

# RCA REVIEW

*a technical journal*

*Published quarterly by*

RCA RESEARCH AND ENGINEERING

*in cooperation with all subsidiaries and divisions of*

RADIO CORPORATION OF AMERICA

---

## INDEX, VOLUME 29

---

### CONTENTS

#### March 1968

	PAGE
A Solid-State Low-Noise Preamplifier and Picture-Tube Drive Amplifier for a 60-MHz Video System .....	3
O. H. SCHADE, SR.	
The Si-SiO <sub>2</sub> Solid-Solid Interface System .....	22
A. G. REVEZS AND K. H. ZAININGER	
Multipath Problems in Communications between Low-Altitude Spacecraft and Stationary Satellites .....	77
S. H. DURRANI AND H. STARAS	
Probability Distribution of Time to Phase Lock for a Second-Order Phased-Locked Loop .....	106
F. S. KEBLAWI	
Quadrature Detection PSK System and Interference .....	122
V. F. VOLERTAS	
RCA Technical Papers .....	141
Authors .....	144

#### June 1968

R-F Sputtering Process .....	149
J. L. VOSSEN AND J. J. O'NEILL, JR.	
Laminated-Ferrite Memories—Review and Evaluation .....	180
R. SHAHBENDER	
An MOS-Transistor-Driven Laminated-Ferrite Memory .....	199
I. GORDON, R. L. HARVEY, H. I. MOSS, A. D. ROBBI, J. W. TUSKA, J. T. WALLMARK, AND C. WENTWORTH	

High-Power Frequency Doublers Using Coupled TEM Lines .....	230
C. SUN	
Multiple-Loop Frequency-Compressive Feedback for Angle-Modulation Detection .....	252
H. HEINEMANN, A. NEWTON, AND J. FRANKLE	
Double-Stream Interaction in a Thin Semiconductor Layer .....	270
B. B. ROBINSON AND B. VURAL	
Manipulation of Sculptured Surfaces .....	281
E. P. HELPERT	
Multispectral Vidicon Camera Study .....	287
J. J. DISHLER AND D. F. LONGCOY	
RCA Technical Papers .....	298
Authors .....	301

### September 1968

MOS Field-Effect Transistor Drivers for Laminated-Ferrite Memories .....	311
W. A. BÖSENBERG, D. FLATLEY, AND J. T. WALLMARK	
Development of a 64-Output MOS Transistor Selection Tree .....	324
J. T. GRABOWSKI	
Monolithic Sense Amplifier for Laminated-Ferrite Memories .....	355
H. R. BEELITZ	
An Experimental Pulsed CdS Laser Cathode-Ray Tube .....	379
F. H. NICOLL	
Electromagnetic Wave Propagation in Superconductors .....	386
P. BURA	
Selection Diversity with Non-Zero Correlations .....	401
A. SCHMIDT	
Adaptive Detection Mode with Threshold Control as a Function of Spatially Sampled Clutter-Level Estimates .....	414
H. M. FINN AND R.S. JOHNSON	
RCA Technical Papers .....	465
Authors .....	468

### December 1968

New Process Technologies for Microelectronics—Foreword .....	473
J. A. AMICK AND N. E. WOLFF	
Isolation Techniques for Fabricating Integrated Circuits	
I. Laminate Substrates .....	475
J. A. AMICK AND A. W. FISHER	
II. Dielectric Refill Techniques and Decal Air Isolation .....	485
A. I. STOLLER AND W. H. SCHILP, JR.	
Relationship Between the Performance of a Linear Amplifier Microcircuit and the Isolation Technique Used in Its Construction ...	507
A. I. STOLLER	
Apparatus for Chemical Vapor Deposition of Oxide and Glass Films ..	525
WERNER KERN	

Diffusion Characteristics and Applications of Doped Silicon Dioxide Layers Deposited from Silane ( $\text{SiH}_4$ ) .....	533
A. W. FISHER, J. A. AMICK, H. HYMAN, and J. H. SCOTT, JR.	
Diffusion Characteristics of Doped Silicon Dioxide Layers Deposited from Premixed Hydrides .....	549
A. W. FISHER and J. A. AMICK	
A Technique for Measuring Etch Rates of Dielectric Films .....	557
WERNER KERN	
DC Sputtering with RF-Induced Substrate Bias .....	566
J. L. VOSSEN and J. J. O'NEILL, JR.	
Additive Processing Techniques for Printed-Circuit Boards .....	582
R. J. RYAN, T. E. McCURDY, and N. E. WOLFF	
Calculation of the Electrical Parameters of Transistor Chips from Measurements Made on Packaged Units .....	600
A. I. STOLLER	
Controlled Oxidation of Silane .....	618
K. STRATER	
Development of P-Channel Enhancement MOS Triodes .....	630
P. DELIVORIAS	

## AUTHORS INDEX, VOLUME 29

	ISSUE	PAGE
Amick, J. A. (Coauthor)—"Isolation Techniques for Fabricating Integrated Circuits I. Laminate Substrates" ...	Dec.	475
(Coauthor)—"Diffusion Characteristics and Applications of Doped Silicon Dioxide Layers Deposited from Silane ( $\text{SiH}_4$ )" .....	Dec.	533
(Coauthor)—"Diffusion Characteristics of Doped Silicon Dioxide Layers Deposited from Premixed Hydrides" .....	Dec.	549
Beelitz, H. R.—"Monolithic Sense Amplifier for Laminated-Ferrite Memories" .....	Sept.	355
Bösenberg, W. A. (Coauthor)—"MOS Field-Effect Transistor Drivers for Laminated-Ferrite Memories" .....	Sept.	311
Bura, P.—"Electromagnetic Wave Propagation in Superconductors" .....	Sept.	386
Delivorias, P.—"Development of P-Channel Enhancement MOS Triodes" .....	Dec.	630
Dishler, J. J. (Coauthor)—"Multispectral Vidicon Camera Study" .....	June	287
Durrani, S. H. (Coauthor)—"Multipath Problems in Communications from Low-Altitude Spacecraft and Stationary Satellites" .....	Mar.	77
Finn, H. M. (Coauthor)—"Adaptive Detection Mode with Threshold Control as a Function of Spatially Sampled Clutter-Level Estimates" .....	Sept.	414
Fisher, A. W. (Coauthor)—"Isolation Techniques for Fabricating Integrated Circuits I. Laminate Substrates" ...	Dec.	475
(Coauthor)—"Diffusion Characteristics and Applications of Doped Silicon Dioxide Layers Deposited from Silane ( $\text{SiH}_4$ )" .....	Dec.	533
(Coauthor)—"Diffusion Characteristics of Doped Silicon Dioxide Layers Deposited from Premixed Hydrides" .....	Dec.	549
Flatley, D. (Coauthor)—"MOS Field-Effect Transistor Drivers for Laminated-Ferrite Memories" .....	Sept.	311

Frankle, J. (Coauthor)—“Multiple-Loop Frequency-Compressive Feedback for Angle-Modulation Detection” . . . . .	June	252
Gordon, I. (Coauthor)—“An MOS-Transistor-Driven Laminated-Ferrite Memory” . . . . .	June	199
Grabowski, J. T.—“Development of a 64-Output MOS Transistor Selection Tree” . . . . .	Sept.	324
Harvey, R. L. (Coauthor)—“An MOS-Transistor-Driven Laminated-Ferrite Memory” . . . . .	June	199
Heinemann, H. (Coauthor)—“Multiple-Loop Frequency-Compressive Feedback for Angle-Modulation Detection” . . . . .	June	252
Helpert, E. P.—“Manipulation of Sculptured Surfaces” . . . . .	June	281
Hyman, H. (Coauthor)—“Diffusion Characteristics and Applications of Doped Silicon Dioxide Layers Deposited from Silane (SiH <sub>4</sub> )” . . . . .	Dec.	533
Johnson, R. S. (Coauthor)—“Adaptive Detection Mode with Threshold Control as a Function of Spatially Sampled Clutter-Level Estimates” . . . . .	Sept.	414
Keblawi, F. S.—“Probability Distribution of Time to Phase Lock for a Second-Order Phased-Locked Loop” . . . . .	Mar.	106
Kern, W.—“Apparatus for Chemical Vapor Deposition of Oxide and Glass Films” . . . . .	Dec.	525
—“A Technique for Measuring Etch Rates of Dielectric Films” . . . . .	Dec.	557
Longcoy, D. F. (Coauthor)—“Multispectral Vidicon Camera Study” . . . . .	June	287
McCurdy, T. E. (Coauthor)—“Additive Processing Techniques for Printed-Circuit Boards” . . . . .	Dec.	582
Moss, H. I. (Coauthor)—“An MOS-Transistor-Driven Laminated-Ferrite Memory” . . . . .	June	199
Newton, A. (Coauthor)—“Multiple-Loop Frequency-Compressive Feedback for Angle-Modulation Detection” . . . . .	June	252
Nicoll, F. H.—“An Experimental Pulsed CdS Laser Cathode-Ray Tube” . . . . .	Sept.	379
O'Neill, J. J., Jr. (Coauthor)—“R-F Sputtering Processes” . . . . .	June	149
(Coauthor)—“DC Sputtering with RF-Induced Substrate Bias” . . . . .	Dec.	566
Revesz, A. G. (Coauthor)—“The Si-SiO <sub>2</sub> Solid-Solid Interface System” . . . . .	Mar.	22
Robbi, A. D. (Coauthor)—“An MOS-Transistor-Driven Laminated-Ferrite Memory” . . . . .	June	199
Robinson, B. B. (Coauthor)—“Double-Stream Interaction in a Thin Semiconductor Layer” . . . . .	June	270
Ryan, R. J. (Coauthor)—“Additive Processing Techniques for Printed-Circuit Boards” . . . . .	Dec.	582
Schade, O. H., Sr.—“A Solid-State Low-Noise Preamplifier and Picture Tube Drive Amplifier for a 60-MHz Video System” . . . . .	Mar.	3
Schilp, W. H., Jr. (Coauthor)—“Isolation Techniques for Fabricating Integrated Circuits II. Dielectric Refill Techniques and Decal Air Isolation” . . . . .	Dec.	485
Schmidt, A.—“Selection Diversity with Non-Zero Correlations” . . . . .	Sept.	401
Scott, J. H., Jr. (Coauthor)—“Diffusion Characteristics and Applications of Doped Silicon Dioxide Layers Deposited from Silane (SiH <sub>4</sub> )” . . . . .	Dec.	533
Shahbender, R.—“Laminated-Ferrite Memories—Review and Evaluation” . . . . .	June	180
Staras, H. (Coauthor)—“Multipath Problems in Communications between Low-Altitude Spacecraft and Stationary Satellites” . . . . .	Mar.	77
Stoller, A. I. (Coauthor)—“Isolation Techniques for Fabricating Integrated Circuits II. Dielectric Refill Techniques and Decal Air Isolation” . . . . .	Dec.	485

—“Relationship Between the Performance of a Linear Amplifier Microcircuit and the Isolation Technique Used in Its Construction” .....	Dec.	507
—“Calculation of the Electrical Parameters of Transistor Chips from Measurements made on Packaged Units” .....	Dec.	600
Strater, K.—“Controlled Oxidation of Silane” .....	Dec.	618
Sun, C.—“High-Power Frequency Doublers Using Coupled TEM Lines” .....	June	230
Tuska, J. W. (Coauthor)—“An MOS-Transistor-Driven Laminated-Ferrite Memory” .....	June	199
Volertas, V. F.—“Quadrature Detection PSK System and Interference” .....	Mar.	122
Vossen, J. L. (Coauthor)—“R-F Sputtering Processes” .....	June	149
(Coauthor)—“DC Sputtering with RF-Induced Substrate Bias” .....	Dec.	566
Vural, B. (Coauthor)—“Double-Stream Interaction in a Thin Semiconductor Layer” .....	June	270
Wallmark, J. T. (Coauthor)—“An MOS-Transistor-Driven Laminated-Ferrite Memory” .....	June	199
(Coauthor)—“MOS Field-Effect Transistor Drivers for Laminated-Ferrite Memories” .....	Sept.	311
Wentworth, C. (Coauthor)—“An MOS-Transistor-Driven Laminated-Ferrite Memory” .....	June	199
Wolff, N. E. (Coauthor)—“Additive Processing Techniques for Printed-Circuit Boards” .....	Dec.	582
Zaininger K. H. (Coauthor)—“The Si-SiO <sub>2</sub> Solid-Solid Interface System” .....	Mar.	22



