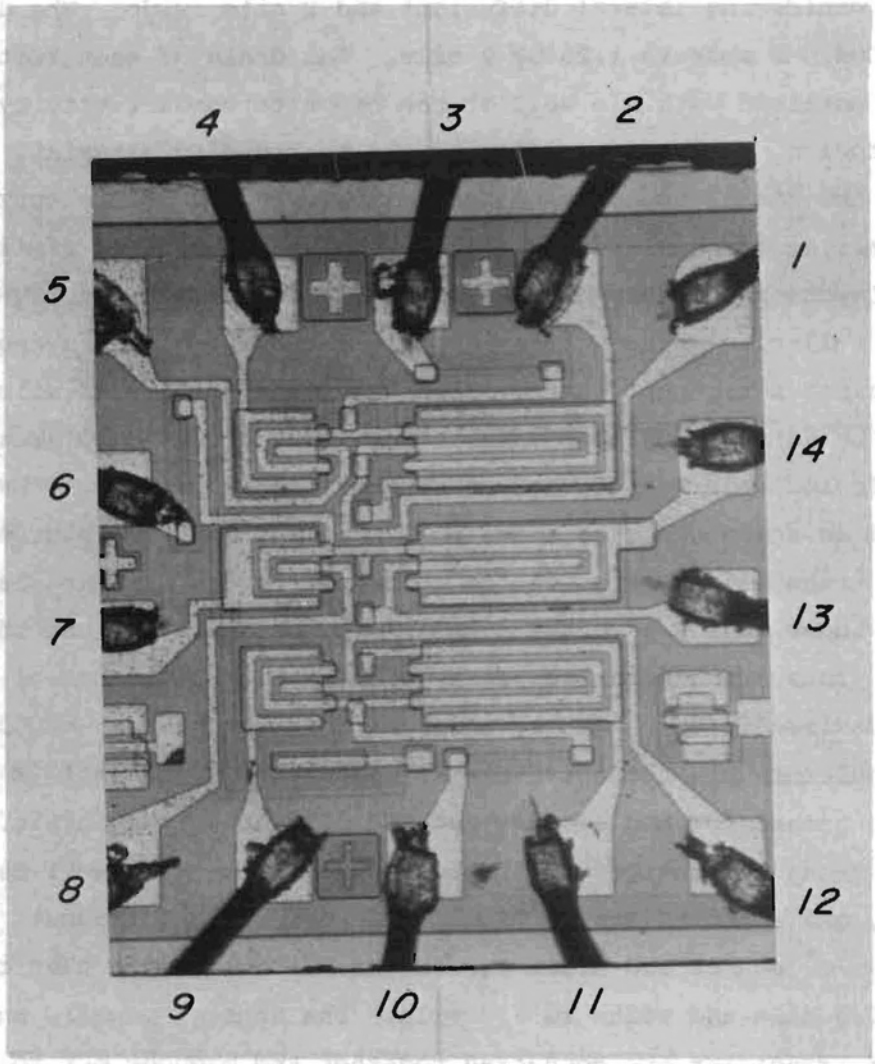


Figure V-2. Photomicrograph of a bulk CMOS dual complementary pair plus inverter circuit.

p^+ and n^+ diffusions. The alignment keys between pads 2 and 3, and between 3 and 4 facilitate alignment of n^+ to p^+ regions, and metal to n^+ and p^+ regions, respectively.

The small test transistors both have channels which measure 0.3 mils long (not considering lateral diffusion) and 2 mils wide. The drain and source diffusions measure 1.25 by 2 mils. The drain of each test transistor is wholly contained within a well of the opposite conductivity type while the source diffusion overlaps the well and the surrounding material. The contact window for the source is large enough to include part of the source diffusion and the material surrounding the well. This arrangement is similar to the bulk CMOS layout previously discussed and illustrated in Fig. IV-12.

The full-size p-channel devices are fabricated in an n-region which is the original n-type epi layer. Each well measures 5 x 11 mils and is surrounded by an n^+ guard band which covers most of the chip area. The p^+ source diffusion is U-shaped, surrounding the drain region on three sides. The channel is again 0.3 mils long, and 20 mils wide. The source region of the center transistor in Fig. V-2 is connected to the n^+ guard band through a contact window adjacent to the n-well. The metalization run then continues to pin 14. This arrangement serves to connect the substrates of all three p-channel devices to V_{DD}. (Note, however, that the sources and substrates of two transistors are not automatically connected together.) A diagram of a p-channel transistor and a cross section view are shown in Fig. V-3.

The n-channel devices are somewhat smaller. The p-well is 5 x 5 mils, 10 μm deep, and doped to $5 \times 10^{16}/cm^3$. The p-well is surrounded by a p^+ guard ring. n^+ source and drain regions are diffused such that the channel length is 0.3 mils and width is 7.7 mils. The source, p-well, and p^+ guard ring of one transistor are connected together and brought out to pin 7. A diagram of an n-channel transistor is shown in Fig. V-4.

Gate input protection devices consist of a series resistor, a distributed diode connected to V_{DD}, and a discrete diode connected to ground. The diagram and circuit layout are shown in Fig. V-5. The resistor is formed from a p^+ diffusion into the n epi layer, which is surrounded in turn by the n^+ guard band. With a sheet resistivity of 50 ohms per square this resistance would be about 800 ohms. Since the n^+ guard band is connected to V_{DD}, there is a distributed diode between the resistor and V_{DD} which has a breakdown

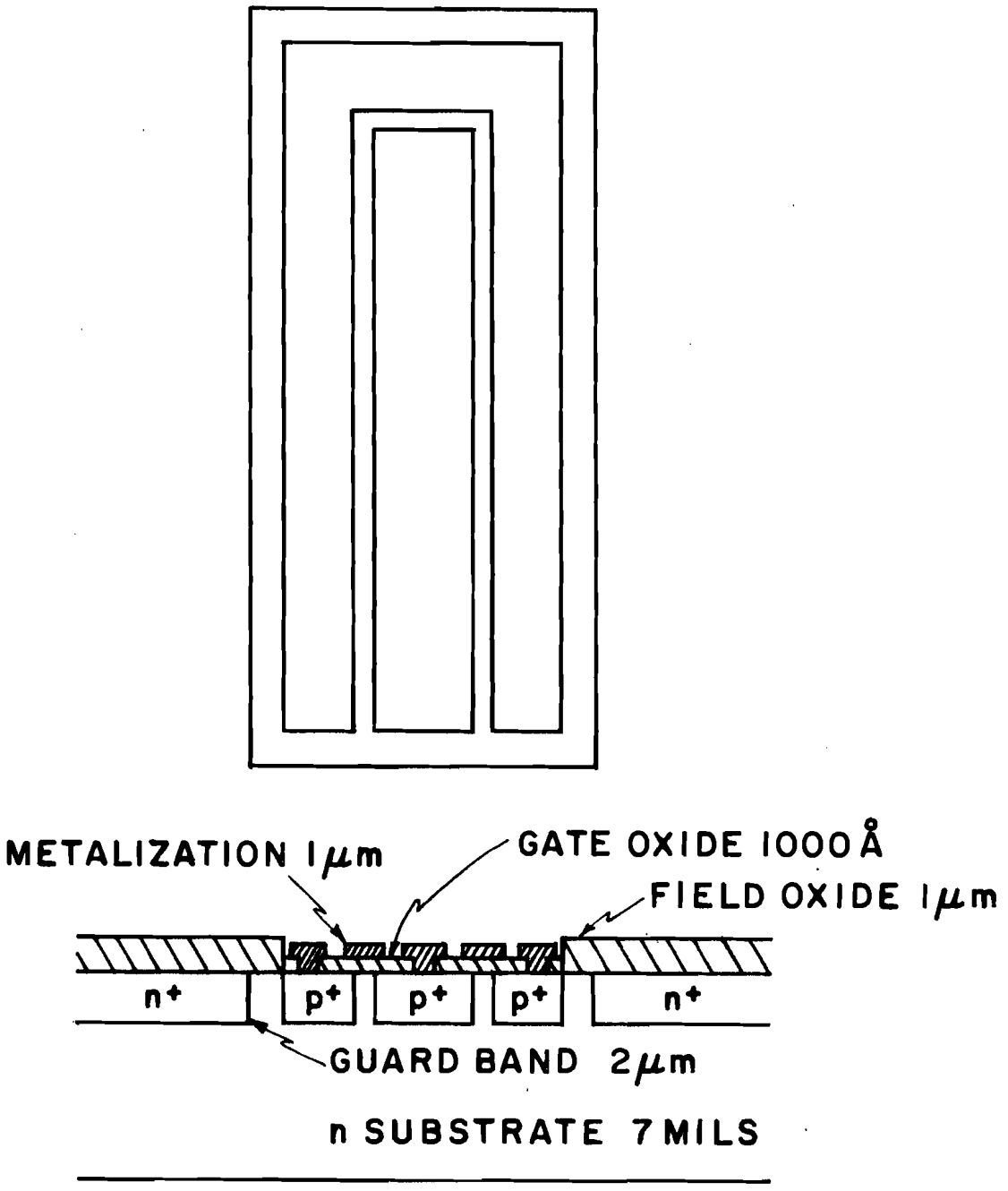


Figure V-3. Layout and cross section of the p-channel transistor.

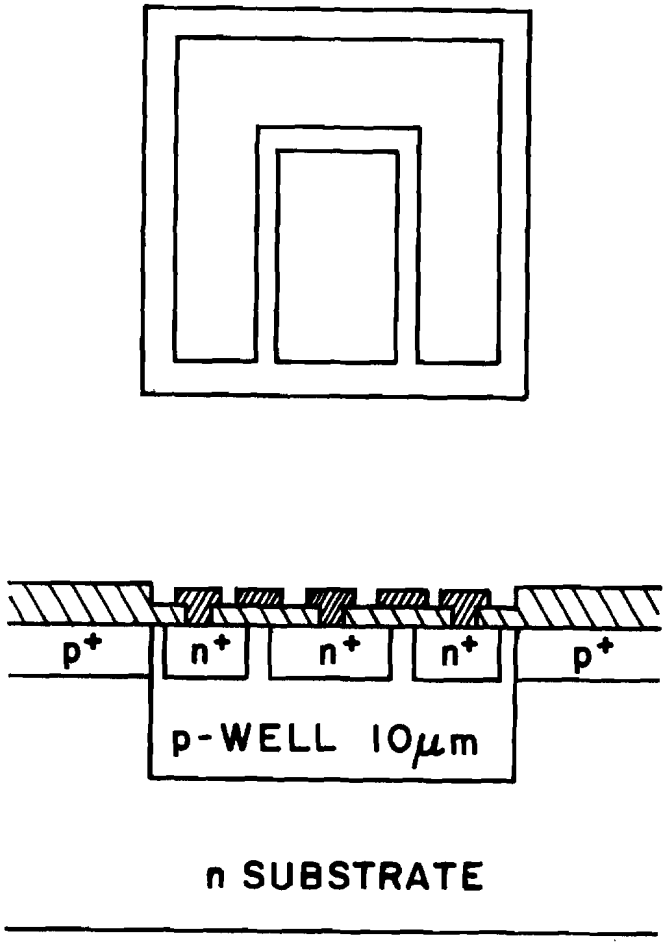


Figure V-4. Layout and cross section of the n-channel transistor.

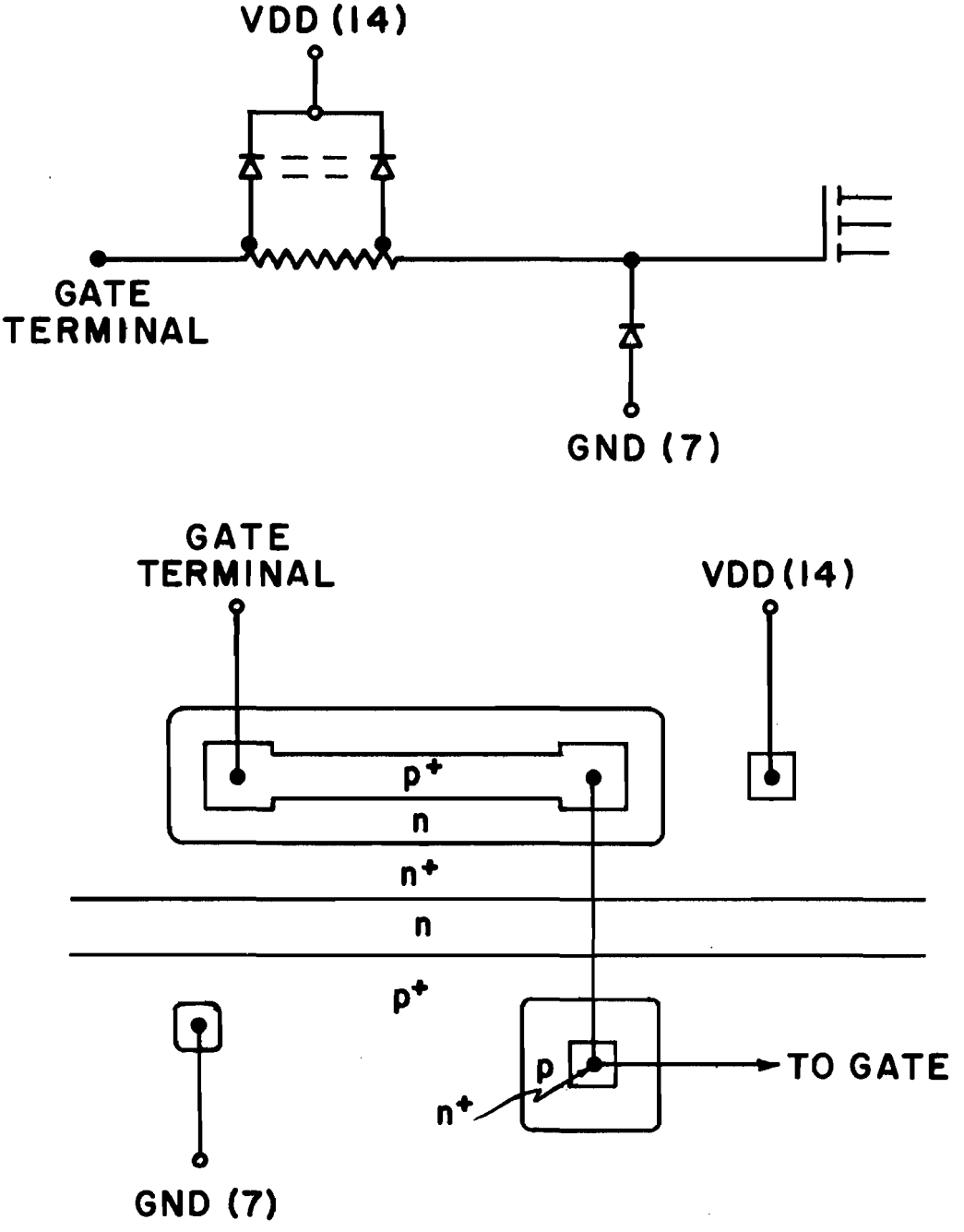


Figure V-5. Diagram and layout of the gate protection circuit.

of roughly 50 volts (the n epi layer has a doping level of $5 \times 10^{15}/cm^3$). The diode from ground to gate is formed from an n^+ diffusion into a small p-well which is surrounded by p^+ material, and hence connected to ground. The breakdown voltage of this diode is about 20 volts (the p-well has a doping of $5 \times 10^{16}/cm^3$).

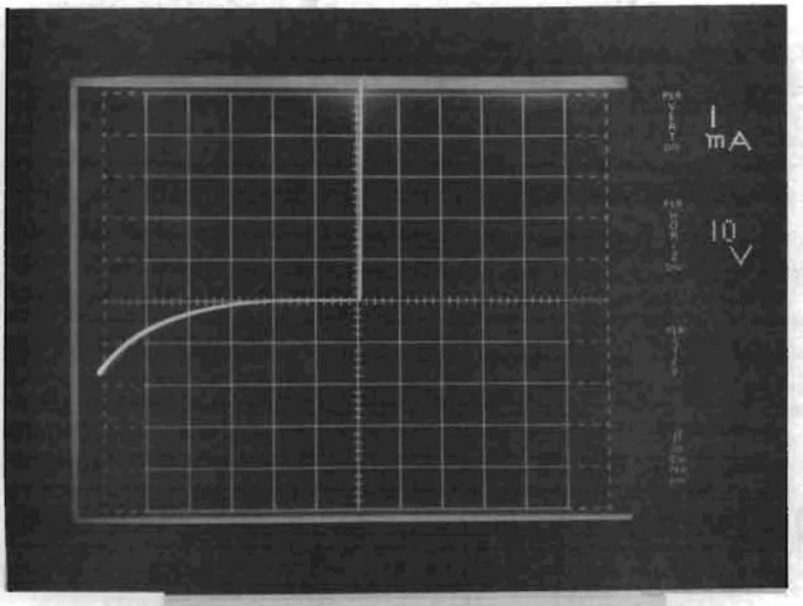
The characteristics of the protection network are shown in Fig. V-6. The upper photograph shows the input characteristic with respect to V_{DD}, while the lower photograph shows the characteristic between the input terminal and ground. V_{DD} has a maximum allowable value of 15 volts, thus it can be seen that input transients greater than 15.5 volts or less than -0.5 volts will be shunted to V_{DD} or V_{SS}, respectively. Typically, gate oxides are capable of withstanding voltages up to 100 volts, so an adequate margin of safety is provided by this protection circuit.

There are several other diodes which restrict the sources and drains from going negative. The diodes are the same as the gate protection diodes between gate and ground, formed by n^+ diffusions into p-regions, hence to the p^+ guard band and to ground. Figure V-7 shows the diodes, discrete and parasitic, which exist for a dual pair.

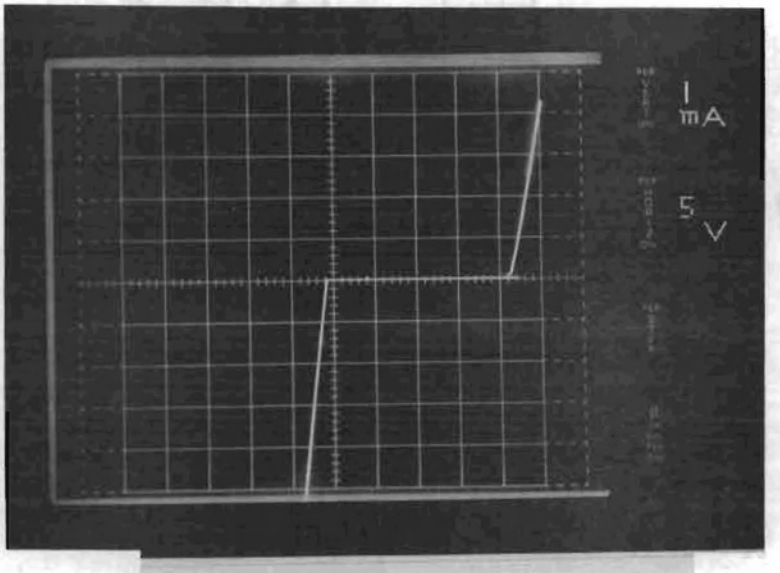
Table V-1 summarizes the structural features of this circuit.

Table V-1. Characteristics of the Bulk CMOS Process Used for the Dual Complementary Pair Plus Inverter

chip size	7x42x42 mils	contact windows	0.4 mils wide
substrate doping	$5 \times 10^{15}/cm^3$	street width	4 mils
p-well depth	10μm	p^+-n breakdown voltage	50 volts
p-well doping	$5 \times 10^{16}/cm^3$	n^+-p breakdown voltage	20 volts
n^+diffusion depth	2μm	gate oxide	1000 Å
n^+ sheet resistivity	5 ohms/square	field oxide	10,000 Å
p^+ diffusion depth	2μm	bonding pads	4x4 mils
metalization thickness	14,000 Å	effective channel length	0.2 mils
metalization width	0.5 mils	n-channel width	7.7 mils
metalization spacing	0.4 mils	p-channel width	20 mils

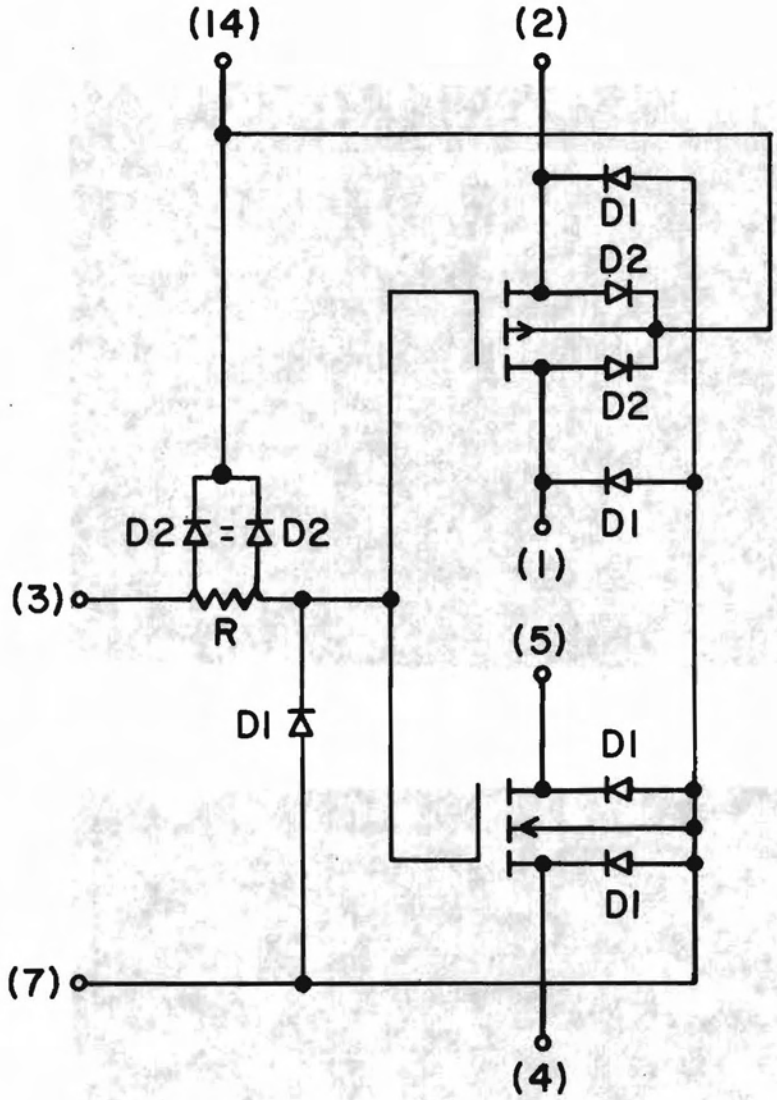


(a) input with respect to V_{DD}



(b) input with respect to ground

Figure V-6. Electrical characteristics of the gate protection circuit.



D1 n^+ TO p WELL
D2 p^+ TO SUBSTRATE
R = 1 K

Figure V-7. Equivalent circuit for a bulk CMOS complementary pair.

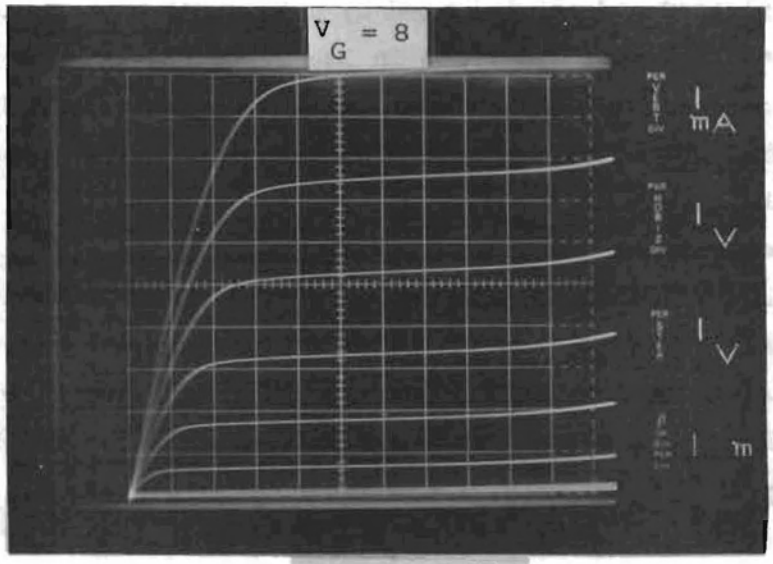
Electrical characteristics of the transistors are shown in Fig. V-8. All three transistors of each type had nearly indentical curves. The n-channel devices have a threshold voltage of 1.2 volts. If the transfer characteristic is plotted and approximated by a square-law characteristic at $V_{DD} = V_{GG} = 5$, $\beta_N = 12\ \mu A/volt^2$ (defined in Eq. IV-2) and $\mu_N = 340\ cm^2/volt\text{-}s$. For the p-channel devices $V_T = 1.0$ volts, $\beta_p = 6.3\ \mu A/volt^2$, and $\mu_p = 180\ cm^2/volt\text{-}s$. Breakdown voltage for the n-channel transistor was 25 volts, and 28 volts for the p-channel transistor.

Much has been said in the literature about the "kink effect" in SOS transistors, and that it could be eliminated by grounding the substrate, meaning the material in the gate region below the induced channel. In bulk MOS similar effects can be seen if the p-well or n-substrate is left floating. Figure V-9 shows this effect for the identical transistors whose characteristics appear in Fig. V-8. Kinks and loops are present, and the overall characteristics are difficult to distinguish from those of SOS devices appearing elsewhere in this report.

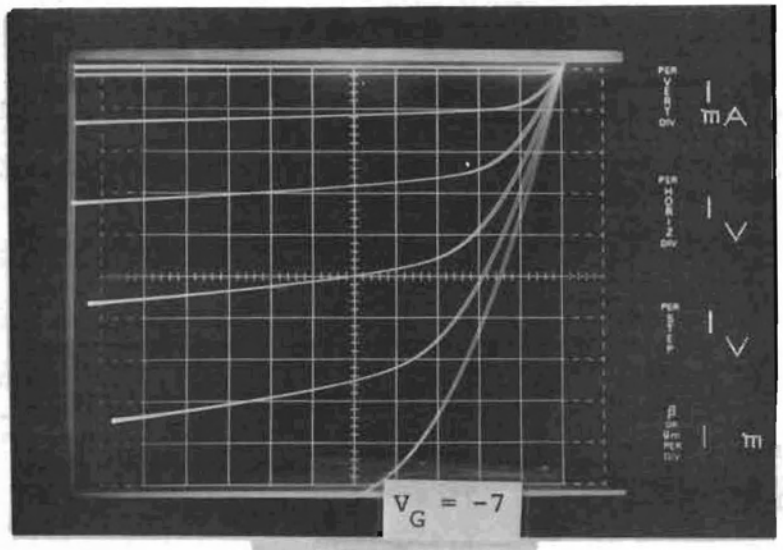
A summary of the electrical characteristics of this circuit appears in Table V-2.

Table V-2. Electrical Characteristics and Specification of Bulk CMOS Devices in the Dual Complementary Pair Plus Inverter

β_N	$12\ \mu A/volt^2$
β_P	$6.3\ \mu A/volt^2$
μ_N	$340\ cm^2/volt\text{-}s$
μ_P	$180\ cm^2/volt\text{-}s$
breakdown voltage n-device	25 volts
breakdown voltage p-device	28 volts
V_T n-device	1.2 volts
V_T p-device	1.8 volts
breakdown voltage $n\text{-}p^+$	50 volts
breakdown voltage $n^+\text{-}p$	20 volts
p^+ sheet resistivity	50 ohms/square
drain current n-device	0.43 mA/mil width at $V_{DD}=V_G=5$
drain current p-device	0.25 mA/mil width at $V_{DD}=V_G=5$

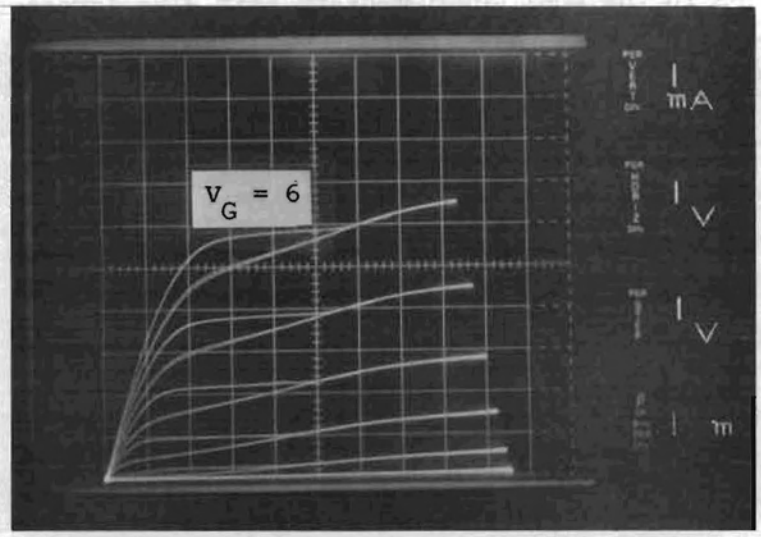


(a) n-channel

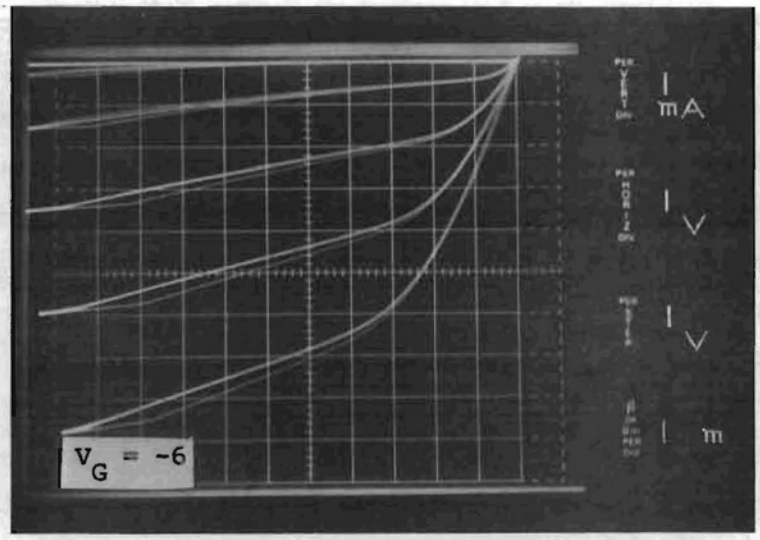


(b) p-channel

Figure V-8. Drain characteristic curves for transistors in the bulk CMOS dual pair.



(a) n-channel



(b) p-channel

Figure V-9. Kink effect in bulk CMOS transistors with floating substrates.

Table V-2 (continued)

Table V-2. Electrical Characteristics and Specification of Bulk CMOS Devices in the Dual Complementary Pair Plus Inverter

capacitance p^+-n, 15 volts	0.04 pF/mil^2
capacitance n^+-p, 15 volts	0.12 pF/mil^2
capacitance, metal over field oxide	0.023 pF/mil^2
capacitance, metal over gate oxide	0.23 pF/mil^2
quiescent dissipation/package	10 nW at 10 volts
dissipation/gate 1 kHz, 10 volts, 15pF	3.5 μW
dissipation/gate 10 MHz, 10 volts, 15 pF	35 mW
propagation delay	20 ns at V_{DD}=10, C_L=15 pF
transition time	30 ns at V_{DD}=10, C_L=15 pF

B. SOS Dual Complementary Pair Plus Inverter

A silicon on sapphire version of the circuit just described is available from Vendor D. A photomicrograph appears in Fig. V-10. Samples which were examined were packaged in a 14 pin frit-sealed ceramic dual-in-line package, date code 7403. Parts can also be supplied in a welded-seal DIP or a plastic DIP. The chip, which measures 48 x 50 x 12 mils thick, is epoxy bonded to the package. Connections to the package leads are by ultrasonically bonded aluminum wire.

Although this circuit is now made by a metal gate, ion implantation process, the circuits discussed here were fabricated by the double epi method. Recall that with this process, an n-type epi layer is grown for the p-channel devices. After the n islands are defined, the rest of the n layer is removed. A p-type layer is then grown, and etched away everywhere except where the n-channel transistors will be. Boron and phosphorus doped glasses are the diffusion sources for the source and drain regions. The fabrication then proceeds as in any other standard process: gate oxide, ohmic etch, metalization, and passivation.

Referring again to Fig. V-10, the devices visible on the chip are three n-channel enhancement mode transistors (at the top of the figure),

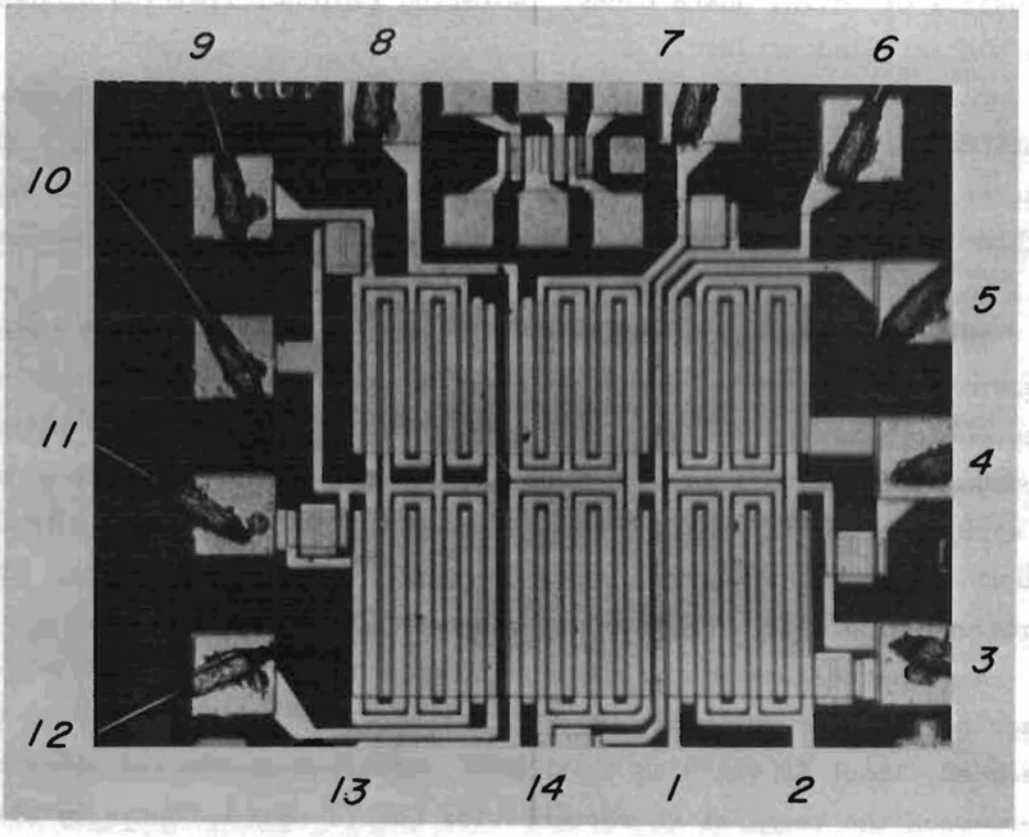


Figure V-10. SOS CMOS dual complementary pair plus inverter.

three p-channel transistors, six gate protection devices, and a test structure (between pads 7 and 8). Not quite visible are the mask alignment patterns between pads 2 and 3. The chip is inherently simpler because of the absence of wells, guard bands, isolating regions, limiting diodes, extra contact windows, etc.