# RCA REVIEW

*a technical journal*

*Published quarterly by*

RCA LABORATORIES

*in cooperation with all subsidiaries and divisions of*

RCA CORPORATION

---

## INDEX, VOLUME 30

---

### CONTENTS

March 1969

	PAGE
Avalanche Diodes as UHF and L-Band Sources .....	3
K. K. N. CHANG	
Large-Signal Transit-Time Effects in the MOS Transistor .....	15
J. R. BURNS	
Resolution Limitations of Electromagnetically Focused Image-Intensifier Tubes .....	36
I. P. CSORBA	
A Potential Solid-State Flying-Spot Scanner .....	53
J. I. PANKOVE AND A. R. MOORE	
The Status of Threshold Logic .....	63
R. O. WINDER	
A Broadband High-Gain L-Band Power Amplifier Module .....	85
R. L. BAILEY AND J. R. JASINSKI	
Close Confinement Gallium Arsenide P-N Junction Lasers With Reduced Optical Loss at Room Temperature .....	106
H. KRESSEL AND H. NELSON	
Stepped-Scanned Ring Array .....	114
J. EPSTEIN AND O. M. WOODWARD	
Electronic Boresight Shift in Space-Borne Monopulse Systems .....	150
H. HONDA	
RCA Technical Papers .....	198
Patents .....	201
Authors .....	203

June 1969

Digit-by-Digit Transcendental-Function Computation .....	209
R. J. LINHARDT AND H. S. MILLER	

	PAGE
Transmission and Reflection Group Delay of Butterworth, Chebychev, and Elliptic Filters .....	248
C. M. KUDSIA AND N. K. M. CHITRE	
High-Frequency Behavior of Microstrip Transmission Lines .....	268
L. S. NAPOLI AND J. J. HUGHES	
Single-Frequency Argon Laser .....	277
L. GOROG AND F. W. SPONG	
Vapor Pressure Data for the Solid and Liquid Elements .....	285
R. E. HONIG AND D. A. KRAMER	
The Effect of Barrier Recombination on Production of Hot Electrons in a Metal by Forward Bias Injection in a Schottky Diode .....	306
R. WILLIAMS	
Current-Voltage Characteristics of Silver-n-Type GaP Schottky Barriers .....	314
C. R. WRONSKI	
Double Injection Electroluminescence in Anthracene .....	322
J. DRESNER	
Switching and Storage Characteristics of MIS Memory Transistors ..	335
J. T. WALLMARK AND J. H. SCOTT	
Theory of the Switching Behavior of MIS Memory Transistors .....	366
E. C. ROSS AND J. T. WALLMARK	
RCA Technical Papers .....	382
Patents .....	385
Authors .....	389

### September 1969

A Theory for the High-Efficiency Mode of Oscillation in Avalanche Diodes .....	397
A. S. CLORFEINE, R. J. IKOLA, AND L. S. NAPOLI	
Characteristics of a Sealed-Off He <sup>3</sup> -Cd <sup>114</sup> Laser .....	422
J. R. FENDLEY, JR., I. GOROG, K. G. HERNQVIST, AND C. SUN	
Low-Radiation-Noise He-Ne Laser .....	429
K. G. HERNQVIST	
The Acoustoelectric Effects and the Energy Losses by Hot Electrons— Part IV Field and Temperature Dependence of Electronic Trans- port .....	435
A. ROSE	
Ionospheric Phase Distortion and Faraday Rotation of Radio Waves ..	475
T. MURAKAMI AND G. S. WICKIZER	
Spectral Analysis of Turbulent Wakes .....	504
D. A. DEWOLF	
Bistatic Clutter in a Moving Receiver System .....	518
E. G. MCCALL	

	PAGE
Miniature Microstrip Circulators Using High-Dielectric-Constant Substrates .....	541
B. HERSHENOV AND R. L. ERNST	
RCA Technical Papers .....	544
Patents .....	546
Authors .....	549

### December 1969

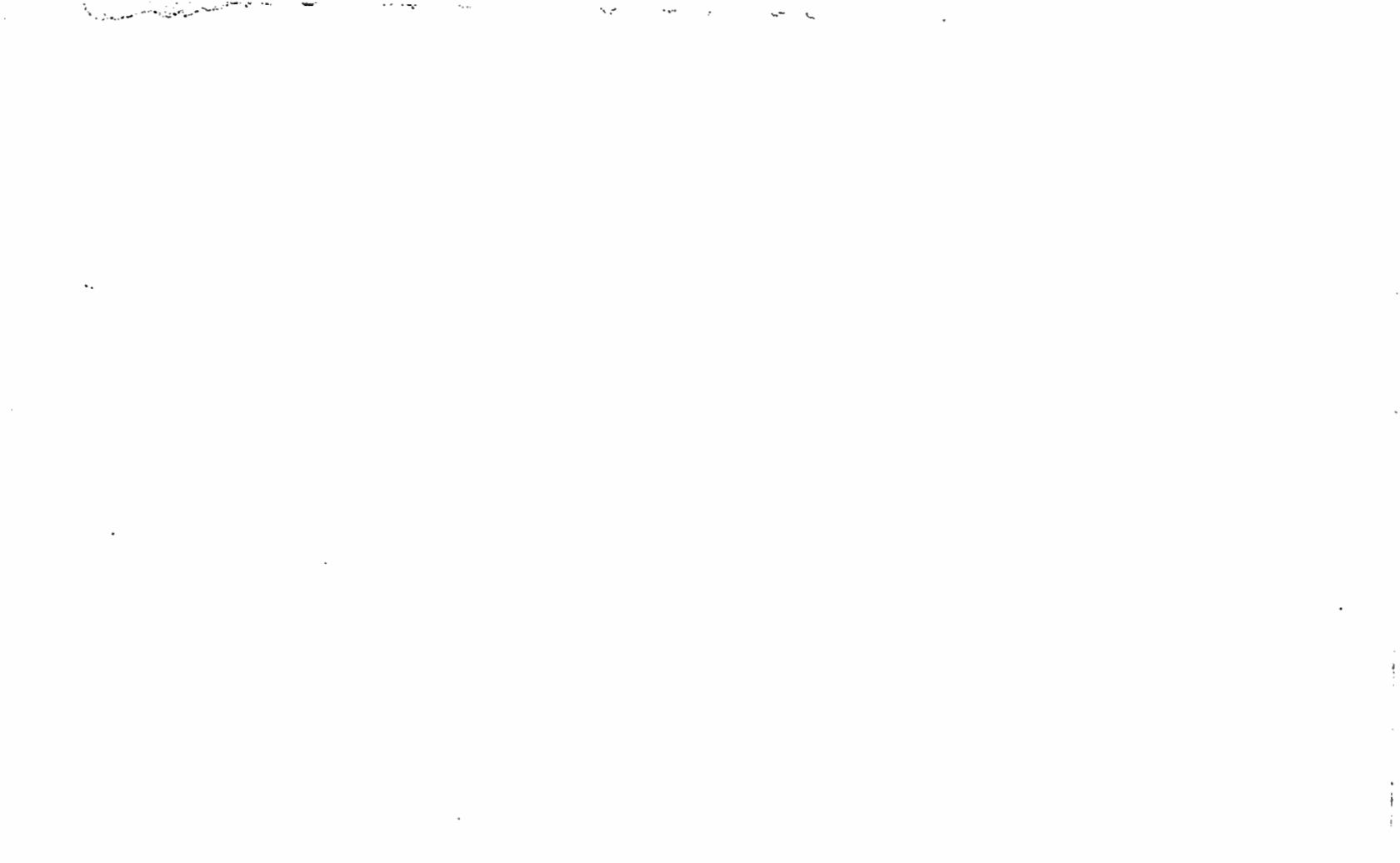
A Photodetector Array for Holographic Optical Memories .....	557
J. M. ASSOUR AND R. D. LOHMAN	
A Reflex Electro-Optic Light Valve Television Display .....	567
D. H. PRITCHARD	
A Technical and Economic Appraisal of the Use of Microwave Energy in the Freeze-Drying Process .....	593
A. C. GRIMM	
Designing High-Frequency Diffused Resistors .....	620
R. E. KLEPPINGER	
Pulsed Heat Conduction in a Layered Semiconductor-Metal Transferred-Electron Oscillator Geometry .....	637
B. S. PERLMAN	
Digital Logic for Radiation Environments: A Comparison of Metal-Oxide-Semiconductor and Bipolar Technologies .....	668
W. J. DENNEHY, A. G. HOLMES-SIEDLE, AND K. H. ZAININGER	
Parallel Processing for Phased-Array Radars .....	709
J. E. COURTNEY AND H. M. HALPERN	
Integrated-Circuit Metalized Plastic Symmetrical Millimeter Trough Waveguide System with Nonreciprocal Elements .....	724
C. P. WEN	
Prediction of the Dielectric Behavior of Titanate Dispersions .....	729
J. M. GUIOR	
RCA Technical Papers .....	733
Patents .....	736
Authors .....	739

### AUTHORS INDEX, VOLUME 30

	ISSUE	PAGE
ASSOUR, J. M. (Coauthor)—A Photodetector Array for Holographic Optical Memories .....	Dec.	557
BAILEY, R. L. (Coauthor)—A Broadband High-Gain L-Band Power Amplifier Module .....	Mar.	85
BURNS, J. R.—Large-Signal Transit-Time Effects in the MOS Transistor .....	Mar.	15
CHANG, K. K. N.—Avalanche Diodes as UHF and L-Band Sources .....	Mar.	3