The test structure consists of a 2x2 mil p-channel transistor, plus a 2x3 mil device which has an n-channel transistor and a separate p^+ diffusion at one end. Channel length is 0.3 mils (less lateral diffusion). Gate metalization is 0.5 mils wide. Source and drain minimum metalization width is 0.3 mils.

The gate protection device is a series connection of 2 p-channel transistors, 2 mils wide, with 0.3 mil channels. These transistors are connected in series. The gates are connected together, and are connected to the common terminal of the two transistors. Two protection devices are used with each complementary pair. One is connected between the gate input terminal and V_{DD}, the other between the gate terminal and ground. The structure and circuit schematic are shown in Fig. V-11.

A p-channel transistor with gate connected to source conducts only leakage current for any value of drain voltage, until the breakdown voltage is reached, about 50 volts in this case. With two such transistors connected back-to-back the terminal characteristics should show essentially zero current, then a breakdown characteristic when the voltage reaches $\pm$ 50 volts. This is shown in Fig. V-12. The limiting voltage is slightly higher than the breakdown voltage of the 3 regular p-channel transistors on this chip, but exhibits the walk-out which is characteristic of p-channel SOS transistors. To verify the composite characteristic a dual pair was connected in the same configuration as the protection device. The terminal characteristics were the same, but the limiting voltage was 10 volts less in magnitude for either polarity. This is probably due to the much larger size of the regular p-channel transistors and the normal variation in breakdown voltage in lightly doped material.

When a dual pair is connected as a simple inverter with the source of the n-channel transistor grounded and the source of the p-channel transistor connected to V_{DD}, the common gate is restricted by the protection devices

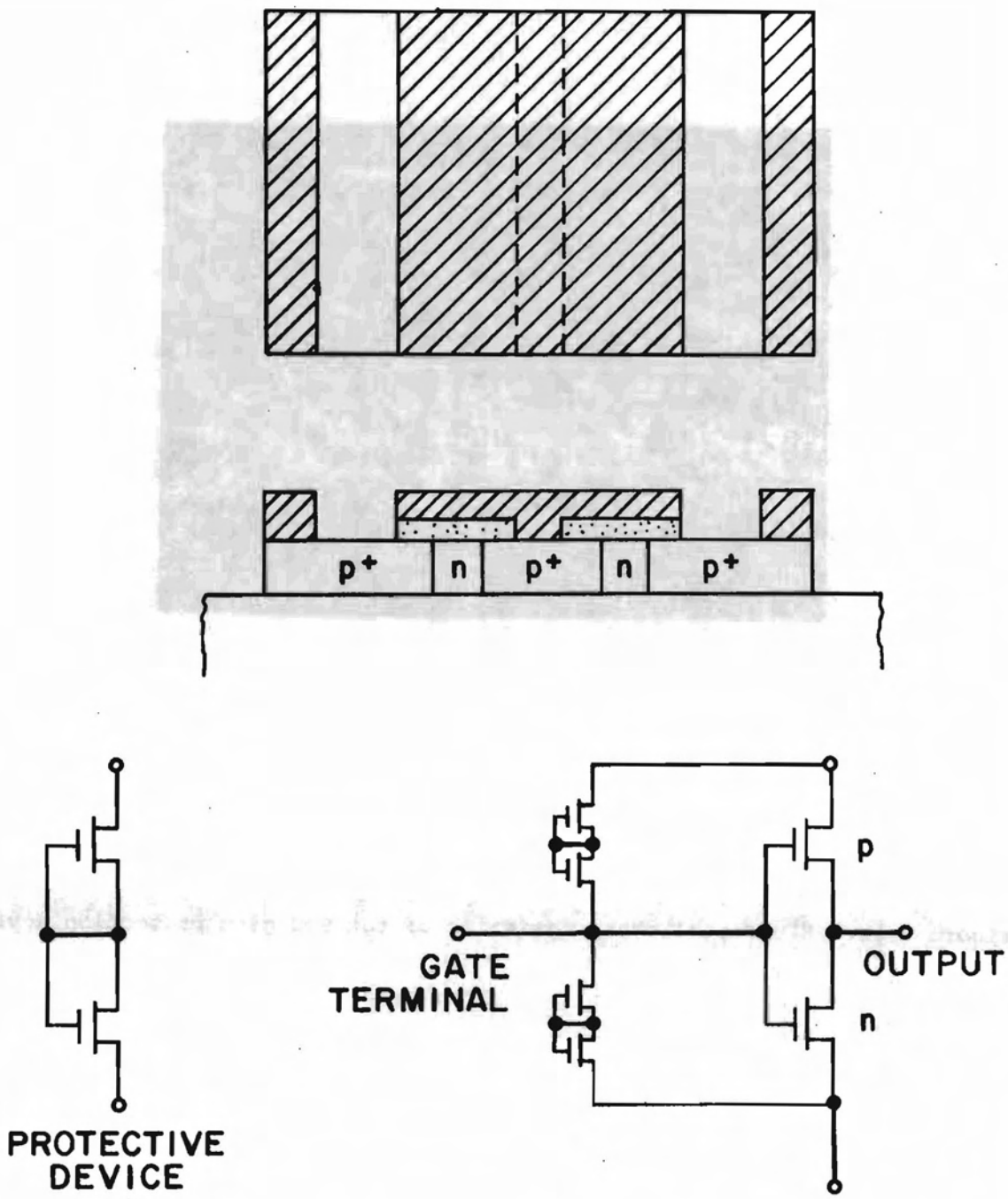


Figure V-11. Gate protection device for the SOS CMOS dual pair plus inverter.

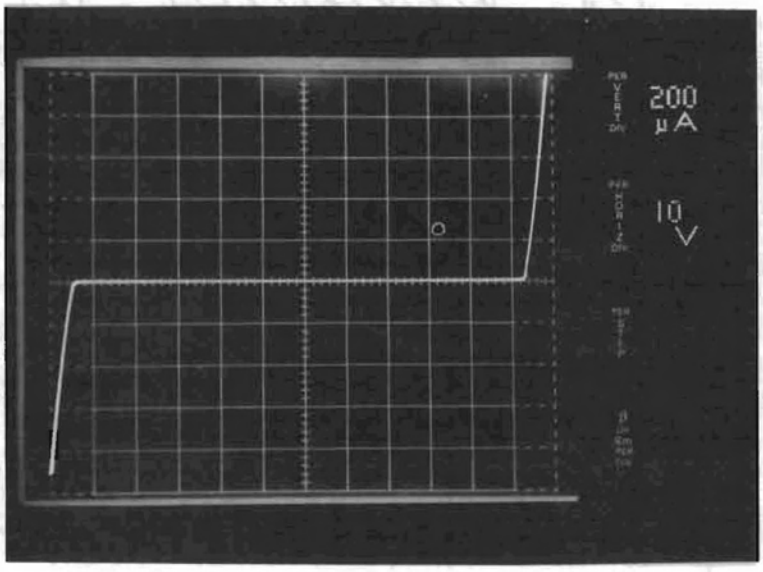


Figure V-12. Electrical characteristics of the SOS gate protection device.

to voltages between +50 and $-(50-V_{DD})$. For input transients more positive than +50 volts, the transient is conducted to ground. For voltages more negative than $-(50-V_{DD})$, the transient is conducted to V_{DD}.

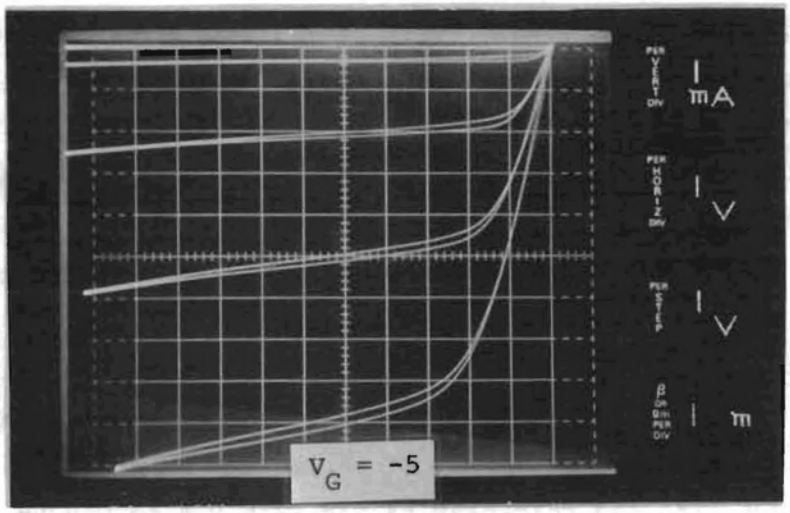
The silicon islands for the full-size p-channel transistors are 7.5x 10 mils. The source, drain, and gate of each transistor are interdigitated to provide a total channel width of 50 mils. As in the test transistors, all channels are 0.3 long, with 0.5 mil gate metalization and 0.3 mil source and drain metalization. The metal to silicon contacts are 0.3 mils wide. Drain characteristics and a construction diagram for a p-channel transistor are given in Fig. V-13. The epi layer originally has a nominal thickness of 1 μm. After processing the thickness ranges from 0.4 to 0.8 μm. Gate oxide is 1200 Å, while the metalization is about 1 μm.

Threshold voltage for the p-channel devices is 1.0 volts, with a breakdown of 45 volts. Approximating the drain current at $V_{DD} = V_G = 5$ volts by a square-law characteristic, $\beta_P = 4.5$ μA/volt^2, and $\mu_P = 150$ cm^2/volt sec. These are about the same as the values obtained for the bulk CMOS p-channel device.

The n-channel transistors have the same structure as the p-channel except that the island dimensions are 7.5x8 mils. Threshold voltage is 0.5 volts, and breakdown voltage is 50 volts. Calculated values of β_N and μ_N are 3.3 μA/volt^2, and 110 cm^2/volt sec, respectively. Drain characteristics are shown in Fig. V-14.

The values of β_N and μ_N are lower than bulk CMOS values, and are, indeed, lower than typical values given by the manufacturer. The drain current is not exactly a square law characteristic, and more typical results could probably be achieved with a more sophisticated $I_D - V_G$ relationship. The calculated values are sensitive to operating point, particularly near the edge of the saturation region. Furthermore, there was a significant variation in characteristics among 5 units which were tested. Unfortunately, not enough testing was undertaken to make definitive statements on this aspect of the electrical performance.

Characteristics are summarized in Tables V-3 and V-4. Table entries represent measured values or characteristics supplied by the manufacturer.



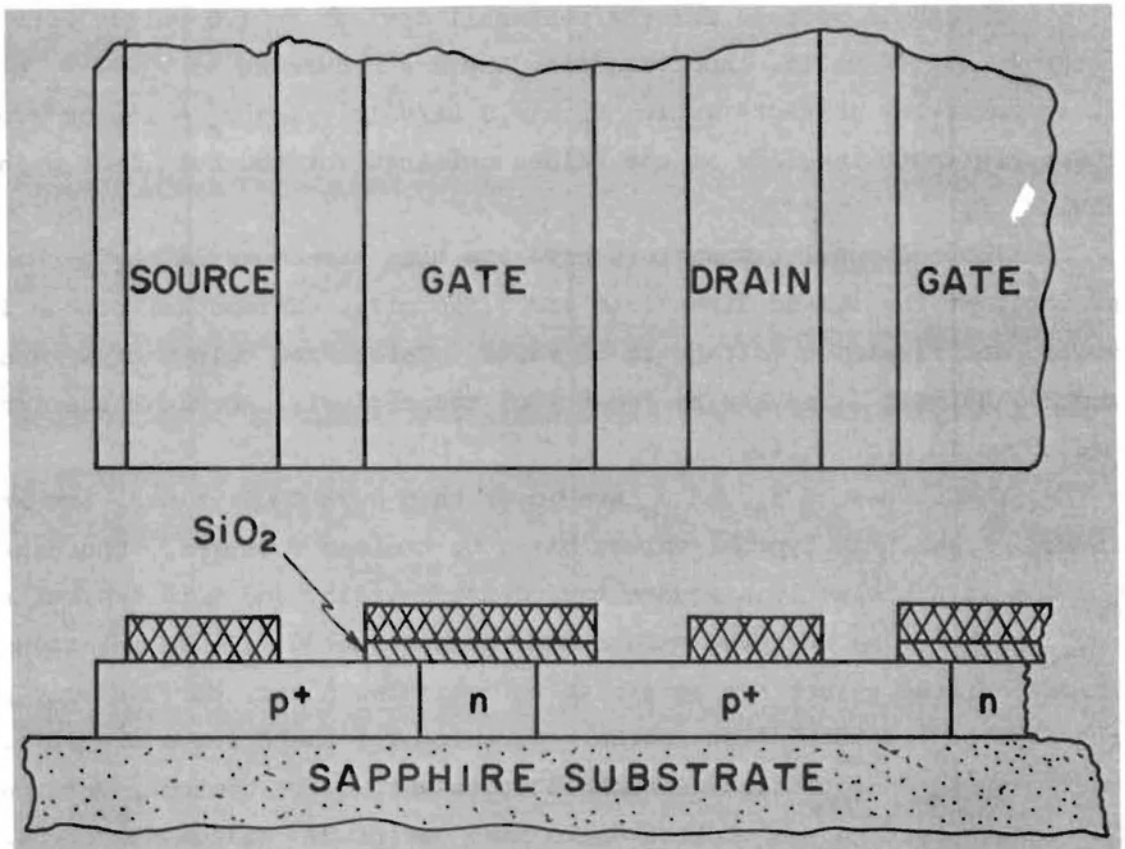


Figure V-13. Electrical characteristics and construction of p-channel SOS transistor.

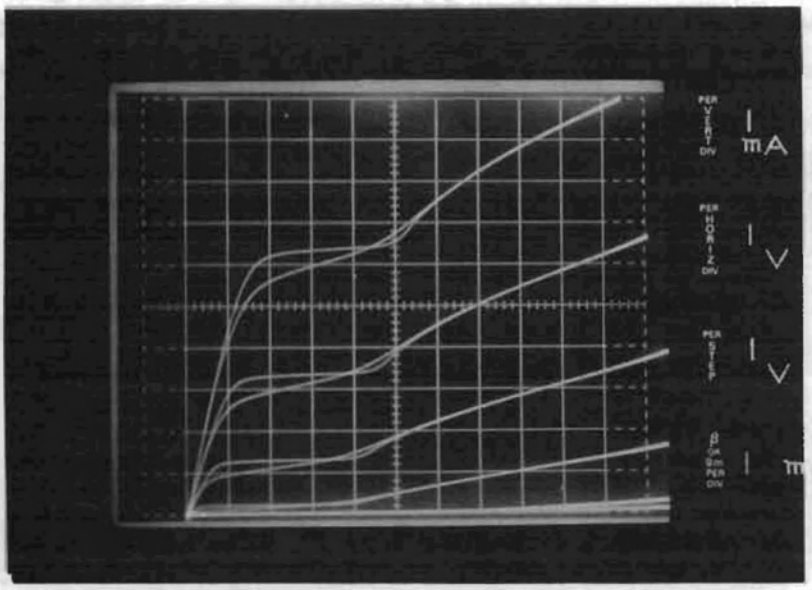


Figure V-14. Electrical characteristics of an SOS n-channel device.

Table V-3. Structural Characteristics of an SOS Dual Complementary Pair Plus Inverter

chip size	48x50x12 mils thick	gate metaliz. width	0.5 mils
epi thickness	0.65 μm	min. metaliz. width	0.3 mils
gate oxide thickness	1200 Å	contact width	0.3 mils
metalization thickness	1 μm	n-channel width	40 mils
effective channel length	0.2 mils	p-channel width	50 mils
p island size	7.4x10 mils	street width	5 mils
n island size	7.5x8 mils	bonding pads	4x4 mils

Table V-4. Electrical Characteristics of an SOS Dual Complementary Pair Plus Inverter

p-channel doping	$10^{16}/cm^3$	V_T p-device	1.0 volts
n-channel doping	$10^{15}/cm^3$	V_T n-device	0.5 volts
β_N	3.3 $\mu A/volt^2$	BV p-device	45 volts
β_P	4.5 $\mu A/volt^2$	BV n-device	50 volts
μ_N	110 cm^2/volt-s	drain current p-device	0.18 mA/mil width at $V_{DD}=V_G=5$
μ_P	150 cm^2/volt-s	drain current n-device	0.16 mA/mil width at $V_{DD}=V_G=5$

leakage current p-device	40 pA/mil width at V_{DD} = 10
leakage current n-device	100 pA/mil width at V_{DD}= 10
quiescent dissipation/package	60 nW at 10 volts
propagation delay	8 ns at V_{DD}= 10, C_L= 15 pF
transition time	10 ns at V_{DD}= 10, C_L= 15 pF

The type of circuit exemplified by the dual pair plus inverter illustrates some of the differences between SOS and bulk CMOS. The structure is less complicated, and delay time, transition time, and probably dynamic power dissipation are improved. Quiescent power dissipation is higher because of the higher leakage currents. Carrier mobilities and

β's are smaller, at least in the devices which were measured here. This is probably the reason for the larger transistor size, in order to provide camparable current drive capability and increase the speed. This, in turn, leads to a slightly larger chip. The detailed description of this circuit in its two varieties was not undertaken to determine the advantages and disadvantages of the two technologies, but rather to emphasize the differences between bulk CMOS and SOS processing, structure, and device characteristics.