	ISSUE PAGE
CHITRE, N. K. M. (Coauthor)—Transmission and Reflection Group Delay of Butterworth, Chebychev, and Elliptical Filters .....	June 248
CLORFEINE, A. S. (Coauthor)—A Theory for the High-Efficiency Mode of Oscillation in Avalanche Diodes .....	Sept. 397
COURTNEY, J. E. (Coauthor)—Parallel Processing for Phased-Array Radars .....	Dec. 709
CSORBA, I. P.—Resolution Limitations of Electromagnetically Focused Image-Intensifier Tubes .....	Mar. 36
DENNEY, W. J. (Coauthor)—Digital Logic for Radiation Environments: A Comparison of Metal-Oxide-Semiconductor and Bipolar Technologies .....	Dec. 668
DEWOLF, D. A.—Spectral Analysis of Turbulent Wakes .....	Sept. 504
DRESSNER, J.—Double Injection Electroluminescence in Anthracene .....	June 322
EPSTEIN, J. (Coauthor)—Stepped-Scanned Ring Array .....	Mar. 114
ERNST, R. L. (Coauthor)—Miniature Microstrip Circulators Using High-Dielectric-Constant Substrates .....	Sept. 541
FENDLEY, J. R., JR. (Coauthor)—Characteristics of a Sealed-Off He <sup>3</sup> -Cd <sup>114</sup> Laser .....	Sept. 422
GOROG, I. (Coauthor)—Single-Frequency Argon Laser .....	June 277
(Coauthor) Characteristics of a Sealed-Off He <sup>3</sup> -Cd <sup>114</sup> Laser .....	Sept. 422
GRIMM, A. C.—A Technical and Economic Appraisal of the Use of Microwave Energy in the Freeze-Drying Process .....	Dec. 593
GUIOT, J. M.—Prediction of the Dielectric Behavior of Titanate Dispersions .....	Dec. 729
HALPERN, H. M. (Coauthor)—Parallel Processing for Phased-Array Radars .....	Dec. 709
HERNQVIST, K. G.—Low-Radiation Noise He-Ne Laser .....	Sept. 429
(Coauthor)—Characteristics of a Sealed-Off He <sup>3</sup> -Cd <sup>114</sup> Laser .....	Sept. 422
HERSHENOV, B. (Coauthor)—Miniature Microstrip Circulators Using High-Dielectric-Constant Substances .....	Sept. 541
HOLMES-SIEDLE, A. G. (Coauthor)—Digital Logic for Radiation Environments: A Comparison of Metal-Oxide-Semiconductor and Bipolar Technologies .....	Dec. 668
HONDA, H.—Electronic Boresight Shift in Space-Borne Monopulse Systems .....	Mar. 150
HONIG, R. E. (Coauthor)—Vapor-Pressure Data for the Solid and Liquid Elements .....	June 285
HUGHES, J. J. (Coauthor)—High-Frequency Behavior of Microstrip Transmission Lines .....	June 268
IKOLA, R. J. (Coauthor)—A Theory for the High-Efficiency Mode of Oscillation in Avalanche Diodes .....	Sept. 397
JASINSKI, J. R. (Coauthor)—A Broadband High-Gain L-Band Power Amplifier Module .....	Mar. 85
KLEPPINGER, R. E.—Designing High-Frequency Diffused Resistors .....	Dec. 620
KRAMER, D. A. (Coauthor)—Vapor Pressure Data for the Solid and Liquid Elements .....	June 285
KRESSEL, H. (Coauthor)—Close Confinement Gallium Arsenide P-N Junction Lasers with Reduced Optical Loss at Room Temperatures .....	Mar. 106
KUDSIA, C. M. (Coauthor)—Transmission and Reflection Group Delay of Butterworth, Chebychev, and Elliptical Filters ..	June 248
LINHARDT, R. J. (Coauthor)—Digit-by-Digit Transcendental-Function Computation .....	June 209
LOHMAN, R. D. (Coauthor)—A Photodetector Array for Holographic Optical Memories .....	Dec. 557
MCCALL, E. G.—Bistatic Clutter in a Moving Receiver System ..	Sept. 518
MILLER, H. S. (Coauthor)—Digit-by-Digit Transcendental-Function Computation .....	June 209

	ISSUE PAGE
MOORE, A. R. (Coauthor)—A Potential Solid-State Flying-Spot Scanner .....	Mar. 53
MURAKAMI, T. (Coauthor)—Ionospheric Phase Distortion and Faraday Rotation of Radio Waves .....	Sept. 475
NAPOLI, L. S. (Coauthor)—High-Frequency Behavior of Microstrip Transmission Lines .....	June 268
(Coauthor)—A Theory for the High-Efficiency Mode of Oscillation in Avalanche Diodes .....	Sept. 397
NELSON, H. (Coauthor)—Close Confinement Gallium Arsenide P-N Junction Lasers with Reduced Optical Loss at Room Temperatures .....	Mar. 106
PANKOVE, J. I. (Coauthor)—A Potential Solid-State Flying-Spot Scanner .....	Mar. 53
PERLMAN, B. S.—Pulsed Heat Conduction in a Layered Semiconductor-Metal Transferred-Electron Oscillator Geometry .....	Dec. 637
FRITCHARD, D. H.—A Reflex Electro-Optic Light Valve Television Display .....	Dec. 567
ROSE, A.—The Acoustoelectric Effects and Energy Losses in Hot Electrons—IV Field and Temperature Dependence of Electronic Transport .....	Sept. 435
ROSS, E. C. (Coauthor)—Theory of the Switching Behavior of MIS Memory Transistors .....	June 366
SCOTT, J. H. (Coauthor)—Switching and Storage Characteristics of MIS Memory Transistors .....	June 335
SPONG, F. W. (Coauthor)—Single-Frequency Argon Laser .....	June 277
SUN, C. (Coauthor)—Characteristics of a Sealed-Off He <sup>2</sup> -Cd <sup>114</sup> Laser .....	Sept. 422
WALLMARK, J. T. (Coauthor)—Switching and Storage Characteristics of MIS Memory Transistors .....	June 335
(Coauthor)—Theory of the Switching Behavior of MIS Memory Transistors .....	June 366
WEN, C. P.—Integrated-Circuit Metalized Plastic Symmetrical Millimeter Trough Waveguide System with Nonreciprocal Elements .....	Dec. 724
WICKIZER, G. S. (Coauthor)—Ionospheric Phase Distortion and Faraday Rotation of Radio Waves .....	Sept. 475
WILLIAMS, R.—The Effect of Barrier Recombination on Production of Hot Electrons in a Metal by Forward Bias Injection in a Schottky Diode .....	June 306
WINDER, R. O.—The Status of Threshold Logic .....	Mar. 63
WOODWARD, O. M. (Coauthor)—Stepped-Scanned Ring Array .....	Mar. 114
WRONSKI, C. R.—Current-Voltage Characteristics of Silver-n-Type GaP Schottky Barriers .....	June 314
ZAININGER, K. H. (Coauthor)—Digital Logic for Radiation Environments: A Comparison of Metal-Oxide-Semiconductor and Bipolar Technologies .....	Dec. 668



# RCA REVIEW

*a technical journal*

*Published quarterly by*

RCA LABORATORIES

*in cooperation with all subsidiaries and divisions of*

RADIO CORPORATION OF AMERICA

---

VOLUME 29

MARCH 1968

NUMBER 1

---

## CONTENTS

A Solid-State Low-Noise Preamplifier and Picture-Tube Drive Amplifier for a 60-MHz Video System .....	3
O. H. SCHADE, SR.	
The Si-SiO <sub>2</sub> Solid-Solid Interface System .....	22
A. G. REVEZS AND K. H. ZAININGER	
Multipath Problems in Communications between Low-Altitude Spacecraft and Stationary Satellites .....	77
S. H. DURRANI AND H. STARAS	
Probability Distribution of Time to Phase Lock for a Second-Order Phased-Locked Loop .....	106
F. S. KEBLAWI	
Quadrature Detection PSK System and Interference .....	122
V. F. VOLERTAS	
RCA Technical Papers .....	141
Authors .....	144

---

© 1968 by Radio Corporation of America  
All rights reserved.

RCA REVIEW is regularly abstracted and indexed by *Abstracts of Photographic Science and Engineering Literature*, *Applied Science and Technology Index*, *Bulletin Signalétique des Télécommunications*, *Chemical Abstracts*, *Electronic and Radio Engineer*, *Mathematical Reviews*, and *Science Abstracts (I.E.E.-Brit.)*.

# RCA REVIEW

---

## BOARD OF EDITORS

### *Chairman*

H. W. LEVERENZ  
*RCA Laboratories*

A. A. BARCO  
*RCA Laboratories*

E. D. BECKEN  
*RCA Communications, Inc.*

G. H. BROWN  
*Radio Corporation of America*

A. L. CONRAD  
*RCA Service Company*

E. W. ENGSTROM  
*Radio Corporation of America*

A. N. GOLDSMITH  
*Honorary Vice President, RCA*

G. B. HERZOG  
*RCA Laboratories*

J. HILLIER  
*RCA Laboratories*

R. S. HOLMES  
*RCA Medical Electronics*

E. C. HUGHES  
*RCA Electronic Components*

E. O. JOHNSON  
*RCA Electronic Components*

H. R. LEWIS  
*RCA Laboratories*

G. F. MAEDEL  
*RCA Institutes, Inc.*

L. S. NERGAARD  
*RCA Laboratories*

H. F. OLSON  
*RCA Laboratories*

K. H. POWERS  
*RCA Laboratories*

J. A. RAJCHMAN  
*RCA Laboratories*

F. D. ROSI  
*RCA Laboratories*

L. A. SHOTLIFF  
*RCA International Licensing*

C. P. SMITH  
*RCA Laboratories*

W. M. WEBSTER  
*RCA Laboratories*

### *Secretary*

C. C. FOSTER  
*RCA Laboratories*

---

## REPUBLICATION AND TRANSLATION

Original papers published herein may be referenced or abstracted without further authorization provided proper notation concerning authors and source is included. All rights of republication, including translation into foreign languages, are reserved by RCA Review. Requests for republication and translation privileges should be addressed to *The Manager*.

# A SOLID-STATE LOW-NOISE PREAMPLIFIER AND PICTURE-TUBE DRIVE AMPLIFIER FOR A 60-MHZ VIDEO SYSTEM

By

OTTO H. SCHADE, SR.

RCA Electronic Components  
Harrison, N. J.

*Summary*—The development of MOS transistors with transconductances of 10 to 15 millimhos makes possible the design of wide-band video preamplifiers providing lower noise levels than preamplifiers using muistor tubes. The low-noise input stage of the transistor preamplifier is a cascode circuit composed of two dual-gate transistors. The large frequency-proportional gain correction made in the third stage depresses the  $1/f$  noise of the transistors below their high-frequency noise level and compensates for the capacitive input load of the preamplifier. Thus, the noise spectrum becomes proportional to frequency and appears as a fine-grain random structure of low visibility on the picture-tube screen. Aperture correction and low-pass filters are used to vary the frequency response and video pass band. The 50-volt video-signal output developed by two 2N3866 planar transistors in a cascode circuit and driven from a 90-ohm line is adequate to modulate a high-definition picture tube.

## WIDE-BAND PREAMPLIFIER DESIGN PRINCIPLES

HIGH-DEFINITION video-system preamplifiers have a number of special design features. If the photoelectric signals from a high-resolution storage surface (such as a  $50 \times 50$  millimeter photoconductor) were fed directly to an external amplifier, the noise introduced by the amplifier would severely limit the resolving power of the system. Therefore, a camera-tube electron multiplier is used as the signal source. The peak signal current in the return beam of a high-resolution vidicon may range from 25 nanoamperes to several microamperes depending on the degree of exposure and the diameter and current density of the reading beam. The peak-to-peak signal-to-rms-noise ratio (SNR) in the electron-multiplier output current may be as high as 40 dB in a 60-MHz video pass band. The maximum useful gain of an external preamplifier that follows the internal electron multiplication is determined by the preamplifier noise level as follows.

The value of the output load of the multiplier is the value of the impedance of a parallel combination of a resistance  $R_1$  shunted by a capacitance  $C_1$ , where  $C_1$  is the sum of the multiplier output capacitance (approximately 12 picofarads), circuit capacitances, and the input capacitance of the first amplifier stage. The value of  $C_1$  is approxi-

mately 25 picofarads. The reactance of this capacitance is  $1/\omega C$  or 106 ohms at 60 MHz. Thus, a constant multiplier current continues to develop signal voltages that are inversely proportional to frequency down to a frequency at which the capacitive reactance begins to exceed the value of the load resistance  $R_1$ . The value of  $R_1$  should be high so that its thermal noise and the transistor-generated  $1/f$  noise is minimized. The value of  $R_1$  is limited to a certain maximum value by the frequency range over which a linear relation of amplifier output signals to multiplier current input can be restored by an amplifier stage having a gain proportional to frequency. For example, for a load resistance of 6200 ohms, the 60-MHz gain must be 58 times higher than the low-frequency signal gain (the gain below 1 MHz). As a consequence, white noise injected by an input stage becomes frequency-proportional noise after the gain/frequency correction. An equivalent rms-noise input voltage can be computed by dividing the output noise by the low-frequency gain of the amplifier.

The theoretical equivalent rms-noise voltage  $E_n$  at the input of a compensated amplifier with a capacitive input is given by

$$E_n = \frac{2.6\pi}{3} R_1 C_1 \times 10^{-10} (R_{eq})^{1/2} \Delta f^{3/2}, \quad (1)$$

where  $R_{eq}$  is the equivalent noise resistance and  $\Delta f$  is the bandwidth.  $R_{eq}$  for an electron-tube amplifier is given by

$$R_{eq} \approx \frac{0.89 T}{T_o g_m} \quad (2)$$

where  $T$  is the emitter temperature in degrees K,  $T_o$  is the ambient temperature, and  $g_m$  is the transconductance.

Because the compensated amplifier restores a constant frequency response, the value of the equivalent signal input voltage becomes the product of  $I_s$  (signal current) and  $R_1$ , and

$$\text{SNR} = \frac{2.13 I_s \times 10^9}{C(R_{eq})^{1/2} (\Delta f)^{3/2}} \quad (3)$$

Thus, for a given signal current and bandwidth,

$$\text{SNR} = \frac{K(g_m/T)^{1/2}}{C_1}, \quad (4)$$

where  $K$  is a numerical factor.

It follows that the total capacitance  $C_1$  should be minimized and that the amplifier input stage should have a high transconductance ( $g_m$ ) and a low emitter temperature ( $T$ ). A cascode stage composed of two nuvistor triodes with a total input capacitance  $C_1$  of approximately 8.2 picofarads for a transconductance of 10 millimhos makes a relatively good input amplifier. An emitter temperature  $T$  of 1050°K and an ambient temperature  $T_n$  of 300°K yield an equivalent noise resistance  $R_{eq}$  of  $3.1/g_m$  or 310 ohms. For the frequency channel bandwidth of 60 MHz, a total input capacitance  $C_1$  of 25 picofarads and a total resistance  $R_1$  of 6200 ohms, Equation (1) yields an equivalent noise voltage  $E_n$  of approximately 0.6 millivolt rms for nuvistors.

MOS transistors have transconductance values similar to those of nuvistor tubes. Their input capacitance is somewhat lower, but, as explained below, the input gate must be protected from high-voltage transients by shunt diodes. Under this condition, MOS transistors have almost the same total capacitance  $C_1$  as the nuvistor triode configuration described above. Junction temperature is uncertain. If a relatively high emitter temperature is assumed ( $T = 480^\circ\text{K}$ ), a transconductance  $g_m$  of 10 millimhos, and Equation (2) remains valid for high-frequency noise in a transistor, the equivalent noise resistance  $R_{eq}$  is 142 ohms and Equation (1) yields 0.405 millivolt rms for the equivalent input noise  $E_n$  for MOS transistors.

The equivalent rms input-noise voltages measured on actual amplifiers are higher in value; with a transconductance of 10 millimhos,  $E_n$  is 1 millivolt for nuvistors and 0.68 millivolt for dual-gate MOS transistors. The reason for the disagreement with computed values is not known. The noise contributed by the second unit of the cascode stage and the following stages has been measured in MOS amplifiers by replacing the first cascode unit with a 500-ohm resistor. When this substitution is made, the amplifier noise level decreases to 37% of the value measured when the first cascode unit is in place. Thus, it can be stated that the amplifier stages following the first cascode unit contribute only 6% to the total rms noise.

Preamplifier noise should be less than the "essential" noise in the multiplier output current. Thus, for an essential SNR of 40 dB with a signal level of 1 volt in the output line of the amplifier, the essential noise is 10 millivolts rms.

The gain of the preamplifier is determined by the permissible noise increase. If an increase of 1.6 dB or 20% is allowed, the amplifier output noise  $E_{no}$  should not exceed the value  $E_{no} = (12^2 - 10^2)^{1/2} = 6.7$  millivolts. Hence, the low-frequency amplifier gain  $g_n$  must be limited in this case to  $E_{no}/E_n = 6.7/0.68$  or approximately 10 for the MOS pre-

amplifier. With this gain figure, the signal input voltage is 0.1 volt and requires a sustaining multiplier signal current  $I_s$  of  $0.1/R_1 = 0.1/6200$ , or about 16 microamperes. It is evident that considerable current gain must be provided by the multiplier to increase small photoelectric signals to this value.

The amplifier gain should be adjustable so that it can be increased to accommodate lower signal-to-noise ratios from the camera tube and slow-scan systems, which require a lower bandwidth. In slow-scan systems, the input resistance  $R_1$  should be increased by the bandwidth

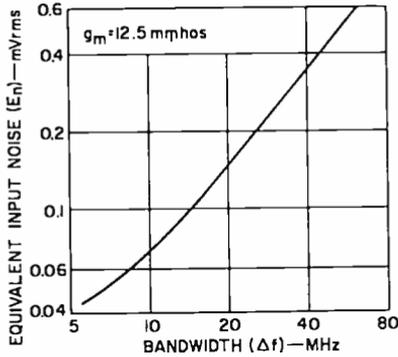


Fig. 1—Equivalent rms-noise input voltage of an MOS cascode preamplifier as a function of bandwidth.

ratio;  $R_1$  should be 62,000 ohms for a bandwidth of 6 MHz. Amplifier gain can also be increased when the frequency bandwidth is reduced, because the essential noise of the camera tube decreases with  $\Delta f^{1/2}$ , whereas the amplifier noise decreases with  $\Delta f^{3/2}$ . The equivalent amplifier input noise  $E_n$  has been determined by measuring the output noise with an rms meter and dividing the value obtained by the low-frequency gain. The input noise  $E_n$  is shown in Figure 1 as a function of video bandwidth  $\Delta f$  for an MOS amplifier with an input transconductance of 12.5 millimhos. The function follows a 1.2 rather than a 1.5 power law.

It is of interest to note that the peak-to-peak noise observed with a high-frequency oscilloscope (Tektronix 585A) appears to have 5 rather than 6 times the rms noise value when observed with continuous sweep rates of 0.1 to 2 microseconds per centimeter. The noise displayed on a picture tube consists only of fine-grained high-frequency noise, because the relatively high  $1/f$  noise of the transistors is depressed below visibility as discussed in the following section.

## MOS TRANSISTOR PREAMPLIFIER

The transfer characteristics of MOS dual-gate transistors are similar in many respects to those of nuvistor tetrodes; although the input capacitance of a transistor is somewhat smaller, 4 to 5 picofarads compared to 8 picofarads for a nuvistor, the output capacitance and the transconductance of both devices are nearly the same. The gate-bias values may require individual adjustment.

Many circuits have an impedance in the source connection. In metal-case transistors this impedance requires separation of the source from the shell and substrate to eliminate high-frequency feedback and a strong frequency dependence on the supply voltage. In early developmental transistors, four insulated leads provided external connections for the source, drain, and both gates. The substrate was connected to the metal shell, which could be grounded. These metal-case transistors, which were originally used in the input and compensation stages of the preamplifier, have been replaced by more recent plastic-case MOS transistors\* that have separate substrate leads. The plastic-case transistors have performed without difficulty.

The design of the transistorized interstage coupling filters for wide-band stages is similar to that used in the nuvistor preamplifier.<sup>1</sup> With this filter design, when the transistor capacitances vary from design values, frequency response can be kept constant by the simple adjustment of two resistance values. The values of the circuit elements in a 60-MHz preamplifier are shown in Figure 2; the frequency response is shown in Figure 3 for the adjustable video pass band.

The circuit in Figure 2 differs from a nuvistor preamplifier in four major respects.

1. Two high-frequency diodes are connected parallel to the input gate of the first stage. Their purpose is to short out charging transients in the coupling capacitor, which has a potential of 1 to 2 kilovolts on the camera-tube side. The gate is held open by the energy gap of the diodes ( $\pm 0.4$  volt).

2. The interstage coupling capacitors have somewhat smaller values than needed for a perfect square-wave response at 60 MHz because the insulated gates draw no "grid" current and therefore cannot limit the surges caused when a sudden change of d-c levels occurs in the camera (for example, during adjustments). The low-frequency limit decreases the tendency toward low-frequency "bounce" and transient signal

---

\* Development Type TA7035.

<sup>1</sup> O. H. Schade, Sr., "A 60-Megacycle Video Chain for High Definition Television Systems", *RCA Review*, Vol. XXVI, No. 2, p. 178, June 1965.





compression or cutoff. A perfect low-frequency response is restored subsequently by a clamp circuit.

3. High-frequency compensation for the capacitive input load is made in the third stage instead of the second stage. This arrangement reduces the low-frequency gain following compensation and depresses the  $1/f$  low-frequency noise of the transistors below their high-frequency noise components. Low-frequency noise takes the form of horizontal "streaks" on a picture-tube screen, and has nearly the same peak

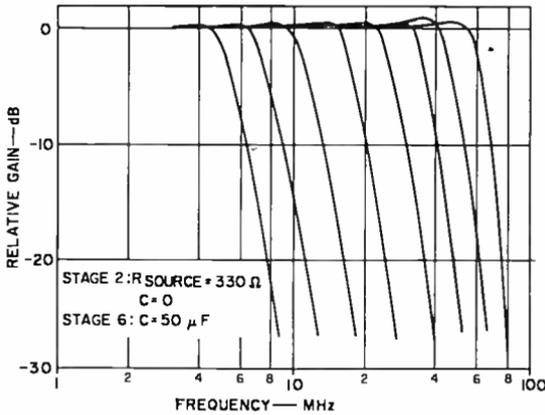


Fig. 3—Preamplifier video pass bands.

amplitude as the high-frequency noise when frequency compensation is made in the second stage. The "streaks" disappear completely when the low-frequency gain following the frequency compensation is not permitted to increase beyond the values shown in Figure 4. The zero-dB level in Figure 4 corresponds to the low-frequency input-signal voltage, which is typically 0.1 volt but which may be as high as 0.25 volt when large aperture correction is needed. When aperture correction is required, the low-frequency gain of the second stage can be reduced by increasing the source resistance. This increase of source resistance prevents signal overload of the third stage, a condition that occurs when the input signal contains spurious pulse signals that exceed the video signal. "White" spikes are clipped by current cutoff in the second stage, which has a dc-restoring diode circuit. The third stage should be able to accommodate 5-volt signals without distortion.

4. The gain/frequency correction is accomplished by low-frequency degeneration in a single nonresonant  $RC$  circuit located in the emitter or source lead of an amplifier. The total shunt capacitance to ground

$C$  should have a value such that the stage gain is degenerated by a factor of 2 at the 60-MHz bandwidth limit. The correct frequency-proportional range and phase correction is obtained when the  $RC$  product has the same value as the input product  $R_1C_1$ . For a transconductance of 10 millimhos,  $C$  is approximately 27 picofarads and  $R$  is about 6 kilohms. The d-c voltage drop for this resistance would be

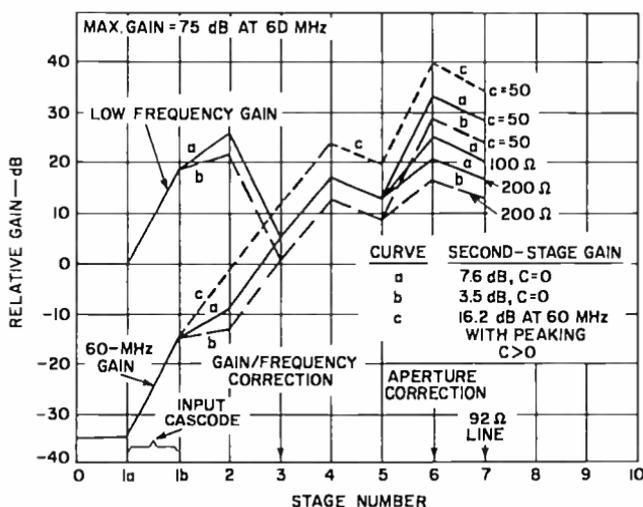


Fig. 4—Relative gain of the preamplifier stages.