C. Quad 2-input NOR Gate

A photomicrograph of an SOS quad 2-input NOR gate appears in Fig. V-15, with a schematic in Fig. V-16. Again the circuit construction is very simple. Each of the vertical structures contains 2 p-channel and 2 n-channel transistors connected in standard NOR gate fashion. The p-channel devices have a channel width of 54 mils - twice as long as the n-channel devices - in order to provide a more nearly symmetrical transfer function as explained in a previous section. Channel length is again 0.3 mil before lateral diffusion is taken into account.

The chip also contains a test structure (the same as discussed above) and 5 gate protection circuits. Four of these are used in conjunction with each gate input pair, and one is connected between V_{DD} and V_{SS} to restrict V_{DD} to positive values.

The protection device is different from the one previously discussed, and is commonly used in SOS circuits. A diagram of this device and the electrical characteristics are shown in Fig. V-17. With $V_{DD} = 15$, for example, the protection circuit would limit the gate input voltage between -15 and +20 volts. Note that input overloads are directed to V_{DD}. There is no protective device connected between the gate and ground as in the two types of dual complementary pairs plus inverters described earlier.

A comparison of this circuit and equivalent bulk CMOS circuits is given in Table V-5.

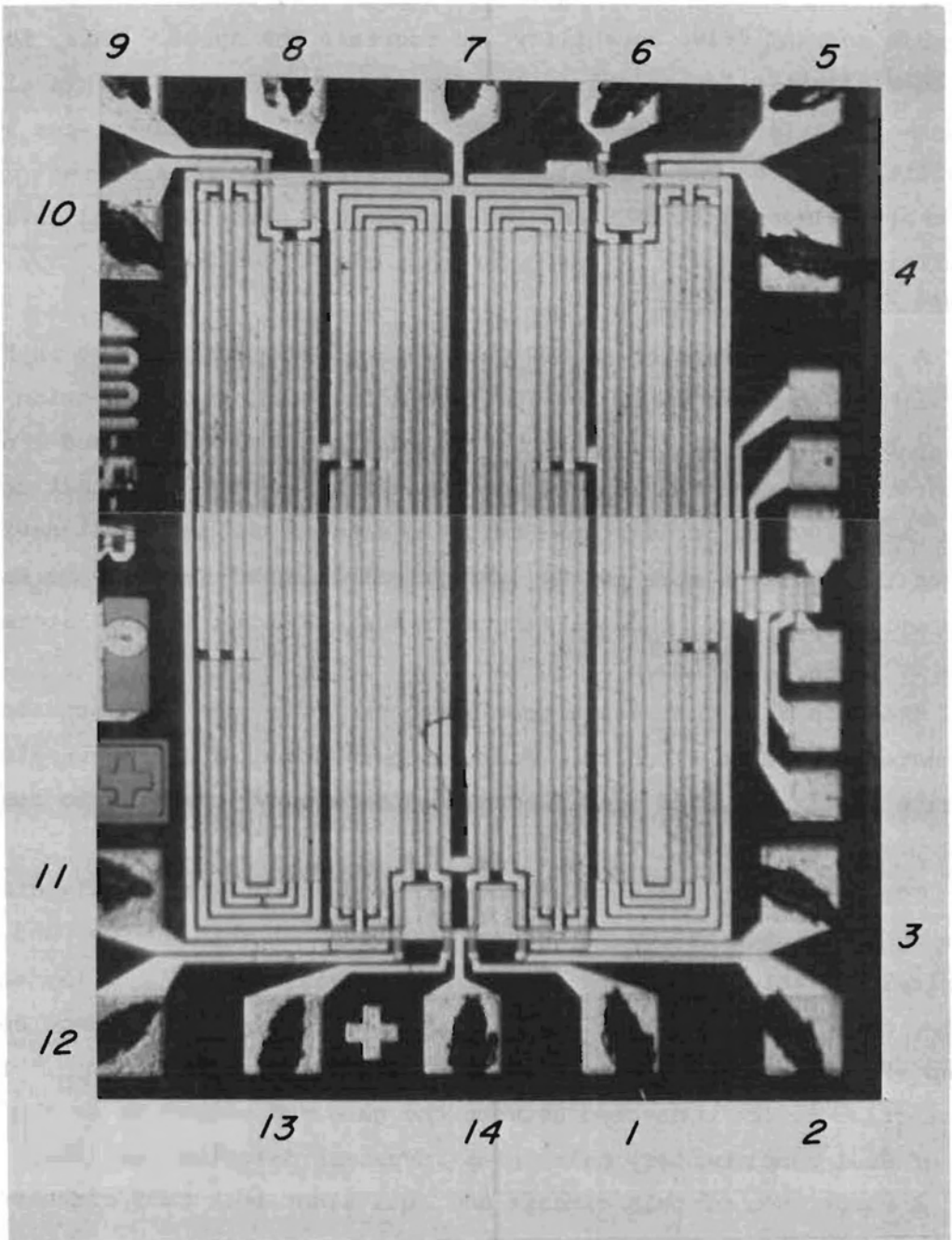


Figure V-15. Photomicrograph of a quad 2 input SOS NOR gate.

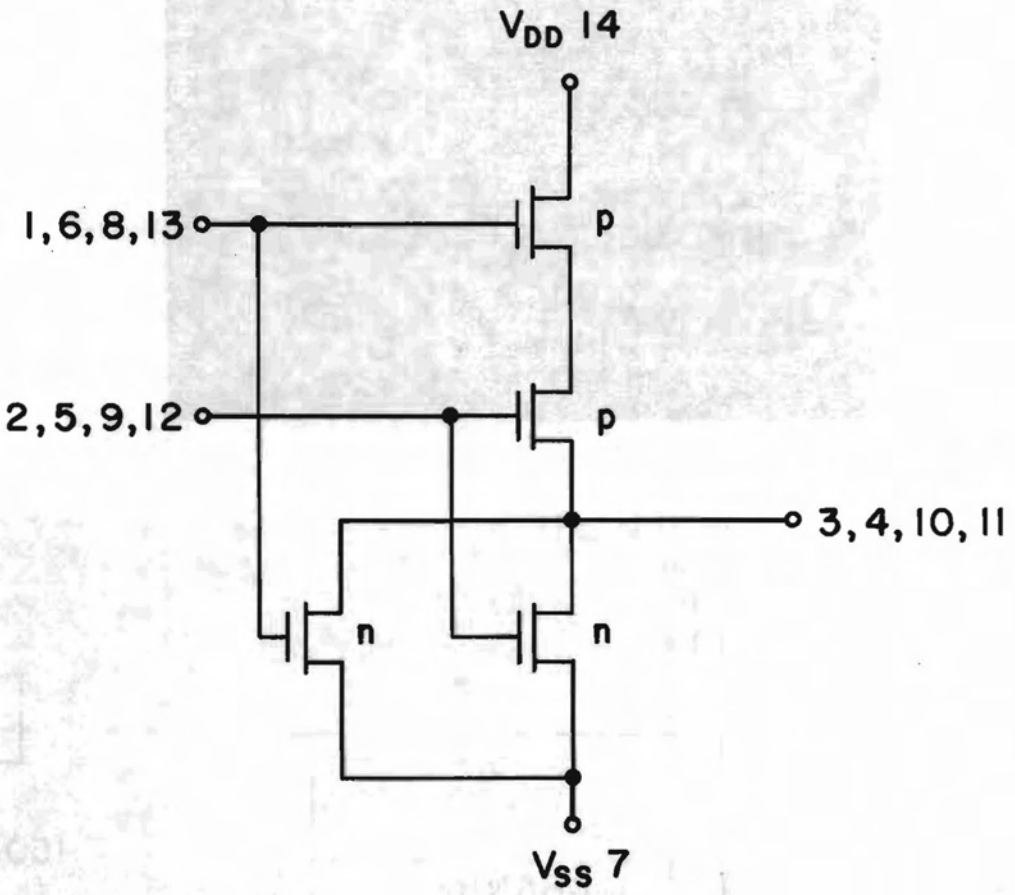
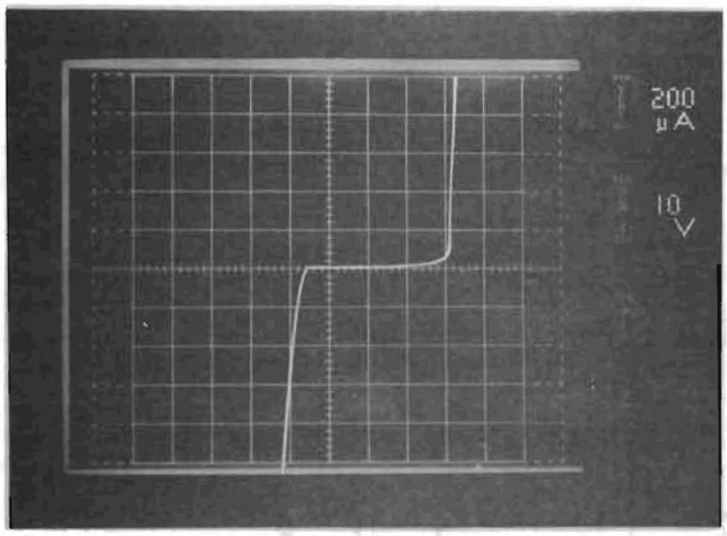


Figure V-16. Schematic diagram of an SOS quad 2 input NOR gate.



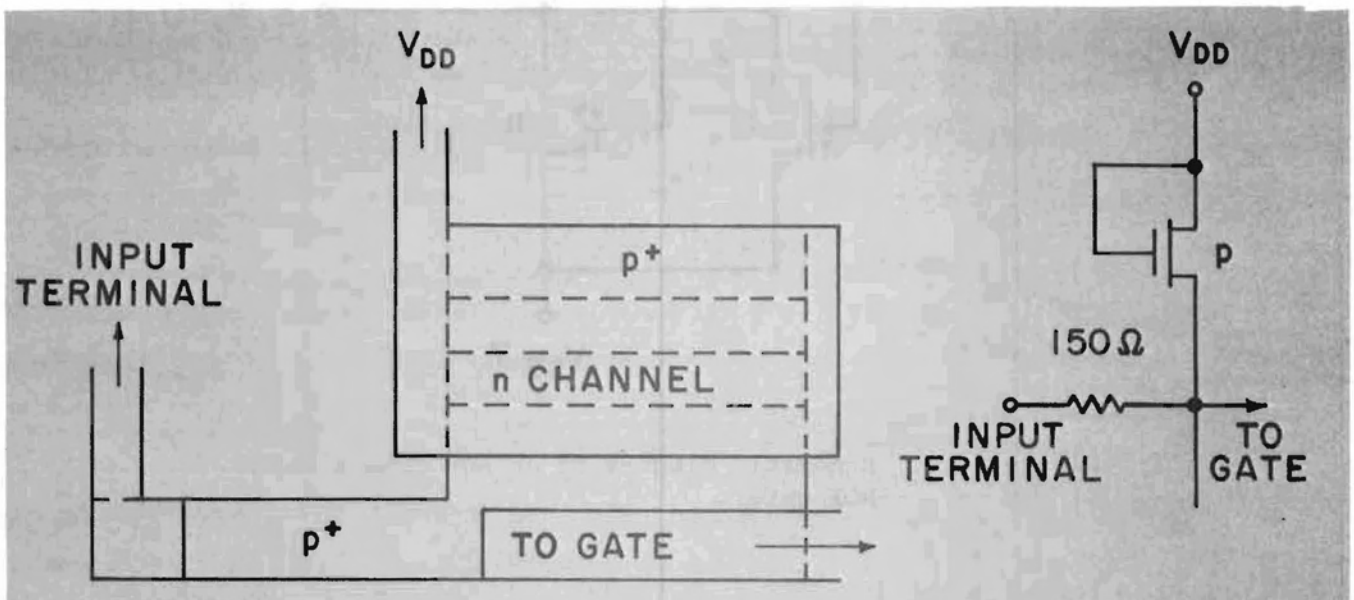


Figure V-17. Electrical characteristics and layout of gate protection device used in SOS NOR gate.

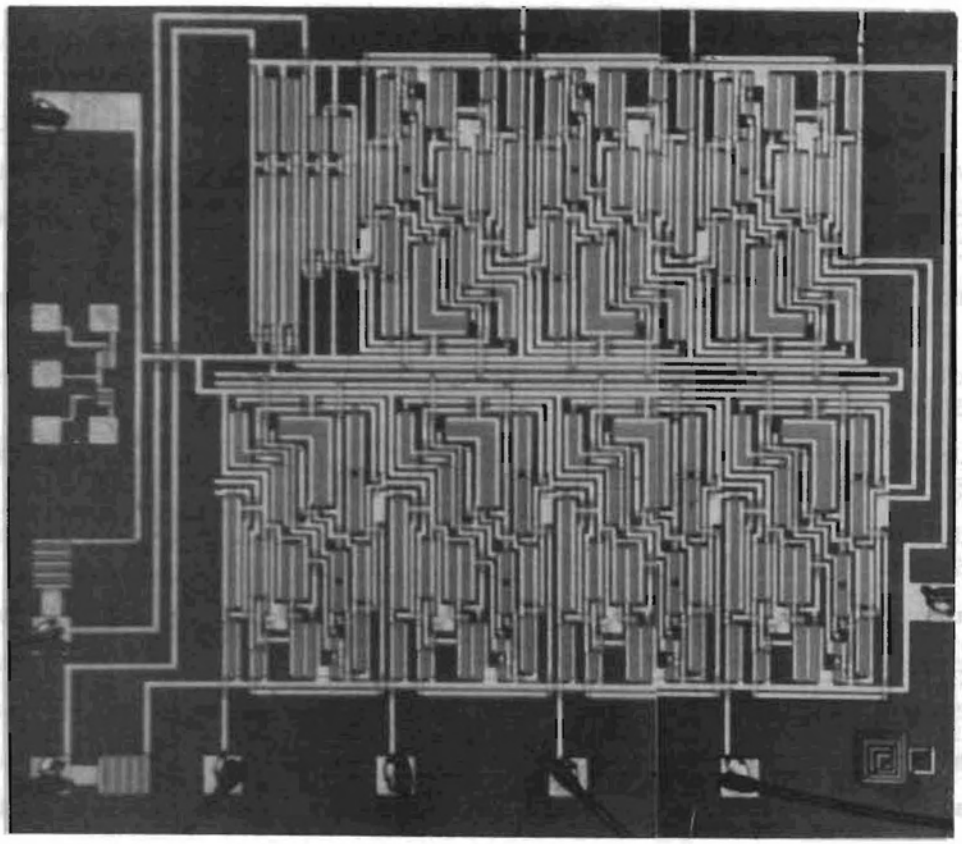
Table V-5. Typical Characteristics of an SOS and Bulk CMOS Quad 2-input NOR Gates From Two Manufacturers

	SOS	Bulk CMOS	Bulk CMOS
chip size	55x70	50x50	60x60 mils
quiescent power/package, V_{DD}=10	30	10	10 nW
transition time, V_{DD}=10, C_L=15 pF	20	35	20 ns
propagation delay, V_{DD}=10 C_L= 15 pF	10	25	25 ns
output drive current, V_{DD}=10, V_O=0.5	3.5	2.5	3.5 mA
output drive current, V_{DD}=10, V_O=9.5 V	2.2	1	2.3 mA
input capacitance	7	5	5 pF
dissipation/gate 1 kHz	-	3.5	- μW
dissipation/gate 10 MHz	-	35	- mW
n-channel width	27	7	- mils
p-channel width	54	16	- mils

The same conclusions are reached for this simple gate circuit. The SOS transistors are considerably larger to provide a higher output current drive, resulting in a larger chip. The high frequency response is improved, however.

D. 7-Stage Binary Counter

A 7-stage binary counter manufactured by Vendor E is shown in Fig. V-18. This circuit was made in several different versions, including a standard double-epi part, a metal gate ion implanted model, and a double-epi silicon gate version. A comparison among several circuits is given in Table V-6.