36 volts with a 6-milliampere source current. The voltage drop is reduced to 4.6 volts, however, by a parallel n-p-n transistor (type 2N3866) that has a high collector impedance and a low collector capacitance; the transistor acts as a constant-current sink. The current in  $R$  is, therefore, 1 milliampere, requiring that the base bias of the current-sink transistor be adjusted for a collector current of 5 milliamperes. The resistor-transistor combination has a current/voltage characteristic of the form indicated in Figure 5; the zero-signal (6-milliampere) operating point lies in the center of the degenerated low-transconductance section.

The signal-gate potential of the MOS transistor is close to 4.6 volts with respect to ground; for frequencies below 1 MHz the source will follow this potential within a fraction of a volt. At zero signal, gate No. 2 is approximately 3 volts positive with respect to the source. Correct adjustment can be checked by applying a low-frequency sine-wave signal of 8 to 10 volts to gate No. 1 and observing the distortion of the output signal. If the gain of the stage is larger than unity, the planar

transistor operates below saturation. This condition is an indication of either too much current in the planar transistor or too little current in the MOS transistor. Current in the planar transistor can be reduced by reducing the base voltage; current in the MOS transistor can be increased by raising the gate potentials. The gate potentials are optimum when positive and negative signal peaks distort simultaneously at maximum signal input. Once adjusted properly, the circuit is very stable.

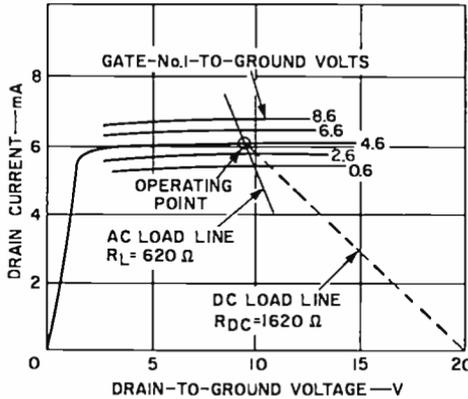


Fig. 5—Output characteristic of the compensating stage equipped with a constant-current sink.

### Aperture Correction

Aperture correction (high-frequency peaking) is used to correct the decreasing high-frequency response of a system component such as a lens, camera, or picture tube. The corrections can be made in stages No. 2 and 6 by inserting and/or bypassing a resistance in the source lead. The frequency/gain correction is made by decreasing the low-frequency gain of the amplifier and then compensating the decrease with an increase of multiplier gain; this method avoids an increase in the amplifier noise level. Figure 6 shows the frequency response of the amplifier for three values of resistance in the source lead of stage 6 and several shunt capacitances. A similar correction is obtainable in stage 2; other corrections can be obtained from the products of stages 2 and 6.

The high-frequency gain increases somewhat without external shunt capacitance when the resistance is switched in, because of the source-to-ground capacitance and the degeneration of the input capacitance of stage 6. A "flat" response is obtained with a source-lead resistance of

320 ohms and zero capacitance in stage 2 along with a capacitance of 50 microfarads in stage 6; or it can be obtained with a source-lead resistor of 180 ohms, zero capacitance in stage 2, and an  $R$  of 100 ohms with a capacitance of zero in stage 6.

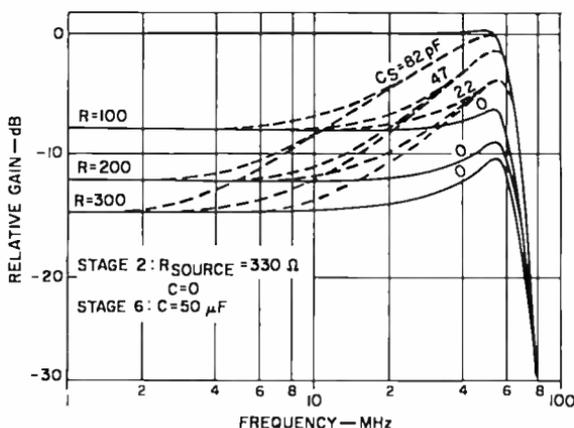


Fig. 6—Frequency response of the preamplifier with aperture correction in stage No. 6.

### Phase Equalizer and Filters

Stage 4 is followed by a phase inverter (stage 5(a)) and emitter follower (stage 5(b)). These stages provide low-impedance sources of opposite phase that drive a phase-correcting circuit. The phase-correcting circuit consists of a series-tuned circuit ( $L_2C_2$ ) and a resistance  $R_2$  as shown in Figure 7(a). With source  $E_1$  operating and  $E_2$  shorted, the circuit is a low-pass filter that has the voltage transfer function  $E_3/E_1$ , shown in Figure 7(b). With source  $E_2$  operating and  $E_1$  shorted, the circuit is a high-pass filter that has an inverse amplitude response  $E_3/E_1$ , when  $E_2 = E_1$ . With both sources operating, the circuit becomes an all-pass circuit that linearizes the phase delay in the cutoff filters.

Without phase correction, the low-pass filters in the 90-ohm line of stage 7 have an impulse response as shown by Figure 7(d). When the high-frequency transient ripple is advanced in the phase corrector, the symmetrical transient shown in Figure 7(c) results. The low-frequency phase delay is controlled by the value of  $R_2$ . Amplitude balance of the transients is obtained when the  $Q$  value of the  $L_2C_2$  circuit is properly adjusted by means of the shunt resistance  $R_s$ . To attain a

balanced ripple frequency, the  $L_2C_2$  product should equal the  $LC$  product of the band-pass filter when  $L$  has a value of about  $0.5 R_2^2C_2$ . The band-pass filters are designed for termination at the far end by a 92-ohm transmission line. Therefore, the emitter resistance  $1/g_m$  of the planar transistor must be increased by a series resistance to 92 ohms, so that the filters are terminated at the driving end. The filter capacitances,  $C$ , are about  $0.9/(f_cR\pi)$ , where  $R$  is 92 ohms and  $f_c$  is the cut-

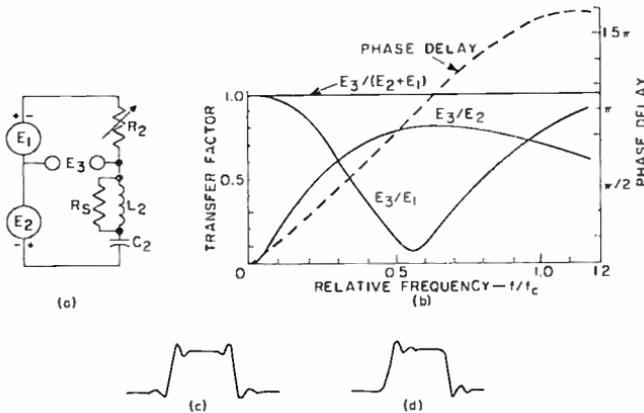


Fig. 7—(a) Circuit, (b) voltage transfer functions, and (c) high-frequency transient of the phase corrector circuit; (d) shows impulse response without correction.

off frequency at the 3-dB-down point. The values of the inductances,  $L$ , are  $R^2C$ . The attenuation characteristics obtained with the phase-corrected low-pass filters are shown in Figure 3. Capacitors  $C$  and  $C/2$  shown in stage 7 of the circuit in Figure 2 are silver-mica types; ceramic capacitors were found to have insufficient  $Q$  values, particularly at lower frequencies (low  $Q$  values cause amplitude losses at mid-band frequencies).

The phase shift introduced by aperture correction in stage 2 or stage 6 compensates the high-frequency phase delay resulting from most causes of decreasing high-frequency response. When the aperture-correction phase shift is incorrect, a high-frequency boost with an inverse phase shift is needed. Aperture correction with an inverse phase shift can be introduced by increasing the signal voltage  $E_2$  of the phase inverter in stage 5(a). The signal voltage is raised through an increase in the value of the output load resistance  $R_3$ , which supplies the high-frequency components to stage 6. Symmetrical transitions with balanced underswing and overshoot, indicating minimum phase error, can

be obtained by suitable combinations of the two types of aperture correction.

### **Circuit Layout**

The circuits are mounted on a copper shelf supported by an aluminum front plate that closes one side of the camera. This plate also provides mounting space for terminals and controls. The first unit of the cascode input stage is mounted at the camera-tube socket and is connected by a short 90-ohm cable to the second unit in the preamplifier. Electrolytic bypass capacitors are pressed into split bushings soldered to the copper plate; this arrangement provides low-reactance grounding. The small interstage filter coils are wound on  $\frac{1}{2}$ -watt or  $\frac{1}{4}$ -watt resistors.

The coupling networks are  $\pi$ -section filters similar in design and adjustment to those used previously in nuvistor preamplifiers.<sup>1</sup> All leads are made as short as possible. An air space of  $\frac{3}{16}$  inch is provided between the coils and transistors and the copper chassis; this configuration provides a low-reactance ground plane and shielding against interstage coupling. To minimize mounting capacitance, the active network is supported by the stand-off insulators for the gate-bias circuits, the grounded shell and substrate leads, and the terminals of the bypass capacitors. The planar transistor in stage 3 is located below the chassis.

The amplifier described is the second model built with MOS transistors. Aside from adjustments of gate-bias voltage to accommodate individual transistor characteristics, the second amplifier model duplicates the frequency response of the first model with no change in circuitry.

### PICTURE-TUBE DRIVE CIRCUIT

The cascode circuit shown in Figure 8 is driven directly from a 90-ohm line. The capacitive load of the picture-tube grid circuit is 7.5 picofarads. The output capacitance of the planar transistor (type 2N3866), including radiator and mounting capacitance, is approximately 9.5 picofarads. This capacitance dictates a value of 400 ohms for the load resistance  $R$  at a video pass band of 70 MHz. The load resistance is located on the transistor side of the series inductance, because the total transistor output capacitance is higher than the picture-tube capacitance.<sup>1</sup> The 0.7-microhenry inductance of the external series coil is approximately two-thirds of the total allowable series inductance to the internal grid capacitance of the picture tube; the grid- and cathode-lead inductances make up the remaining third.



maximum video-signal output including blanking is 50 volts. The zero-signal power dissipation in the second unit is 2.2 watts. Potentials and currents are indicated in Figure 8. The case temperature of the power transistor is less than 100°C because of radiation and conduction cooling that takes place through a standoff insulator.

Preliminary circuit tests can be made with the full series inductance  $L$  of about 1.2 microhenries and with an oscillograph probe of 7 picofarads capacitance substituted for the load capacitance of the picture tube. When the external series inductance is mounted on the picture-tube chassis close to the tube pins, it must be decreased in value because of the inductances of the tube leads.

The true frequency response at the internal grid-cathode capacitance of the tube can be measured by replacing the dc-restorer circuit with a grid resistor and using the screen brightness as an intensity indicator. The picture-tube beam is biased below cutoff at zero signal so that light excitation is obtained only from the positive peak when sine-wave signals are applied. A constant-modulation transfer is indicated when the mean screen illumination remains constant as observed visually or with a photoelectric meter. A constant response can be obtained by adjusting the series inductance and its damping resistor. The beam focus of the picture tube does not affect response measurement.

#### OVER-ALL STABILITY

The maximum over-all voltage gain of the video system with aperture correction from preamplifier input to picture-tube grid is approximately 100 dB at 60 MHz. For complete stability, i.e., zero feedback between camera and picture monitor, low-reactance ground planes must enclose the camera and amplifier circuits so that ground-potential differences in these units are minimized. All voltage and power cables (including deflection and focus supplies) supplying the camera or monitor units must be equipped to bypass r-f to the respective ground planes at their point of entrance to prevent feedback by radiation. It is impossible to eliminate the ground reactance of interconnecting cables at 60 MHz. For example, a "dead" 8-inch video cable soldered to the ground plane of the camera-tube socket housing and shorted at the other end (the end that protrudes from the housing) is a sufficiently good antenna to pick up feedback from the picture-tube. The feedback disappears when the outside cable end is grounded to the camera housing.

The presence of regenerative feedback below the oscillation point can be detected by observing the high-frequency noise displayed on the

monitor tube at maximum gain; each noise-pulse shock excites the system and causes it to "ring" at the cutoff frequency. The ringing gives rise to clusters or bursts of decaying noise pulses rather than randomly distributed single pulses.

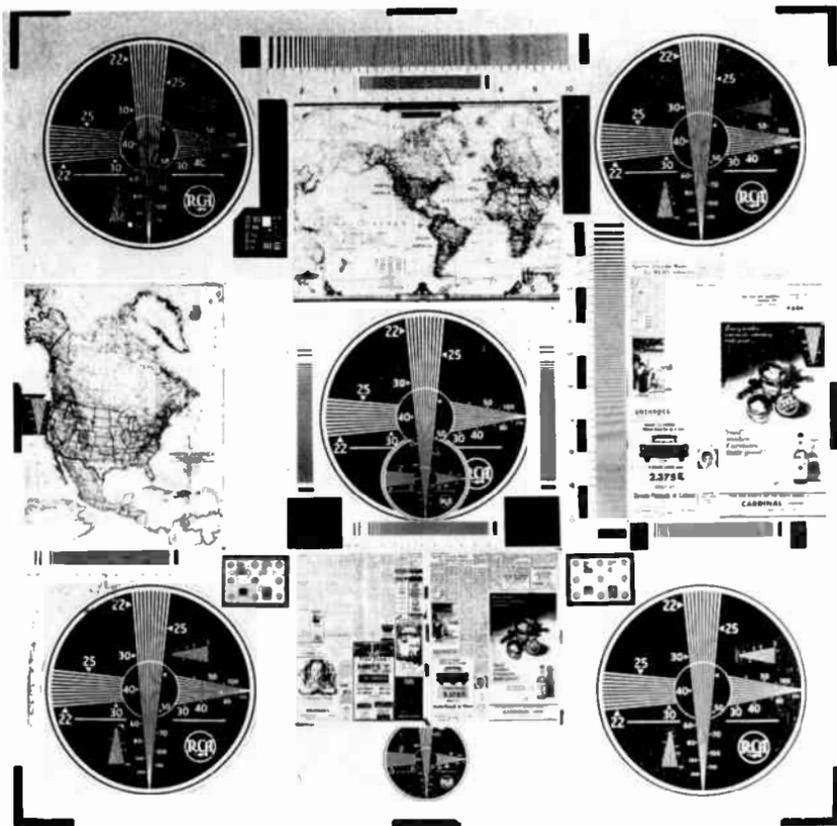


Fig. 10—Photograph of test pattern.

With adequate bypassing of supply leads and with good ground planes (not the common socket-lead ground braid) the 60-MHz video system is completely stable and free from interference.

#### CONCLUSIONS

A complete video system with a pass band of at least 60 MHz can be built with high-transconductance dual-gate MOS transistors and

high-frequency planar transistors. The system should include processing amplifiers for blanking insertion and gamma correction.

A system with these specifications was built by the author; its performance is demonstrated by the photographs in Figures 10 through 13. These photographs show the spatial frequency response of a video

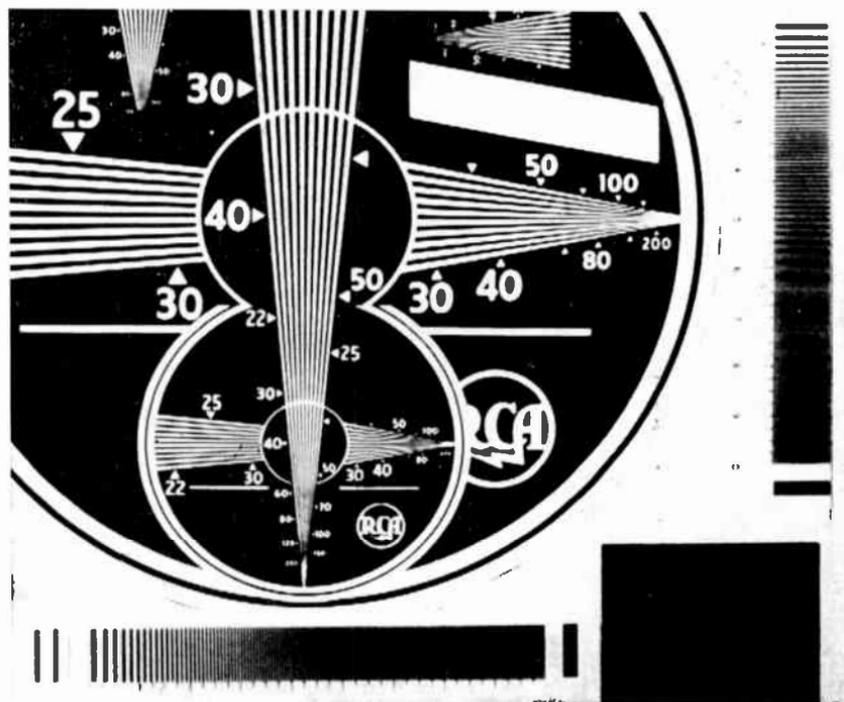


Fig. 11—Enlarged section of monitor photograph showing center of 1760-line test-pattern image for a bandwidth of 60 MHz.

system to a test pattern; the test pattern itself is shown in Figure 10. Picture-tube deflection was expanded by a factor of 1.8 to reduce signal degradation caused by the scanning beam and by the photographic recording process used to obtain the pictures shown in Figures 11 through 13. The video signals were generated by a developmental high-definition 4.5-inch return-beam vidicon camera with a 1760-line scanning raster, 60 fields per second, 3 to 1 interlaced. The bandwidth is 60 MHz in Figure 11, 33 MHz in Figure 12, and 16 MHz in Figure 13. The index numbers in the large circle, when multiplied by 20, give the total number of lines in the picture. The numbers on the small fre-

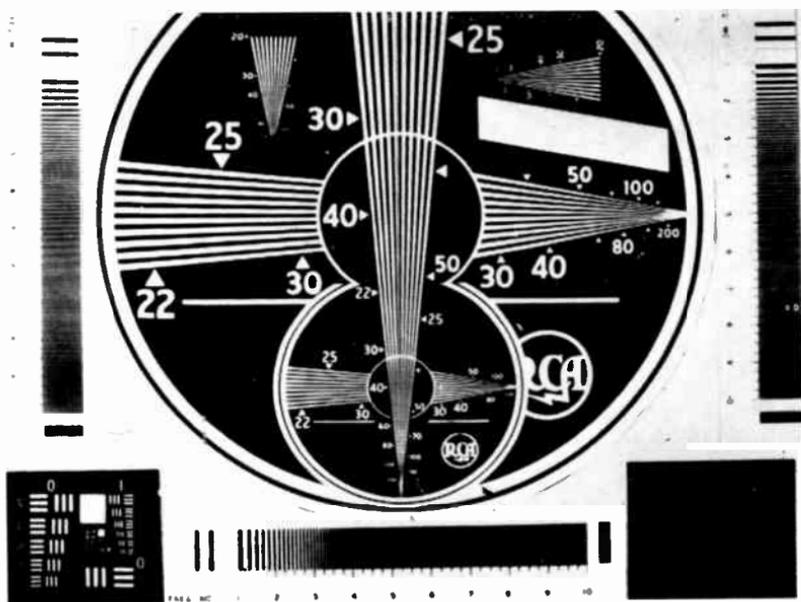


Fig. 12—Enlarged section of monitor photograph showing center of 1760 line test-pattern image for a bandwidth of 33 MHz.

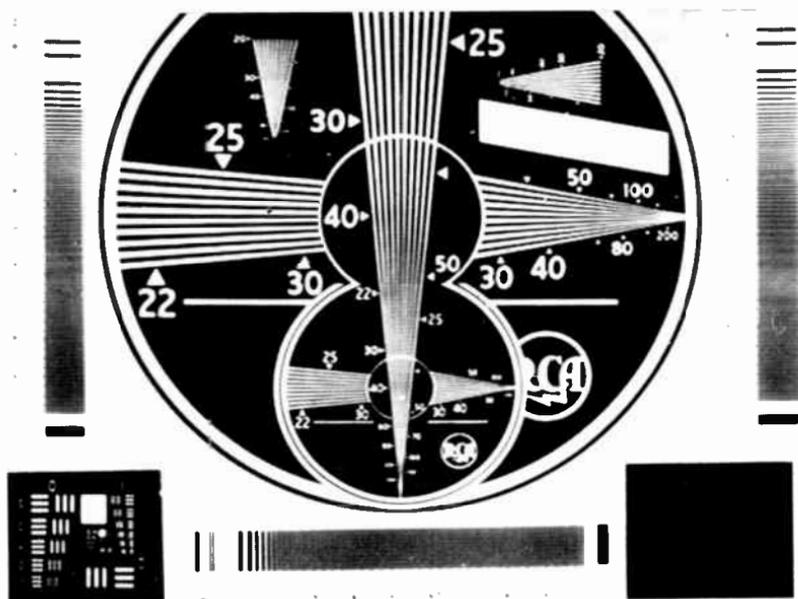


Fig. 13—Enlarged section of monitor photograph showing center of 1760-line test-pattern image for a bandwidth of 16 MHz.

quency wedges and in the small circle must be multiplied by 50 to obtain the total number of lines in the picture dimension.

The visible resolving power of the complete video system is shown in Figure 11 and extends to  $60 \times 50 = 3000$  lines, which corresponds to a cutoff frequency  $f_c$  of 64 MHz. The two-section low-pass filters used in the system have a slower roll-off, resulting in a horizontal resolving power of  $40 \times 50 = 2000$  lines for the 33-MHz filter and  $23 \times 50 = 1150$  lines for the 16-MHz filter.

# THE Si-SiO<sub>2</sub> SOLID-SOLID INTERFACE SYSTEM

BY

A. G. REVESZ AND K. H. ZAININGER

RCA Laboratories,  
Princeton, N.J.

*Summary*—The Si-SiO<sub>2</sub> interface has emerged as one of the best understood solid-solid interface systems. Its properties are reviewed here with special emphasis on electric charge distribution. A qualitative outline of this distribution is followed by a description of methods of preparing SiO<sub>2</sub> films. Thermal and, to a lesser extent, anodic oxidation of silicon are stressed. The influence of oxidation conditions and post-oxidation treatments on the properties of the interface is then discussed in detail, as are various aspects of stability behavior. A description of the "bulk" properties of noncrystalline SiO<sub>2</sub> films is then given on the basis of the quantum chemistry of the Si-O bond and the spatial distribution of  $\pi$ -bonds. The evidence presented shows that with the use of proper techniques, a high degree of perfection in both the Si-SiO<sub>2</sub> interface and the oxide films can be obtained. This is of great importance in achieving desired characteristics and stability behavior in silicon devices and integrated circuits.