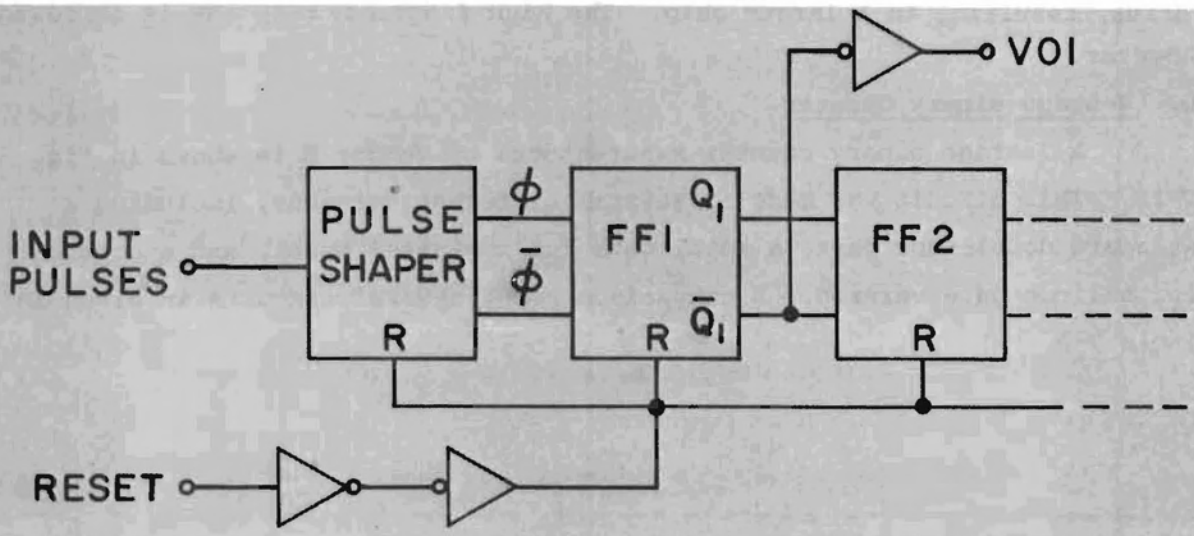


Figure V-18. Photograph and block diagram of an SOS 7-stage binary counter.

Table V-6. Characteristics and Performance of 7-Stage Binary Counters

	SOS, Manufacturer D	SOS, Manufacturer E	Bulk CMOS
chip size	-	90x100x10*	75x85x7 mils
quiescent power V_{DD}=10	10	200	5 μW
dissipation at 1 kHz, V_{DD}=10, C_L=15 pF	-	-	40 μW
dissipation at 10 MHz, V_{DD}=10, C_L=15 pF	-	40	40** mW
propagation delay, V_{DD}=10, (input clock to 1st output)	45	46	80 ns
maximum clock frequency, V_{DD}=10	16	20	7 MHz

* other types ranged from 120x120x10 to 70x80x10 mils

** extrapolated

E. Quad Bilateral Switch

This circuit, shown in Fig. V-19, is an example of an ion implanted silicon gate device. The chip measures 55x55x12 mils. The circuit was manufactured by Vendor F and represents a rather sophisticated example of recent SOS design and fabrication technique. Apart from the novel design features, the process appears to be typical. Starting material is 1 μm n-type silicon, which is ion implanted to invert the channel of the p-devices and adjust the threshold voltage. Unfortunately, electrical performance characteristics were not available at the time of this writing.

F. SOS Test Chip

A special test chip, made by still another manufacturer, Vendor G, is shown in Fig. V-20. The chip, 200x200x12 mils, is made by an aluminum gate, ion implant process. Individual components on the chip include, among others,

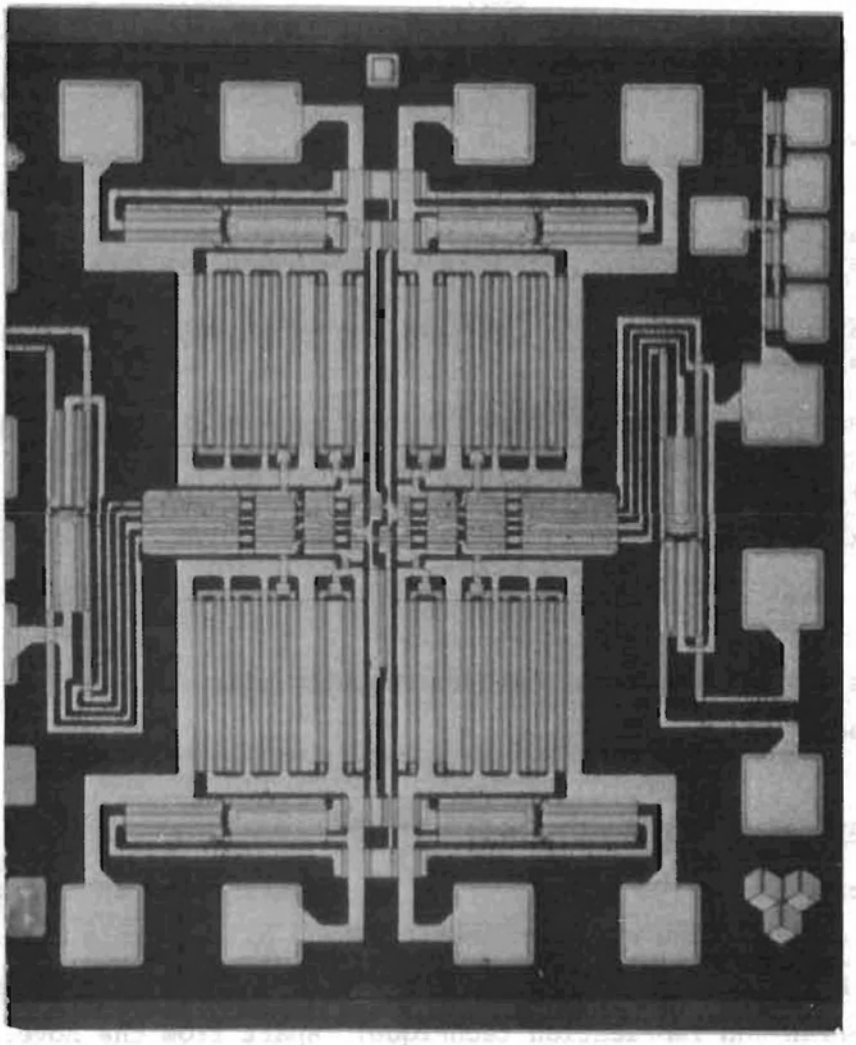


Figure V-19. Photograph of an SOS quad bilateral switch.

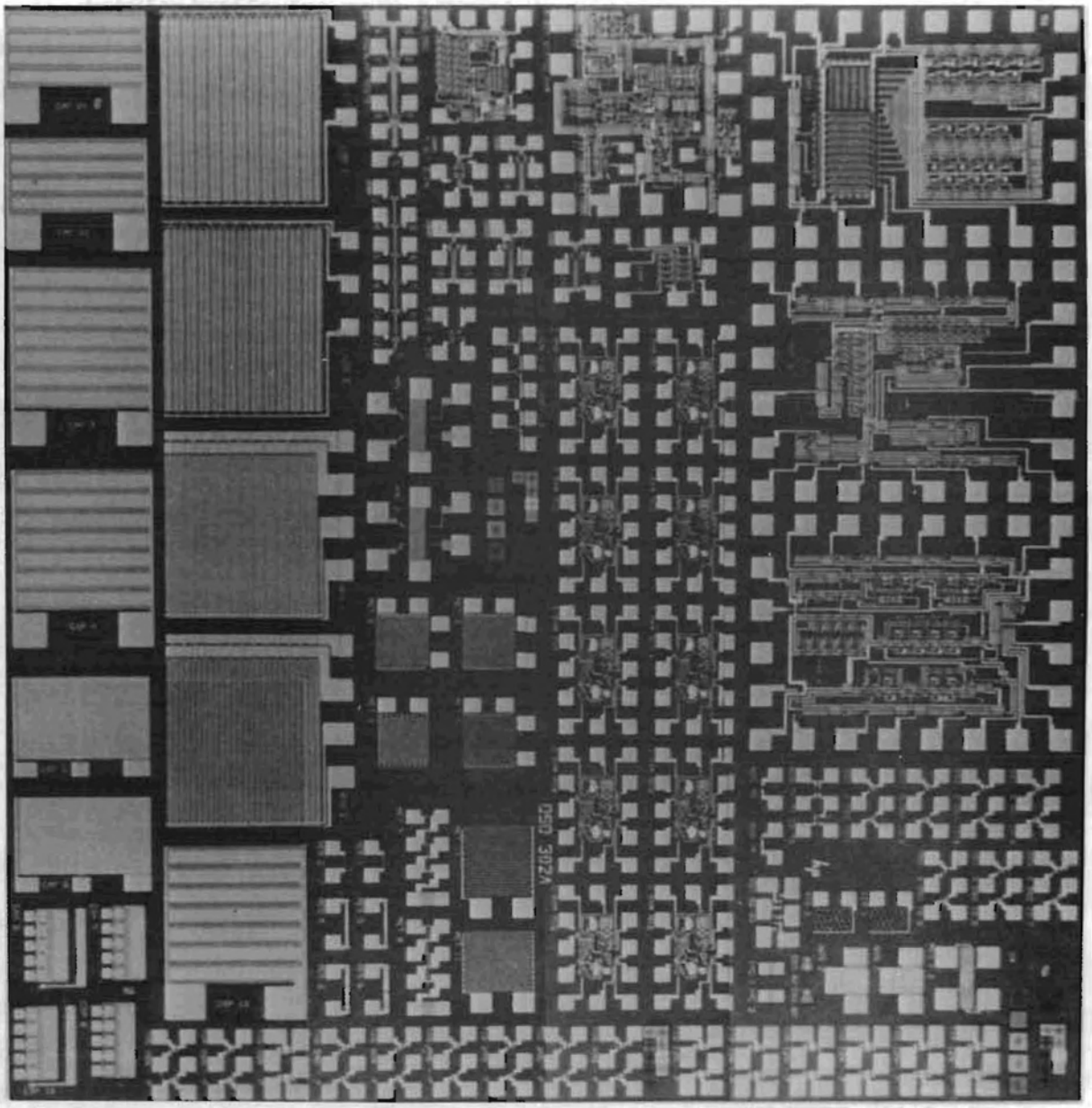


Figure V-20. Photograph of a complex SOS evaluation chip.

capacitors, resistors, diodes, transistors, Hall effect structures, metalization runs, etc. The small test transistors have channel widths varying from 4 to 16 μm, and channel lengths from 3 to 12 μm (mask opening). A variety of circuits, including static and dynamic rams are also included.

Minimum metalization width is 0.2 mils, and the smallest pads are 2x2 mils. These dimensions are about half those of standard integrated circuits.

G. Decade Counter, and 14-stage Binary Counter

Two other examples of MSI integrated circuits are shown in Fig. V-21. These circuits are compared to bulk CMOS versions in Tables V-7 and V-8.

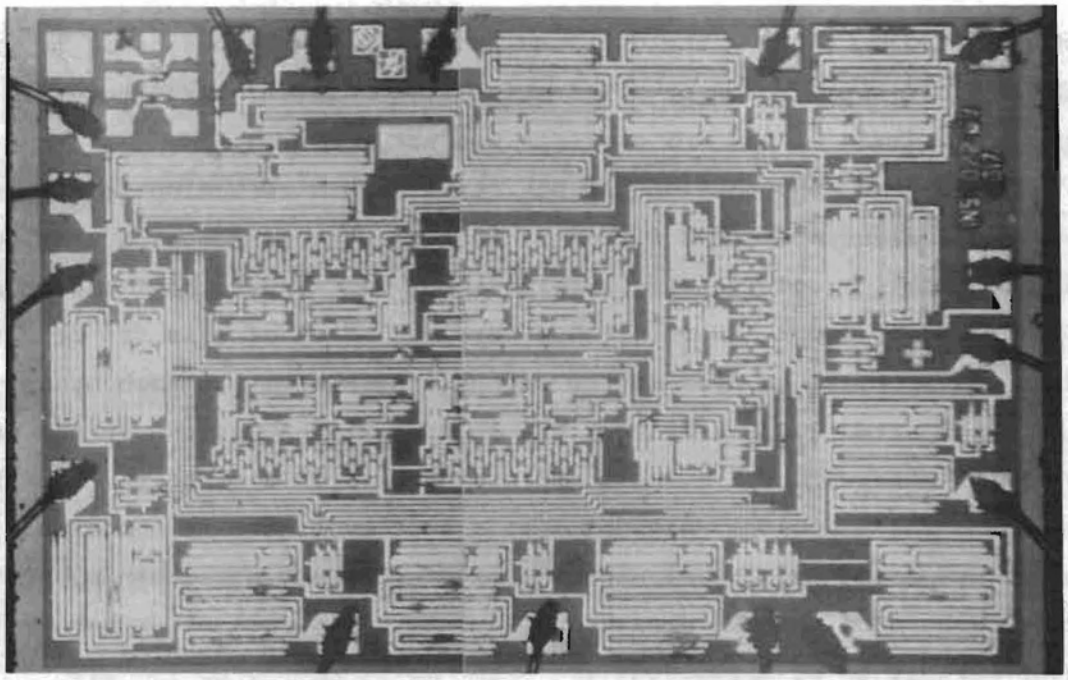
Table V-7. Characteristics of Decade Counters

	SOS	Bulk CMOS	Bulk CMOS
chip size	75x110	78x80	76x82 mils
quiescent dissipation at 10 volts	10	5	2.5 μW
dynamic dissipation at 10 volts, 1 MHz	-	7	- mW
propagation delay, CLOCK to CO	70	125	125 ns
max clock frequency	13	5	5 MHz

Table V-8. Characteristics of 14-stage Binary Counters

	SOS	Bulk CMOS	Bulk CMOS
chip size	98x103	71x83	85x87 mils
quiescent dissipation at 10 volts	2	10	10 μW
dynamic dissipation at 10 volts, 1 MHz	-	4	- mW
propagation delay, CLOCK to 1st output	80	150	130 ns
max clock frequency	16	7	8 MHz

For the small scale or medium scale integrated circuits discussed so far, the often stated advantages of SOS are not necessarily realized. The improvement in speed has been amply demonstrated. In each example the increase in maximum clock frequency or decrease in propagation delay has been



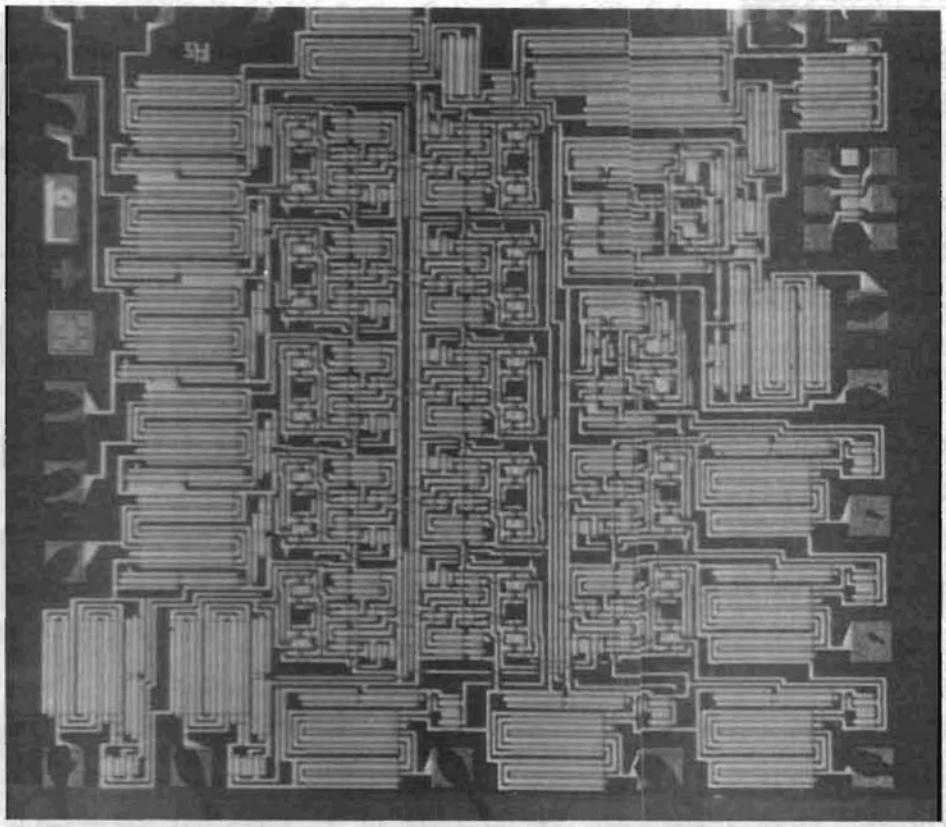


Figure V-21. Photographs of SOS counters.
Upper, decade counter. Lower, 14-stage counter.

a factor of two or three. The quiescent power, on the other hand, is greater for the SOS circuits, again by a factor of two or three. Unfortunately, the dynamic power drain of the SOS circuits was not measured.

The simpler construction has not resulted in a reduction in chip size over the equivalent bulk CMOS circuit. Even for the 14-stage counter, which represents a fairly complex circuit with about 375 transistors, the chip area increases by about 50% in the SOS version. Much of this can be attributed to the increase in area of the output transistors for each stage, to provide the desired current drive. The channel widths for the output drive pairs are 20 mils versus 5 mils in one of the bulk CMOS circuits. Apart from this aspect, the supposedly higher packing density of SOS is not simply an automatic by-product of the elimination of p-wells and guard bands, but must be achieved by good circuit design and layout. This will be emphasized in the next circuit to be discussed.

H. 256 Bit Static RAM

Figure V-22 is a photograph of a 256x1 bit CMOS SOS static random access memory. The circuit has on-chip decoding and is compatible with TTL, CMOS, and other SOS components. It is intended for use in scratch pad, buffer, and main memory applications where high speed and low power are required. Standby power dissipation is 40 μW, and operating dissipation is 30 mW. Typical access time (address to output delay) is 180 ns. Read cycle time is 180 ns and write cycle time is 140 ns. A block diagram of the circuit is shown in Fig. V-23.

The chip measures 120x160 mils. The memory cell array is about 75x75 mils, or 22 $mils^2$/cell. This is much larger than typical cell sizes in modern LSI memories, but is smaller than some memory cells in comparable 256 bit RAMS. The cell is a standard 6-transistor cell, shown in Fig. V-24.

In terms of speed, dissipation, and chip size, this circuit compares reasonably well with other CMOS 256 bit memories. It appears, then, that this level of integration is required before SOS technology shows any significant advantage over ordinary MOS or CMOS circuits. It should be realized though, that within any technology it is possible to enhance performance simply by design improvements. To illustrate this point, this SOS RAM was redesigned at the time of a changeover to an ion implanted aluminum

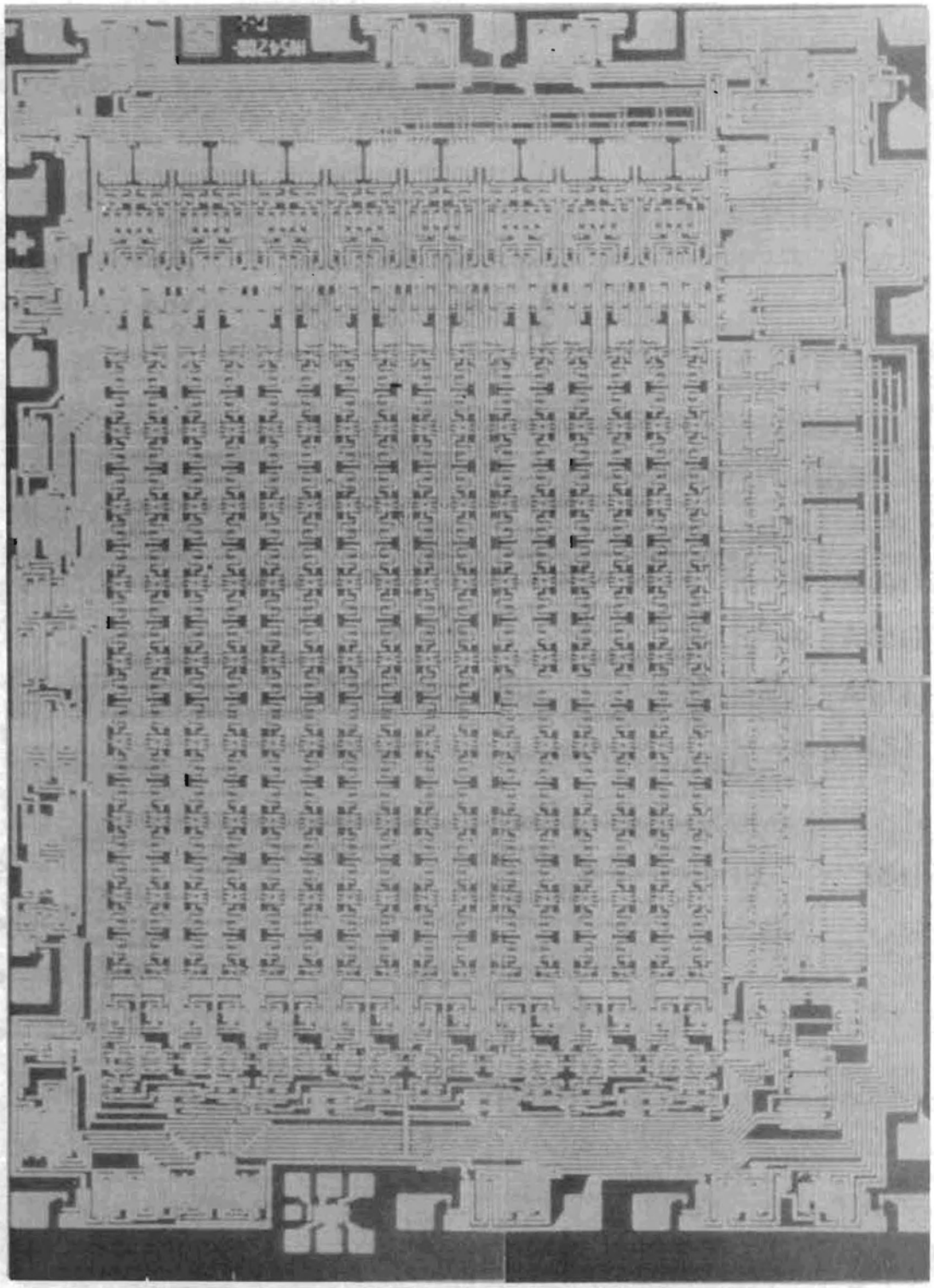


Figure V-22. Photomicrograph of an SOS 256-bit RAM.

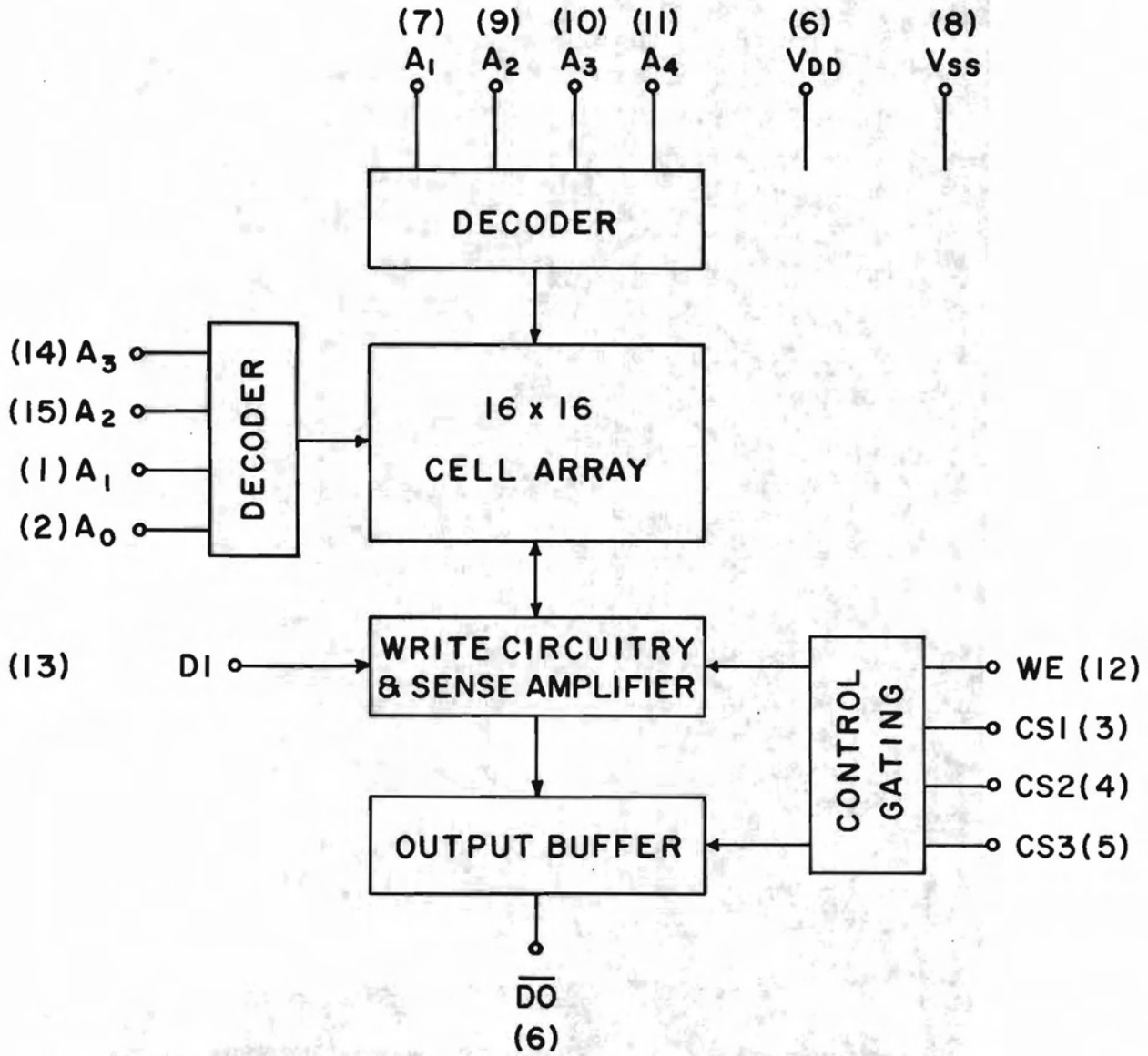
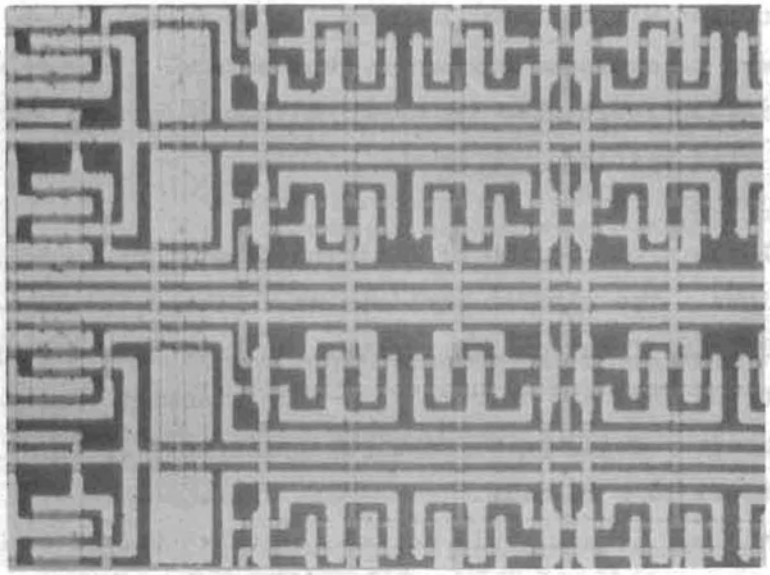


Figure V-23. Block diagram of a 256-bit SOS RAM.



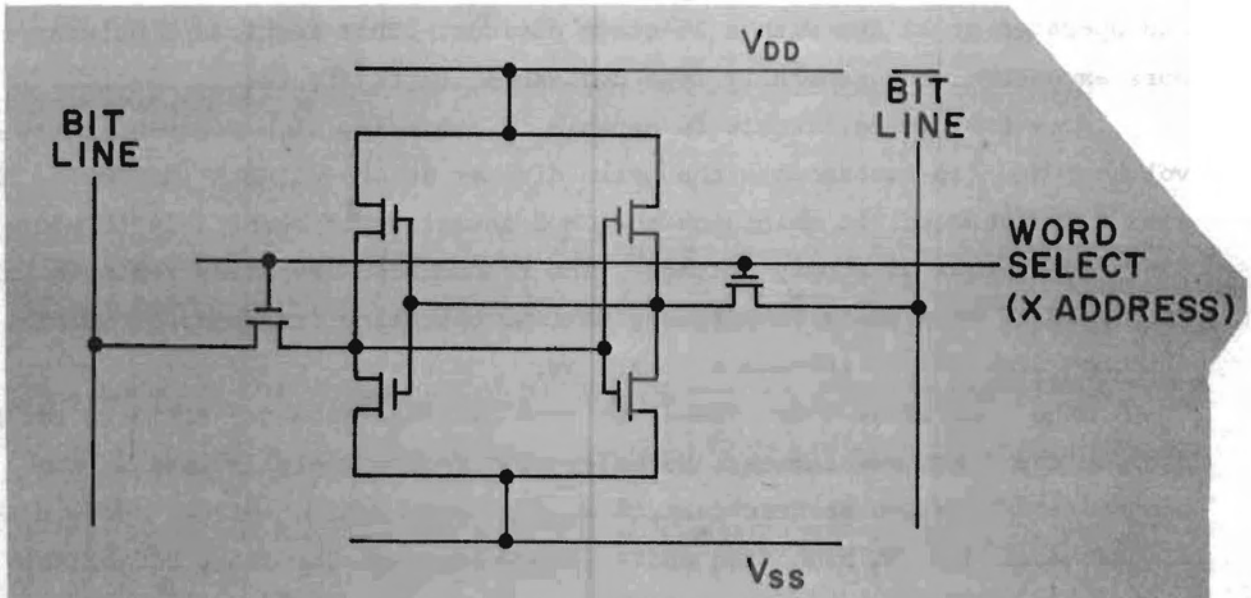


Figure V-24. Photograph and diagram of a memory cell in a 256-bit SOS RAM.

gate process. Although the circuit was functionally the same, a significant reduction in chip size was achieved.

The new version is shown in Fig. V-25. The smaller size is readily apparent by comparing the figure with Fig. V-22, both of which were taken at 50X. The new chip size is 110x110 mils, a reduction in area of nearly 40%. This reduction was achieved by both a reduced cell size, 13 mil^2/cell, and by a decrease in overhead circuitry real estate. Presumably the speed and power dissipation were also improved, although figures were not available at the time of this writing.

I. SOS Timing Circuit

Figure 26 is a photograph of a timing circuit designed for electronic watch applications. The circuit consists of a Pierce oscillator section for use with a 4.194 MHz AT-cut quartz crystal, an inverter, 7 flip-flop stages to count down to 32 kHz, and 16 more flip-flop stages to count down to 0.5 Hz for stepping motors. Total power dissipation is only 15 μW at V_{DD} = 1.5 volts. The chip is 76 x 88 mils.

SOS technology is admirably suited for this type of application. The low power drain is a necessity for 1 year battery operation and dictates the use of MOS circuitry. Previously used MOS circuits were limited in frequency and operated at 32 kHz with a 15-stage divider. This required a bulkier, more expensive, and generally less desirable quartz crystal.

The SOS timing circuit is capable of achieving high frequencies at low voltages and can accommodate the extra divider stages without increasing chip size. Furthermore, it maintains the CMOS advantage of being able to operate over wide ranges of supply voltage. The recommended operating range is 1.1 to 5 volts. At 5 volts the typical maximum operating frequency is 80 MHz, although dissipation increases to 1.6 mW.

The circuit is also unusual in that dynamic frequency dividers can be used in the first few stages. Normally only static logic is used in SOS because the high and erratic leakage current complicates refresh cycle considerations. At the high frequencies encountered at the front end of the timing circuit refresh is not a problem, and dynamic stages require fewer components.

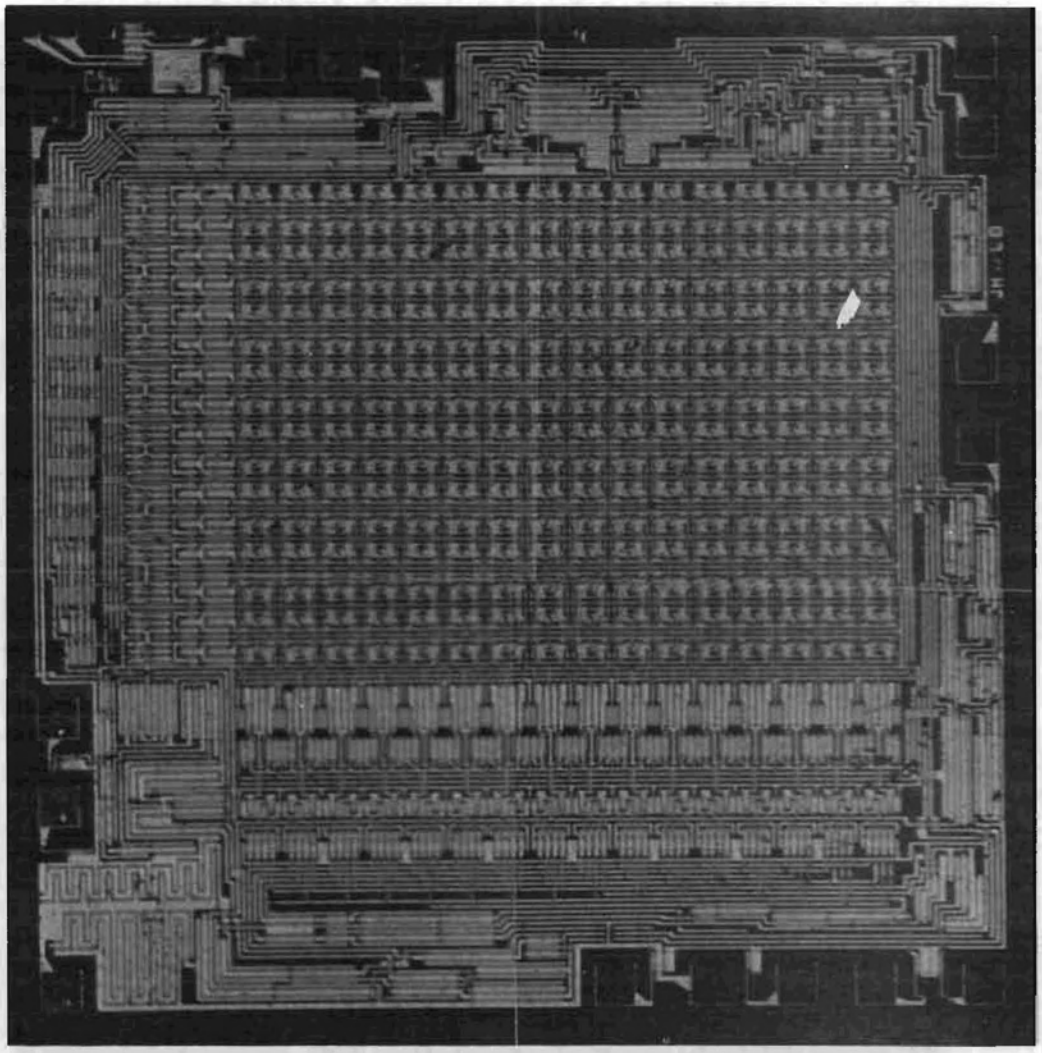


Figure V-25. Photograph of an improved SOS 256-bit RAM using a silicon gate, ion implant process.