## 1. INTRODUCTION

SILICON-BASED discrete devices and especially integrated circuits (IC's) play an essential and ever increasing role in the electronics industry as evidenced by the following data. The estimated dollar volume of silicon-based IC's in the U.S. for 1967 was about \$250,000,000 and that of silicon transistors and diodes roughly \$520,000,000. Silicon dioxide films play a vital role in the manufacture of these devices and ICs, mostly as diffusion masks and protective films, but also in stabilizing the properties of the silicon surface. Furthermore, in contrast to bipolar devices, the basic mechanism of operation of Insulated-Gate Field-Effect Transistors (IGFET) is controlled by the Si-SiO<sub>2</sub> interface.<sup>1</sup>

From a theoretical viewpoint this system showed great promise since a large amount of work had already been done on the silicon-gas and silicon-vacuum interfaces.<sup>2</sup> However, until the advent of the Metal-Oxide-Silicon (MOS) device, the effects of passivating insulator films used on the surface of bipolar transistors and other devices were only minimally understood. In MOS devices, however, the film is an active, integral part of the device, and device operation and characteristics are very sensitive to its properties. Considerable sophistication

in the fabrication process is required to make good MOS insulator films and to produce useful insulator-semiconductor interfaces. This is, basically, the reason for the current interest in semiconductor-insulator interface physics.

In this paper we review the properties of the Si-SiO<sub>2</sub> interface especially from the standpoint of the distribution of electric charges.\* The aim has been to obtain a nearly perfect Si-SiO<sub>2</sub> interface. This goal has been essentially achieved, and SiO<sub>2</sub> on Si is, at the present, certainly one of the best understood solid-solid interface systems.

The organization of this paper is as follows. After some general remarks on surfaces, the behavior of the space-charge region at the silicon surface is discussed. This discussion is followed by a description of charge distribution in the Si-SiO<sub>2</sub> system and a brief summary of the MOS capacitance method for determining this distribution. This method proved to be extremely useful in studying various interface phenomena, including charge transport through the oxide. Since the perfection of the Si-SiO<sub>2</sub> interface is largely determined by the oxide preparation, thermal and anodic oxidation of silicon is discussed in Section 3, together with a brief mention of some deposition methods. Then the influence of oxidation conditions and post-oxidation heat treatments on the interface properties are discussed (Section 4). Specifically, it is pointed out that there is no inherent limitation in obtaining a nearly perfect Si-SiO<sub>2</sub> interface. The virtual elimination of interface states made it possible to investigate ion- and electron-transport processes in SiO<sub>2</sub> films at a level lower than ever before reached. These processes are very important from the standpoint of stability of MOS devices; they are discussed in Section 4. The detailed structure of SiO<sub>2</sub> films plays an important role in the various transport processes. The noncrystallinity of the oxide, the nature of the Si-O bond, and defects in SiO<sub>2</sub> are treated in Section 5.

## 2. CHARGE DISTRIBUTION IN THE Si-SiO<sub>2</sub> SYSTEM

The space-charge region in silicon largely determines the characteristics of MOS devices and is responsible for the many anomalies encountered in bipolar devices. An understanding of the behavior of this

---

\* A massive but uncritical bibliography on this subject has recently been published.<sup>3</sup> Although this was claimed to be a complete catalogue of works published before August 1, 1967, a cursory check showed the omission of several pertinent papers, for instance, on room-temperature oxidation of silicon,<sup>4</sup> ellipsometric studies on SiO<sub>2</sub>,<sup>5,6</sup> and interface properties.<sup>7-10</sup> More recent studies, which are not discussed here, have been published in a special MOS issue of *IEEE Trans. ED-14* (Nov. 1967) and are to appear in the March issue of *Trans. Met. Soc. AIME* (1968).

space-charge region is, therefore, essential from the standpoint of the control of charge distribution in MOS structures. In these structures, a real insulator replaces the idealized charge-free insulator that is usually assumed in semiconductor surface physics. This change results in considerable complexity in the charge distribution. This is briefly outlined, following which a summary of MOS capacitance and its measurement is given.

Since space-charge phenomena in silicon, including various aspects of MOS capacitance, have been treated at length in several books<sup>1,2,11,12</sup> and papers,<sup>13-15</sup> the following discussion is intended only to provide a qualitative understanding that will be helpful in subsequent sections.

## 2.1 General Remarks on Surfaces

There is some ambiguity as to what is meant by the surface of a solid. One can define the surface in the manner of Gibbsian thermodynamics as a dividing *plane* between two bulk phases. In reality, however, the surface must generally be considered as a *region* because the mutual interaction of the phases is not localized in a plane. The surface, in this sense, is a three-dimensional region of discontinuity between different media in which one or more properties characterizing each medium changes. (The term *medium* is used here in a broad sense, and can refer to components, phases, crystallographic orientations, or simply to a drastic difference in impurity content within a phase.) The change generally occurs within a region that is narrow compared to the spatial extent of the system considered. Since different properties will, in general, change in different manners, the depth of the surface region, even within a given system, depends on the property under consideration. For instance, in the case of a silicon-gas interface, the region where the type of atoms and/or crystal structure change is much narrower than the region in the silicon where the density of electrons (holes) differs from that of the bulk. The most common interfaces are solid-liquid, solid-gas, and liquid-gas. Here the change in the state of aggregation occurs over atomic dimensions. These interfaces are extensively treated in the literature, but information on solid-solid interfaces is usually sketchy and scattered.

Although our main concern here is a solid-solid interface, it is expedient to mention briefly other interfaces with solids, especially solid-vacuum and solid-gas systems, because these are well understood. Many concepts that have been developed in connection with these interfaces were later adapted to solid-solid systems. A common feature of these interface systems is that the more or less perfect periodicity of

the crystal potential (taking a single crystal as an example) is disturbed or even completely lost at the interface. Also, the surface atoms are usually displaced from their ideal lattice positions, thus giving rise to a two-dimensional surface structure that deviates from the bulk. All these phenomena lead to the appearance of localized electronic states at the surface called *surface states* or *interface states*.

Although atomic and electronic processes are, by and large, intimately interrelated, one of the two is usually emphasized in dealing with a given problem. Thus, for example, impurity segregation at an interface and epitaxy are generally discussed without considering electronic phenomena, whereas a metal-semiconductor contact is treated on the basis of electron distribution without paying too much attention to structural details at the interface. A knowledge, or at least a qualitative understanding, of the electric potential, electronic wave functions, energy levels, and statistics of electrons at interfaces is a prerequisite to an understanding of complex phenomena such as adsorption, surface reactions, etc. Unfortunately, our knowledge of the structure dependence of electronic properties of surfaces is very rudimentary.

## 2.2 The Surface Space-Charge Region in Silicon

### 2.2.1 Surface Space Charge and Capacitance

Si-SiO<sub>2</sub> as a two-component solid-solid interface system is best approached by first considering a silicon-vacuum interface. The silicon-vacuum interface, like the silicon-gas interface, is conceptually simpler and has already been investigated extensively.<sup>2</sup> Furthermore, the behavior of the space-charge region in silicon can serve as a connecting link between these interface systems because phenomenologically, it is independent of the exact nature of the interface. The space-charge region in silicon, in this sense, is well understood, and the distribution of charges in it as a function of applied field is uniquely determined by the combination of Poisson's equation and Maxwell-Boltzmann statistics. Depending on the polarity and value of the field applied perpendicularly to the surface, the following cases can be distinguished (taking n-type silicon as an example):

- (1) For a positive field the density of electrons (majority carriers in this case) is higher in the space-charge region than in the bulk. This is the *accumulation regime*.
- (2) For zero field the carrier density is uniform throughout the Si crystal. This is the *flat band* condition.
- (3) For a negative field of moderate value, electrons are repelled

from the surface and the positive space charge consists of uncompensated ionized donors. This is the *depletion regime*.

- (4) If the negative field exceeds a given value that depends on the resistivity of silicon, the density of holes (minority carriers in this case) equals or exceeds the bulk density of electrons. Thus, a p-type layer is formed within the space-charge region of the n-type silicon. This is called the *inversion regime*.

If surface states or similar disturbances are present in the vicinity of the phase boundary, then, in principle at least, the space-charge region in silicon will again show these four regimes. However, it is possible that some of the regimes cannot be realized in practice. Such a situation exists, for instance, if the density of surface states is so high that the charges trapped there essentially shield the interior of the semiconductor; that is, the behavior of the semiconductor surface is similar to that of a metal. This makes the control of the space-charge region by a field impossible. Realization of this problem was brought about by failure of efforts to construct a simple semiconductor device based on such control; this was the point of departure for intensive investigations on semiconductor surfaces.<sup>16</sup>

The space charge in the silicon surface per unit area,  $Q_{sc}$ , is normally derived under a certain set of assumptions, including room temperature, nondegeneracy, uniform doping, homogeneity, and semi-infinite geometry. It is an unambiguous function of the semiconductor surface potential,  $\psi_s$ , or surface barrier,  $V_s$ , and a differential capacitance can be associated with it. This *space-charge capacitance* per unit area, is defined as

$$C_{sc} = \frac{\partial Q_{sc}}{\partial V_s}. \quad (1)$$

$C_{sc}$  is also uniquely determined by  $V_s$  or  $\psi_s$  which, in turn, depend on the applied field. The value of  $C_{sc}$  is frequency independent in the accumulation and depletion regimes as long as the frequency is lower than the reciprocal of the dielectric relaxation time of silicon (about  $10^{11}$  for 10 ohm-cm material). In the inversion regime, however,  $C_{sc}$  takes on different forms depending on whether or not the minority carriers can follow the applied bias and/or a-c signal. This is discussed further in Section 2.3.2 when the MOS capacitance is described.

If surface states are present, then charges are distributed among these states and the silicon space-charge region. The extent of the ionization of the surface states depends on their energy level relative

to the Fermi level. One can also associate a differential capacitance,  $C_{ss}$ , with the charge in the surface states,  $Q_{ss}$ . This *surface-state capacitance* can be defined by

$$C_{ss} = \frac{\partial Q_{ss}}{\partial V_s} \quad (2)$$

Since the total surface charge,  $Q_s$ , is the sum of  $Q_{sc}$  and  $Q_{ss}$ , the semiconductor surface capacitance,  $C_s$ , is

$$C_s = \frac{\partial Q_s}{\partial V_s} = C_{sc} + C_{ss} \quad (3)$$

Thus the semiconductor surface capacitance is a parallel combination of space-charge and surface-state capacitances. Whereas  $C_{sc}$  is a unique function of  $V_s$  for a given impurity concentration and temperature,  $C_{ss}$  depends on the particular spatial and energy distribution of surface states and their occupancy at a given external field.

### 2.2.2 Surface Conductance

The concept of surface conductance becomes important when current flow along a direction parallel to the surface is considered, as, for instance, in field-effect studies or in field-effect transistors. However, most investigations of the Si-SiO<sub>2</sub> interface, including our own, have been carried out using MOS capacitors, where surface conductance does not play any role. Therefore, surface conductivity is not discussed in this paper, but it should be understood that at a semiconductor surface, the conductivity generally differs from that in the bulk. The difference occurs for two reasons:

- (1) As mentioned previously, the carrier densities in the space charge region take on values dictated by the surface potential. Except for flat-band conditions, i.e., when  $\psi_s = 0$ , they are different from the bulk values.
- (2) For the case of large positive or negative surface potentials, a potential well exists for one or the other type of carrier and, under certain circumstances, this can reduce the mobility at the surface because of surface scattering.

It is generally found that in MOS transistors with a good Si-SiO<sub>2</sub> interface the surface mobility is about one half of the bulk mobility.