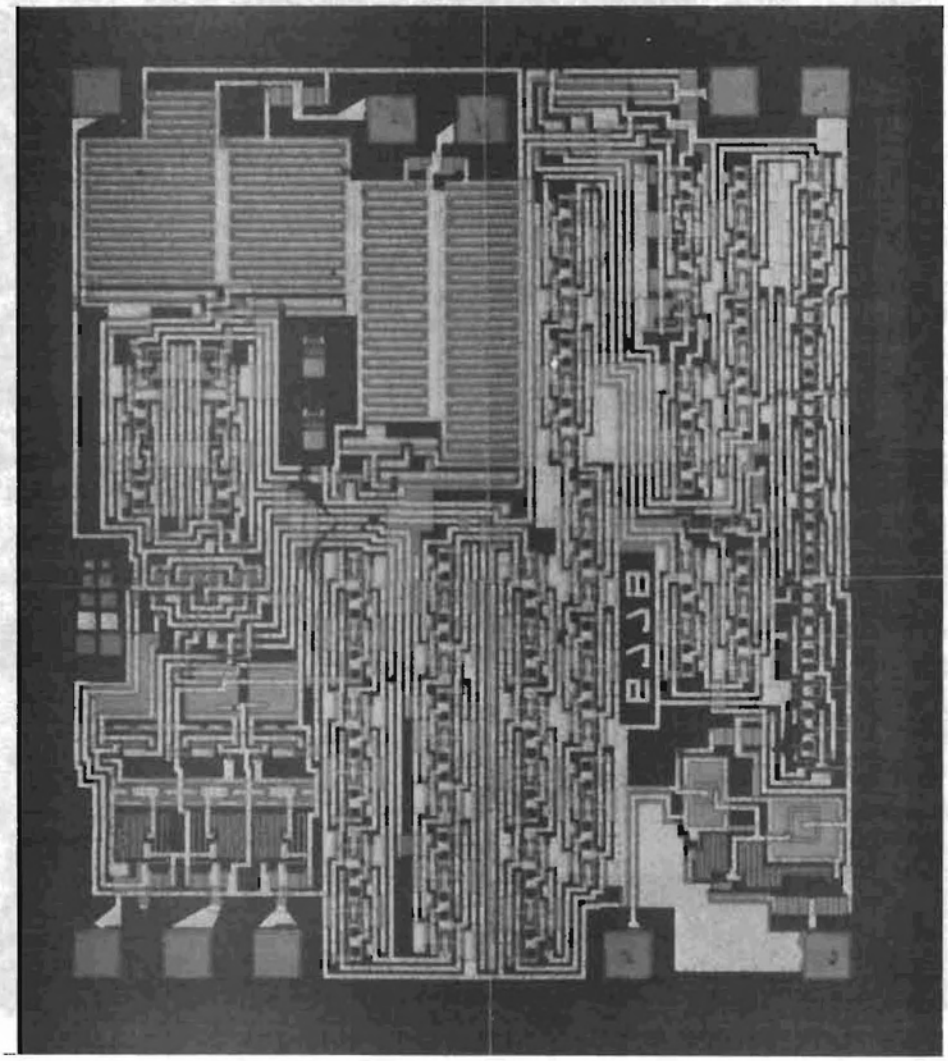


Figure V-26. Photomicrograph of an SOS timing chip.

SECTION VI

DISCUSSION

The purpose of this section is to summarize the status of silicon on sapphire technology and point out some areas where further study needs to be undertaken. Of course the technology is by no means a mature one, and considerable growth and evolution can be expected in the next few years.

A. SOS Processing

In Section III the different methods of fabricating SOS circuits were described in detail. In the summary of that chapter it was stated that the silicon gate, ion implant process will probably be accepted as the standard method. The technological advantages over the earlier double epi method can be summarized as follows.

1) Better frequency response. With self-aligned gates, the gate to source and gate to drain capacitances are substantially reduced. Since this is the major contribution to the inverter node shunt capacitance, reductions here will pay immediate dividends. Maximum frequencies can be doubled or tripled over comparable aluminum gate structures.
2) Greater packing density. Elimination of gate to source and drain overlap provides a small improvement in area. Additional economy results from the tighter design rules which can be permitted. Less polysilicon to metal spacing is required than metal to metal spacing. A more important advantage is the availability of another level of interconnection. A total of three layers is available; p^+ or n^+ silicon, polysilicon, and aluminum.
3) Better circuit characteristics. Ion implantation allows better control over threshold voltage absolute value and matching. Use of a p^+ polysilicon gate will reduce the threshold voltage to low enough levels to make low voltage applications feasible.
4) Processing advantages. With ion implantation only one epitaxial layer is needed. This is important because uniform, controllable epitaxial growth is still a problem in SOS technology. The elimination of one high temperature process is also an advantage. Finally, the critical gate oxide is grown early in the processing sequence and immediately covered with polysilicon. This eliminates possible

contamination from aluminum and affords a certain degree of protection during subsequent processing and handling.

Since most SOS circuits are complementary MOS, it is useful to compare processing in these two technologies.

SOS	Bulk CMOS
7 masks	7 masks
1 epitaxial deposition	
1 ion implant	
1 polysilicon deposition	
1 diffusion	3 diffusions
4 oxidations	3 oxidations
1 metalization	1 metalization
1 glassivation	1 glassivation

It is difficult to be precise in comparing the two processes because some companies perform a high temperature anneal after ion implant, some strip and regrow oxides after diffusions while others grow the oxide during drive-in, etc. On the basis of the steps listed here the two processes have about the same complexity. Some people have concluded that wafer processing costs are about the same in either technology provided that the facilities are producing the same number of wafers per week. Others have concluded that since there are no deep diffusions in SOS manufacturing the amount of throughput for a given production facility will be greater when processing SOS. On this basis SOS wafer processing costs would be less. More experience needs to be accumulated to verify this, however.

B. Yield and Area Considerations

Although good information on die sort yields is lacking, one would expect figures comparable to those for bulk CMOS since the processing steps are roughly the same. One manufacturer reports yields ranging from 1% for circuits 5×10^4 mil^2 in area to 20% for 1×10^4 mils^2 (53). These figures apply to circuits fabricated for a special test program and not to commercial devices. For commercial parts 1×10^4 mil^2, yields as high as 55% have been obtained.

Certainly the yield will increase as manufacturers improve their processes. SOS has a potentially higher yield than bulk CMOS because of two factors: (1) the absence of a field region, and (2) the smaller size.

Because silicon is removed over much of the surface of an SOS circuit the probability of a photo defect is substantially reduced. In the same way there are no silicon defects in the field region. Since the total area per gate or per cell can be smaller in SOS, this further reduces the probability of defects.

Comparing area reductions made possible by various processes, some manufacturers estimate that aluminum gate SOS circuits can be 30% smaller than bulk devices. Silicon gate structures effect even further reductions since some aluminum to aluminum spacing requirements can be reduced, silicon gate has shorter channel lengths, and conductor crossovers can be reduced. Ion implantation can further decrease the device active area. In LSI the packing density and chip area are ultimately limited by contact cuts, line spacing, line widths, etc., which are the same for any technology. Any improvement in basic processing steps such as photolithography will therefore affect all technologies.

Some of the problems still confronting SOS processing engineers are non-uniform epi material, poor control over silicon etching, and poor metal coverage over silicon steps. On the other hand, SOS processing can be faster than bulk CMOS because the diffusion schedules are shorter and because a thick field oxide is not necessarily required. If processing times are shorter then the probability of defects is reduced and yield is increased.

It appears that the potential yield of SOS circuits is greater than bulk CMOS, but this must still be demonstrated in volume production.

C. Reliability

SOS circuits have certain characteristics which point to possible reliability hazards. One of these concerns the high and variable leakage currents. As already pointed out this precludes the use of SOS circuits in dynamic logic. Another instability, previously discussed, is the edge effect associated with the gate conductor overlapping the silicon island edge. One of the contributions to the instability is thought to be caused by a preferential change in the fixed charge at the edge interface which causes a hysteresis effect when the edge current exceeds the channel current plus leakage current. This instability has been observed in all SOS devices except those with no edges and those to which no contact to the edge region

is allowed. A second instability occurs at high temperature and is associated with charge transport through the gate oxide. This problem is especially pronounced when aluminum is in contact with the gate oxide, and appears to be eliminated by a polysilicon gate.

Additional sources of reliability problems are some of the same difficulties which contribute to poor yields, namely, non-uniform epi material and poor step coverage of metalization. Metalization problems were encountered early in development, and are still recognized as a source of difficulty.

Gate dielectric breakdown was a serious problem in early MOS manufacture. To eliminate this hazard all circuits utilize some type of gate protective device, and are shipped in some form of special conductive package to protect against static charge. Not only should full gate protection be insisted upon, but protective networks should be connected between V_{DD} and V_{SS} to suppress power supply transients that could damage gates connected to the power supply busses.

A severe handicap to reliability estimation in SOS circuits is the lack of data on the functional operating life. What little information that is available now is not very encouraging. One report (53) lists 20 failures out of 88 devices with a total of 359,000 device hours. About half were functional failures, due to metalization opening up over silicon steps or due to gate dielectric breakdown. The other failures were due to excessive increases in leakage currents. These devices were not commercial parts, therefore these figures can be misleading. Other figures cite results of an operating life test for a commercial dual D-type flip-flop which yielded a failure rate of 0.04%/1000 hours at the 60% confidence level for 188,000 device hours (extrapolated to 25°C) and 0.02%/1000 failure rate for 138,000 device hours for life storage test (again extrapolated to 25°C) (54).

Far more reliability data must be assembled before reliability can be estimated with any assurance.

D. Radiation Effects

Radiation effects in silicon integrated circuit devices can be classified as displacement damage or electronic ionization effects. Displacement damage is mainly concerned with the creation of defects in the silicon which can act as scattering centers, recombination sites, and trapping levels and

with the release of bound contaminants in the silicon dioxide. Electronic ionization effects can create electron-hole pairs in silicon and silicon dioxide layers. The major effects on MOS circuits of displacement damage are the decrease in life-time due to recombination centers, the reduction in carriers due to trapping, and the release of sodium and other ions in silicon dioxide. Electronic ionization generates conductance modulation effects in bulk silicon which may be significant in SOS devices, photocurrents in space charge regions, and carrier trapping in silicon dioxide, at the silicon dioxide-silicon interface, and at the silicon-sapphire interface. Electronic ionization may create a build-up of positive charge in the oxide near the silicon dioxide-silicon interface which can cause gate threshold voltage shift.

It has been generally assumed that SOS circuits have superior radiation resistant properties as compared to similar bulk devices. The reason for this assumption is that SOS devices are constructed in approximately 1μm silicon layers on insulating substrates. Most high energy radiation will pass through the silicon film and terminate harmlessly in the sapphire substrate, and since the field region silicon is removed, the radiation-induced field inversion and photocurrents should be eliminated. The initial data which is limited and unpublished seem to support these assumptions. It has also been assumed that SOS devices would have similar gate threshold problems as bulk MOS devices. The current state of radiation hardening for gate threshold stability in CMOS includes the use of aluminum implants of 10^{15} ions/cm^2 at 10 kw in the n-channel devices, chrome doped oxides, boron implanted oxides, and Al_2O_3 gate dielectrics. It is reasonable to assume that these techniques will be equally effective on SOS.

One report (55) has investigated the radiation resistance of n-channel and p-channel devices on sapphire and bulk silicon from several manufacturers. From the limited test sample, three generalizations were made. First, as expected, the SOS devices are inherently more resistant to photocurrent effects than are their counterparts in bulk silicon. Also, using gate threshold measurements as a function of time after radiation, the change in gate threshold voltage was sometimes much greater initially and did not recover as well as in bulk devices. Finally, a few SOS n-channel devices

exhibited high leakage currents after irradiation, which is thought to be due to silicon-sapphire interface changes or to threshold lowering due to charging of the sapphire which acts as a second gate.

The current status of radiation hardening of SOS devices is unclear, but there are indications that more study and development work will be necessary to achieve the initial SOS claims. In the important area of gate threshold voltage control, it may be that SOS devices are somewhat inferior to bulk devices. How SOS devices will respond to current gate threshold hardening techniques is not yet known.

E. Comparison With Other Integrated Circuit Families

To more fully appreciate the applications where SOS circuitry has an advantage over existing technologies it is helpful to list the performance characteristics of commonly used digital integrated circuit families at three levels of complexity: logic gates (such as NAND gates, NOR gates, inverters, etc.); medium scale integrated circuits (such as shift registers, ripple counters, adders, etc.); and large scale integrated circuits such as large RAMS.

In Table VI-1 the gate dissipation and propagation delay of several logic families are listed. Of course, the table entries represent only average values and vary among manufacturers. Performance also depends on such variables as power supply voltage, fan-in, fan-out, etc. Nevertheless, most of the listed values are taken from readily available compilations. No attempt is made to compare gate areas (typically 50 x 50 mils) since this has little significance for small logic arrays where so much of the chip area is devoted to bonding pads, scribe lanes, etc. Furthermore, most SSI circuits represent an early stage of development when packing density was not a prime consideration.

The table shows that SOS is slightly faster than bulk CMOS but slower than standard TTL. Static dissipation is greater than bulk CMOS but far less than bipolar logic. Dynamic dissipation is less than all logic families (including bulk CMOS) up to about 1 MHz. Of course, only two characteristics are being compared in this table. Many other features, such as power supply requirements, off-chip drive, noise immunity, and fan-out must be considered for a given application. In all of these, except off-chip drive, CMOS (both bulk and SOS) is outstanding.

Table IV-1. Performance Characteristics of Logic Families

Logic Family	DTL	Low Power TTL	Standard TTL	Bulk CMOS	SOS	High Speed TTL	Schottky TTL	ECL
Gate Dissipation, mW	12	1	10	10^{-5} static[1]	2×10^{-5} static[2]	22	20	55
Propagation Delay, ns	30	30	10	25	20	6	3	1

[1] 30 mW at 10 MHz

[2] 20 mW at 10 MHz

Table VI-2. Characteristics of MSI Integrated Circuit Families

Logic Family	PMOS	Bulk CMOS	TTL	SOS	ECL
Dissipation, mW	20	.01 static[1]	100	.05 static[2]	700
Maximum Frequency, MHz	2	5	30	50	500

[1] 5 mW at 1 MHz

[2] 5 mW at 1 MHz

Shift registers, counters, and adders represent another level of integrated circuit complexity. Chip sizes are about 80 x 80 mils, with perhaps as many as 300 devices per chip. Representative values for dissipation and maximum frequency are listed in Table VI-2.

Figures listed in the table are only rough estimates, but serve to point out that SOS is capable of excellent frequency response at this level of complexity. Only ECL surpasses SOS in this respect. It should be noted that SOS circuits are new on the market, and hence are able to take advantage of the most recent process improvements, design rules, etc. As pointed out earlier in this report, a redesign of an older product line would result in improved performance also. In power dissipation both SOS and bulk CMOS enjoy miniscule static losses, and dissipate less than standard TTL up to about 20 MHz. As far as chip sizes are concerned, these appear to be larger in SOS versions at the present time.

Large scale integrated circuits, specifically 1024 bit memories, are difficult to compare because of widely varying requirements - some are designed for high speed operation whereas others, using the same technology, fulfill low power requirements. Generally speaking, bipolar RAMS of this size feature access times of 50 ns or lower and dissipation of 600 mW or so. PMOS RAMS are the slowest, with access times of 100-300 ns. Power dissipation is 2-70 mW standby, 200-400 mW operating. NMOS is faster, 60-150 ns, with 1-10 mW standby power and 100-500 mW operating. There are few bulk CMOS memories on the market, and these are relatively slow. SOS memories which have been announced have access times of about 100 ns with unusually low dynamic power dissipation, 20 mW.

One of the difficulties encountered when a comparison is attempted is the continued advances in the state-of-the-art. No sooner is a new development announced than someone makes design improvements on an older part to improve its performance. At every large conference one is assailed with new methods of fabricating bipolar circuits with less power dissipation, making MOS devices which are faster, reducing cell size, or eliminating one or more masks. This fast-paced evolution attests to the virility of the industry and the creativity of designers, even as it makes technical evaluations somewhat tenuous.

It appears that silicon on sapphire will be useful in certain MSI and LSI applications, especially those where ordinary CMOS is being used now. Even single channel MOS, though less vulnerable, may eventually surrender much of its market to SOS incursions. Certainly SOS demonstrates low stand-by and operating power, and sufficient speed even now to surpass the fastest MOS circuits and even some bipolar circuits.

SOS will not replace dynamic memories in the foreseeable future because of the problem with leakage and refresh. Also there is little reason to use SOS for small scale logic - it is not as fast as bipolar logic, has poor drive capability, and cannot capitalize on low power dissipation in SSI (except where battery drain is important).

SOS can be used to advantage in static memories, watch circuits, microprocessors, calculators, and to a lesser extent in linear, consumer, and specialty applications. The potential markets are there; whether or not SOS will fill them depends on whether costs can be reduced, reliability improved and on such unpredictables as the health of the economy, who among industry leaders commits themselves to SOS, the appeal of the new technology to potential customers, etc. One of the unknowns is the development of new technologies, such as integrated injection logic or merged transistor logic* before SOS becomes firmly established.

F. Market Outlook

The success or failure of SOS technology is invariably connected with the health of the integrated circuit industry as a whole. During the 1973-74 parts shortage SOS circuits were sold as fast as they could be produced. Their higher price tag was no impediment to parts-starved buyers. In 1974 the commercial sales of SOS circuits amounted to 2-3 million dollars. When the economic slow-down occurred midway through 1974 SOS circuits were the first to be eliminated from the buyers' lists because of high prices and unproven reliability. SOS price projections for late 1974 and 1975 were competitive, especially for new ion implanted LSI products until severe price cutting affected the CMOS market. Whereas the average price of a CMOS gate in early 1974 was well above 50 cents, it declined to less than 20 cents in

*This is the topic of another report in preparation.

early 1975. This unexpected and severe price reduction, coupled with the plight of the economy in general, cash flow problems, production problems, and the financial distress of many electronic component and equipment manufacturers has clouded the issue. Any company which expects to be successful in SOS must have good technical know-how in the MOS business plus a sound financial position.

Growth sectors of the integrated circuits industry can be most readily identified by the WEMA industry figures and forecasts listed in Table VI-3. The figures show that the most rapidly growing segment of the industry is NMOS, followed closely by CMOS. If we assume that CMOS applications and some single channel MOS markets are vulnerable to SOS penetration, the outlook for SOS growth seems very good indeed. In fact, some industry sources have predicted an SOS market of 200 million dollars by 1979. Other sources have disagreed - claiming that this figure is too conservative!

Table VI-3. Free World Markets for Integrated Circuits in Millions of Dollars

	1973	1974	1975	1976	1977
DTL	120	115	105	95	80
TTL	560	700	725	850	995
ECL	65	75	95	135	180
Other Bipolar	70	70	65	60	55
Linear	415	480	395	615	760
PMOS	460	520	440	595	650
NMOS	13	90	110	234	417
CMOS	55	115	160	240	320

The very nature of the market poses an interesting dilemma. Companies who are most likely to succeed in SOS are those who have expertise and high volume sales in MOS, yet these same markets are those most likely to be taken over by SOS.

Some major manufacturers have already announced that they do not intend to become suppliers of SOS products, citing high costs and alternative technologies. A cautious attitude on the part of the industry as a whole could mean that SOS would never become firmly established. The longer the delays in winning a sizeable share of the market, the less the chances become, since state-of-the-art improvements in existing families and development of new ones tend to reduce or eliminate the performance advantage that SOS has. Furthermore, until high volume production gets underway, unit costs will remain high. The picture is thus the classic chicken-or-the-egg situation.

Most SOS manufacturers feel that significant growth will occur in late 1975, especially in watch circuit applications and static RAMS. The next year will probably be crucial in setting the pace for this technology.

G. Availability of Parts

Although more than fifteen companies possess SOS manufacturing capability (56), not all of them are currently fabricating SOS circuits. Furthermore, some have only research and development programs, some produce parts under contract only, and others produce circuits strictly for in-house consumption. Only two companies have commercial off-the-shelf products. Among the parts presently available are:

MSI logic circuits (counters, scalars, etc.)
bilateral switches
1024 bit RAMS (in several configurations)
timing circuits for watch applications

During 1976 it is expected that microprocessors and 4K static RAMS will become available.

Customer interest in SOS products has increased during the last half of 1975 to the point where sales of 1K RAMS are production limited at present. This comes at a time when suppliers are converting from 2" to 3" production lines. Increasing interest is also apparent in circuits for communications applications such as phase-locked loops, multipliers, FFT analyzers, etc. There are no definite plans to develop such products in the immediate future, however.

To reiterate earlier statements, it appears that production and demand are building up together, with each more or less dependent on the other.

H. Summary and Recommendations

Silicon on sapphire is a technology with distinct performance advantages. From the user's point of view the demonstrated advantages when used in digital, large scale integrated circuit applications are

- high speed
- very low standby power
- low operating power
- high noise immunity
- wide range of power supply voltages
- complete isolation between devices

Potential advantages, still to be demonstrated, are

- low cost
- high reliability
- improved radiation resistance

Lower cost, compared to bulk CMOS, can only come about when volume production is achieved - as a result of faster processing, high packing density, high yield, and lower substrate prices. Better reliability appears likely but adequate statistics are not yet available. High tolerance to transient radiation effects exists now, but steady state radiation hardness will probably not be much better than bulk CMOS.

The main drawback to SOS is the limited availability of circuits due in part to high substrate prices, need for special production know-how, and industry-wide lack of commitment. In addition the problems with high leakage, kink effects, and other instabilities have not been completely solved.

In order to achieve improvements in user-oriented areas such as performance, price, reliability, and availability of parts, and to take full advantage of the technology the following recommendations are made.

1. Further work should be done to improve wafer manufacturing methods with a view toward improving quality and lowering costs.
2. Better methods of growing heteroepitaxial silicon on sapphire must be devised. Problem areas are
   a) growing films of uniform thickness
   b) growing films of uniform electrical characteristics
   c) improving the electrical characteristics such as

i) better lifetime

ii) better mobility

iii) lower leakage currents

iv) better control of kink effects and instabilities.

3. A better understanding of reliability problems must be obtained. Therefore programs should be initiated to

   a) identify failure modes and mechanisms inherent to SOS

   b) obtain good statistical data for reliability prediction.

4. SOS is recommended for procurement in LSI digital applications where its capabilities can make a distinct improvement in system performance. Otherwide it is not recommended simply as a substitute for other integrated circuits because of limited second-source suppliers.

5. Parts should be procured from suppliers with a well-established manufacturing process, a stable financial situation, and a commercial product line to ensure reliability and availability over a long equipment service life.

REFERENCES

1. H. M. Manasevit, "A Survey of the Heteroepitaxial Growth of Semiconductor Films on Insulating Substrates," J. Cryst. Growth, 22, 125-148, 1974.

2. H. M. Manasevit and W. I. Simpson, Proc. Am. Phys. Soc., Edmonton, Alberta, Canada, 1963.

3. H. M. Manasevit and W. I. Simpson, "Single Crystal SOS Substrates," J. App. Phys., 35, 1349-1351, April, 1964.

4. P. H. Robinson and C. W. Mueller, "The Deposition of Silicon Upon Sapphire Substrates," Trans. Met. Soc. AIME, 236, 268-274, March, 1966.

5. J. Mercier, "Physical and Electrical Investigations on Silicon Epitaxial Layers on Sapphire Substrates," J. Electrochem. Soc., 117, 666-671, May, 1970.

6. D. J. Dumin, "Electrical Properties of Silicon Films Grown Epitaxially on Sapphire," J. App. Phys., 38, 1909-1914, 1967.

7. G. W. Cullen, G. E. Gottlieb, and C. C. Wang, "The Epitaxial Growth of Silicon on Sapphire and Spinel Substrates: Suppression of Changes in the Film Properties During Device Processing," RCA Review, 31, 355-371, June, 1970.

8. G. W. Cullen, "The Preparation and Properties of Chemically Vapor Deposited Silicon on Sapphire and Spinel," J. Cryst. Growth, 9, 107-125, 1971.

9. G. W. Cullen and J. F. Corboy, "A Comparison of the Semiconducting Properties of Thin Films of Silicon on Sapphire and Spinel," J. Electrochem. Soc., 121, 1345-1350, October, 1974.

10. P. Rai-Choudhury, D. K. Schroder, T. M. S. Heng, and G. G. Sweeney, "Impurity Distribution and Electrical Properties of Silicon on Sapphire Films," Electrochem. Soc., 1974 Fall Meeting, Extended Abstracts, 339-341, October, 1974.

11. F. P. Heiman and P. H. Robinson, "Silicon-On-Sapphire Epitaxial Bipolar Transistors," Solid State El., 11, 411-418, 1968.

12. R. S. Ronen and P. H. Robinson, "Recent Advances in Thin-Film Silicon Devices on Sapphire Substrates," Proc. IEEE, 59, 1506-1510, October, 1971.

13. C. W. Mueller and P. H. Robinson, "Grown-Film Silicon Transistors on Sapphire," Proc. IEEE, 52, 1487-1490, December, 1964.

14. E. J. Boleky, "The Performance of Complementary MOS Transistors on Insulating Substrates," RCA Review, 31, 372-395, June, 1970. See also J. R. Burns, et.al., "Advanced CMOS Circuit Techniques," Interim Report, WPAFB, Contract No. F33615-69-C-1499, May, 1970.

15. A. K. Rapp, "SOS: New Design Tools and Techniques," 1973 Int. Solid State Circuits Con., 170-171, February, 1973.

16. R. A. Kjar and G. Kinoshita, "Transient Photocurrents in SOS Structures," IEEE Trans. Nuc. Sci., NS-20, 315-318, December, 1973.

17. A. Miller and H. Manasevit, J. Vac. Sci. Tech., 3, 68, 1966.

18. J. D. Filby and S. Nielsen, "Single Crystal Films of Silicon on Insulators," Brit. J. App. Phys., 18, 1357-1382, September, 1967.

19. T. S. La Chapelle, A. Miller, and F. L. Morritz, Progress in Solid State Chemistry, Vol. 3, H. Reiss, ed., Pergamon Press, London, 1967.

20. C. C. Wang, G. E. Gottlieb, G. W. Cullen, S. H. McFarlane, III, and K. H. Zaininger, "Heteroepitaxy of Silicon on Stoichiometric Spinel," Trans. Met. Soc. AIME, 245, 441-450, March, 1969.

21. G. W. Cullen and C. C. Wang, "A Comparison of the Hole Mobility and Early Growth of Epitaxial Silicon on Flame Fusion, Flux, and Czchrolski Spinel," J. Electrochem. Soc., 118, 640-644, 1971.

22. P. H. Robinson and D. J. Dumin, "The Deposition of Silicon on Single Crystal Spinel Substrates," J. Electrochem. Soc., 115, 75-78, January, 1968.

23. W. H. Bragg and W. L. Bragg, X-rays and Crystal Structure, Bell & Sons, London, 1916.

24. L. Pauling and S. B. Hendricks, "The Crystal Structures of Hematite and Corundum," J. Amer. Chem. Soc., 47, 781-790, 1925.

25. R. E. Newnham and Y. M. DeHaan, "Refinement of the alpa Al_2O_3, Ti_2O_3, V_2O_3, and Cr_2O_3 Structures," Z. Krist., 117, 235-237, 1962.

26. J. E. A. Maurits, "Fabrication of Sapphire Substrates for Heteroepitaxy," Electrochem. Soc., 1974 Fall Meeting Extended Abstracts, 319-320, 1974.

27. P. F. Becker, "Surface Hardening of Sapphire and Rutile Associated with Machining," J. Am. Cer. Soc., 57, 2, 107-8, 1974.

28. Y. S. Chiang and G. W. Looney, "The Epitaxial Growth and Properties of Silicon on Stoichiometric Spinel and Sapphire Substrates Using Silane-Helium Mixtures vs. Silane-Hydrogen Mixtures," J. Electrochem. Soc., 120, 550-553, 1973.

29. P. H. Robinson, and R. O. Wance, "The Chemical Polishing of Sapphire and Spinel," RCA Review, 34, 616-629, 1973.

30. R. W. Bicknell, B. A. Joyce, J. H. Neave, and G. V. Smith, "The Epitaxy of Silicon on Alumina - Structural Effects," Phil. Mag., 14, 31-46, July, 1966.

31. J. Mercier, "Properties of Low-Temperature Heteroepitaxial Thin Silicon Films Grown in Helium Atmosphere," J. Electrochem. Soc., 118, 962-966, June, 1971.

32. C. C. Chang, "Silicon-on-Sapphire Epitaxy by Vacuum Sublimation: LEED-Auger Studies and Electronic Properties of Films," J. Vacuum Sci. and Tech., 8, 500-511, May-June, 1971.

33. C. Weissmantel, O. Fieldler, G. Hecht, and G. Reusse, "Ion Beam Sputtering and its Application for the Deposition of Semiconducting Films," 13, B59, 366, Nov. 15, 1972.

34. D. J. Dumin and P. H. Robinson, "Carrier Transport in Thin Silicon Films," J. App. Phys., 39, 2759-2765, 1968.

35. E. C. Ross and G. Warfield, "Effects of Oxidation on Electrical Characteristics of Silicon-on-Sapphire Films," J. App. Phys., 40, 2339-2344, 1969.

36. W. E. Ham, "Surface Charge Effects on the Resistivity and Hall Coefficient of Thin Silicon-on-Sapphire Films," App. Phys. Lett., 21, 440-443, Nov., 1972.

37. R. J. Kriegler, "Neutralization of Na^+ Ions in 'HCl-Grown' SiO_2," App. Phys. Lett., 20, 449-451, June, 1972.

38. R. S. Ronen and P. H. Robinson, "Hydrogen Chloride and Chlorine Gathering: An Effective Technique for Improving Performance of Silicon Devices," J. Electrochem. Soc., 119, 747-752, June, 1972.

39. R. J. Kriegler, Y. C. Cheng and D. R. Colton, "The Effect of HCl and Cl_2 on the Thermal Oxidation of Silicon," J. Electrochem. Soc., 119, 338-392, March, 1972.

40. D. J. Dumin and R. S. Silver, "Diffused Diodes in Silicon-on-Sapphire," Solid State Elect., 11, 353-363, March, 1968.

41. P. W. Cook, D. L. Critchlaw, and L. M. Terman, "Comparison of MOSFET Logic Circuits," IEEE J. Solid State Circuits, SC-8, 348-355, Oct., 1973.

42. D. J. Dumin, "The Current Status of Silicon-On-Sapphire," Electrochem. Soc. 1974 Fall Meeting, Extended Abstracts, 321-322, October, 1974.

43. M. Pomper and J. Tihanyi, "Ion Implanted ESFI MOS Devices with Short Switching Times," IEEE J. Solid State Circuits, SC-9, 250-256, Oct., 1974.

44. S. R. Hofstein, "An Analysis of Deep-Depletion Thin-Film Transistors," IEEE Trans. Electron Devices, ED-13, 846-855, Dec., 1966.

45. W. N. Carr and J. P. Mize, MOS/LSI Design and Application, McGraw-Hill Book Company, New York, 1972.

46. H. C. Lin, J. L. Halson, and P. J. Hayes, "Shielded Silicon Gate Complementary MOS Integrated Circuits," IEEE Trans. Electron Devices, ED-19, 1199-1207, Nov., 1972.

47. E. J. Boleky and J. E. Meyer, "High-Performance Low-Power CMOS Memories Using Silicon-on-Sapphire Technology," IEEE J. Solid State Circuits, SC-7, 135-145, April, 1972.

48. C. T. Sah, R. N. Noyce, and W. Shockley, "Carrier Generation and Recombination in P-N Junctions and P-N Junction Characteristics," Proc. IRE, 45, 1228-1243, Sept., 1957.

49. D. Farrington, "Anomalous Current-Voltage Characteristics of Floating Bulk Masfet's," Electrochem. Soc. 1974 Fall Meeting, Extended Abstracts, 345-347, October, 1974.

50. J. Tihanyi and H. Schlotterer, "Properties of ESFI MOS Transistors Due to the Floating Substrate and the Finite Volume," International Electron Devices Meeting Technical Abstract, 39-42, December, 1974.

51. D. W. Flately and M. E. Ham, "Electrical Instabilities in SOS/MOS Transistors," Electrochem. Soc. 1974 Fall Meeting, Extended Abstracts, 487-489, October, 1974.

52. T. Klein, "Technology and Performance of Integrated Complementary MOS Circuits," IEEE J. Solid State Circuits, SC-4, 122-130, June, 1969.

53. D. S. Woo, et.al., "Manufacturing Methods for Silicon Devices on Insulating Substrates," Technical Report AFML-IR-516-3 (VI), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio, January 30, 1975.

54. "SOS/CMOS Reliability Data," Inselek Co., Princeton, N.J.

55. M. Simon, Final Technical Report, Contract N00014-73-C-0236, Research Triangle Institute, 1974.

56. Silicon on Sapphire, available from Dataquest, 480 California Avenue, Palo Alto, CA 94306.

57. Y. S. Touloukian, et.al., Thermophysical Properties of Matter, Vol. 2, Thermal Conductivity - Non-metallic Solids, IFI/Plenum Press, New York, 1970.

58. Y. S. Touloukian and E. H. Buyco, Thermophysical Properties of Matter, Vol. 5, Specific Heat - Non-metallic Solids, IFI/Plenum Press, New York, 1970.

59. IEEE SOS Technology Workshop, South Lake Tahoe, California, Sept., 1975.

60. R. L. Nielsen, D. T. Butcher, and R. S. Ronen, "Electrical and Radiation Characteristics of MOS/SOS Linear Devices," IEEE Trans. Nuc. Sci., NS-22, 2629-2633, Dec., 1975.