### 2.3 The Si-SiO<sub>2</sub> Interface

Since practical silicon surfaces are generally covered with a very thin (up to about 30-50 Å) oxide film that forms even at room temperature, this oxide film was considered part of the surface region where structural deviations from the bulk occur. The localized electronic states associated with this surface region are generally divided into fast and slow surface states, depending on the speed at which they interact with the silicon space-charge region. The slow states are generally associated with the outer surface of the thin oxide film. The occupation by electrons (holes) of these states can usually be influenced by the ambient, indicating that they participate or even originate in adsorption processes. The fast states, of course, result from the disturbance of the periodicity of the silicon crystal potential. Calculation of the electron (hole) distribution in the silicon space-charge region involves taking into account the occupation statistics of the surface states in addition to use of Poisson's equation and Maxwell-Boltzmann statistics. If the field or the ambient changes with time, then the distribution of the surface states from the standpoint of their response time must also be considered.

#### 2.3.1 Silicon Dioxide as a Nonideal Insulator

The main concern of semiconductor surface physics and chemistry is to control the properties of surfaces in such a manner that a charge distribution, established at zero applied field, will not vary with time and will also be insensitive to changes in ambient. An important step in this direction was achieved when Atalla et al found that the slow surface states could be largely eliminated and the density of fast states significantly reduced by thermal oxidation of silicon.<sup>17</sup> This process led to improved performance of bipolar silicon devices, to a new technology (i.e., the so-called planar technology), and to the development of MOS devices. The working principle of MOS devices is based on the modulation of the space-charge region by an external field applied by means of a metal electrode on top of the insulator. The presence of a relatively thick (about 1000 Å) oxide film instead of a thin film introduces various complications. Charges *within* the oxide, including those at the metal-oxide interface, can no longer be neglected in comparison with charges in the silicon space-charge region. Thus, the silicon-gas interface system, one of the subjects of classical studies of semiconductor surface physics and chemistry, proved to be inadequate. The new system had to be investigated as a solid-solid interface where processes *within* the oxide play an important role. In other words, the

oxide could no longer be considered as a perfect, charge-free insulator, but had to be treated as a substance whose properties, especially as manifested in electrical interactions with the silicon and/or metal electrode, had to be investigated and controlled in detail. One has to consider the defect structure of the insulator<sup>6</sup> and admit the possibility of a non-zero charge density within the insulator.

These charges can be due to traps in the insulator that electronically interact with silicon or metal, and they are, therefore, similar to surface states but with a three-dimensional distribution. Other types of charges can also exist in the oxide. These are the fixed or "built-in" charges; i.e., charges that do not participate in electron transfer processes across the interfaces. These "built-in" charges may be associated with ionic defects in the oxide or with interface states at the Si-SiO<sub>2</sub> interface that are outside the Si forbidden band. In the former case, an electronic transfer takes place between a defect and silicon or metal at some point during processing, for instance, during the oxide growth or after electrode deposition, but does not occur later. Some of these charges may be associated with mobile defects in the oxide, so that under a d-c field, especially at elevated temperature, they might move. In this case, these charges will affect the silicon surface potential in accordance with the laws of electrostatics. Occasionally, these mobile ionic defects can modify the distribution of interface states or even discharge at one of the interfaces. The discharge process is, of course, associated with a current flow normal to the interface.

Another possibility for current is the injection of electrons or holes at the interface(s) and their transport through the oxide. We see thus, that the replacement of an ideal, charge-free insulator by a real one (e.g., SiO<sub>2</sub>) in the MOS structure can potentially lead to a large variety of phenomena, ranging from somewhat modified effects of surface-state type traps to various charge transport processes in the insulator.

### 2.3.2 The MOS System

In the preceding section we have given a qualitative description of the space-charge region in silicon and its dependence on the electric field at the surface. The role of the metal electrode on top of the insulating film is to make possible the establishment of a field normal to the silicon surface. Again, the interface between metal and insulator may not be ideal. There is an electronic interaction between the metal, insulator, and semiconductor, and even at zero applied voltage there is a charge separation between metal and silicon. This is due to a difference of the electrochemical potential of electrons in these three phases.

A schematic energy diagram of the MOS system is shown in Figure 1.

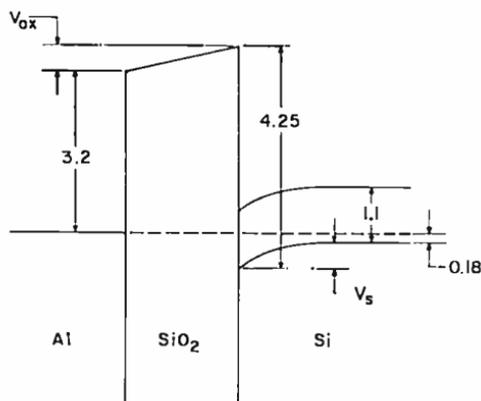


Fig. 1—Energy-band diagram for an MOS system. The metal is aluminum, the silicon is p-type, and the doping level is  $10^{16} \text{ cm}^{-3}$ . The energy values are given in eV, the distance is in arbitrary scale. It is assumed that no surface states or surface dipoles are present. The Si-SiO<sub>2</sub> and Al-SiO<sub>2</sub> barrier energy values were taken from References (19) and (20), respectively. The Al-Si work-function difference for this particular doping level is  $-0.95 \text{ eV}$ .<sup>20</sup> The total voltage drop is the sum of voltage drops in the oxide,  $V_{ox}$ , and in the silicon space charge,  $V_s$ , but its distribution among  $V_{ox}$  and  $V_s$  is a function of the ratio of oxide thickness to silicon Debye length. If this ratio is much larger than unity,  $V_s$  approaches zero, whereas if it is much less than unity,  $V_{ox}$  is very small. (The Debye length of silicon for  $10^{16} \text{ cm}^{-3}$  doping level is  $600 \text{ \AA}$ ).

The capacitance of an MOS structure is given by

$$\frac{1}{C} = \frac{1}{\left( \frac{\partial Q_t}{\partial V_s} \right)} = \frac{1}{C_{ox}} + \frac{1}{C_{sc} + C_{ss}}, \quad (4)$$

where  $Q_t$  indicates total charge, and  $C_{ox}$  is the geometrical capacitance of the oxide, i.e., the capacitance per unit area between the metal electrode and the silicon surface (the capacitance that would be measured if the semiconductor were replaced by a metal). Thus, the measured capacitance consists of the geometric capacitance in series with the parallel combination of the space-charge and surface-state capacitances.

In an idealized case, when the insulator is like a vacuum (i.e., charge free) but with a different dielectric constant, the work function difference between metal and silicon is zero and no surface dipoles exist, the  $C$ - $V$  curve is called "ideal." Such a curve is shown in Figure 2.

Curve (1) in Figure 2 holds if the minority carriers follow both a-c and d-c signals; curve (2) if the minority carriers cannot accumulate at the surface (Schottky depletion layer capacitance); and curve (3) if they follow d-c but not a-c signals.

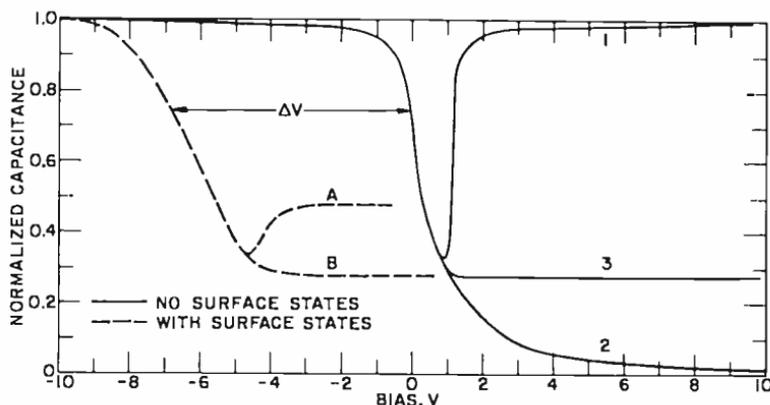


Fig. 2—Normalized MOS capacitance versus bias for 10-ohm-cm p-type silicon with 1000 Å SiO<sub>2</sub> film. The solid curve is for the case of no surface states, no surface dipoles, and no work-function difference between metal and silicon, i.e., it is an ideal curve. If the work-function difference is not zero, then a new ideal curve can be defined by shifting this curve along the voltage in accordance with the work-function difference, provided surface dipoles can still be neglected. In the inversion regime, curve (1) holds if the minority carriers follow both a-c and d-c signals, curve (2) if they cannot accumulate at the surface (Schottky depletion layer capacitance), curve (3) if they follow d-c but not a-c signal. The dashed curve is an experimental result for an identical structure with surface states. Curve (A) corresponds to curve (1) in the ideal case except for the limited response of the minority carriers; curve (B) is analogous to curve (3).  $\Delta V$  is a measure of the density of ionized surface states (see Equation (5)). The bias or gate voltage is equivalent to the voltage on the metal with respect to silicon.

Neglecting the complications introduced by surface dipoles, the following cases can be differentiated in a real situation.

(1) If no surface states are present but the work function difference between metal and Si is non-zero, there will be a parallel shift of the C-V curve without distortion along the voltage axis by an amount equivalent to the work function difference.

(2) If, in addition to a non-zero work function difference, surface states are present that do not follow the a-c signal and whose occupation is independent of applied bias,\* the C-V curve will undergo an

\* These states could either be at the interface but outside the Si forbidden band, or deep enough in the oxide that they do not communicate with the silicon. The charge in both of these states is often referred to as oxide charge.<sup>21</sup>

additional translation along the voltage axis. This shift is a measure of the charge in surface states. Only under these conditions is it possible to define a new ideal  $C$ - $V$  curve that takes the work-function difference into account. It was shown that under proper experimental conditions these requirements can be met.<sup>20</sup>

(3) If interface states are present that do not follow the a-c signal but whose occupation does depend on the d-c bias (i.e., states that are within the forbidden gap of Si), the  $C$ - $V$  curve will be displaced and distorted along the voltage axis, but the ratio of minimum to maximum capacitance will remain unchanged. In this case there is no simple way of defining a new ideal  $C$ - $V$  curve by excluding the effect of the work-function difference, because the charge distribution caused by the work-function difference is a somewhat complicated function of the ratio of oxide thickness to the Si Debye length.

(4) If interface states are present that do follow the a-c signal, they will displace and distort the  $C$ - $V$  curve as well as change the ratio of minimum to maximum capacitance. These changes depend on the energy and spatial distribution as well as response time of these states. No significant information concerning the density of states can be obtained in this case by the MOS capacitance method.

In cases (2) and (3), the difference between the measured and ideal  $C$ - $V$  curves,  $\Delta V$ , can be used to calculate the total charge in surface states,  $Q_{ss}$ , at any particular surface potential within a certain range. The density of ionized interface states, given as  $\text{cm}^{-2}$ , is determined from

$$N_{ss}(\psi_s) = \frac{Q_{ss}(\psi_s)}{q} = \frac{C_{ox}\Delta V(\psi_s)}{q}, \quad (5)$$

where  $q$  is the electronic charge. If the actual curve is to the left of the ideal one,  $Q_{ss}$  is positive, if it is to the right, then  $Q_{ss}$  is negative, corresponding to donor and acceptor states, respectively. If, in case (2), the corrected ideal curve is used, then the result corresponds to the actual density of the states. If the correction for the work-function difference is not made, or if case (3) prevails, then Equation (5) gives an "effective density of surface states."

In the discussion of MOS capacitance, loss mechanisms were neglected. Depending on the measuring frequency, the nature of the space-charge region, the response time distribution of the surface states, generation and recombination velocities of minority carriers, etc., various loss mechanism can be operative. The analysis of these processes is difficult and the equivalent circuit is complicated.<sup>22</sup>

### 2.3.3 MOS Measurements

Measurement methods used to investigate properties of MOS systems can be divided into two groups, depending on whether electrical or other physical properties are to be determined. For electrical properties, a variety of methods are available to determine the density, capture cross section, and response time of interface states, surface recombination, and surface mobility. The MOS capacitance method is by far the most frequently employed technique, especially for determining the interface state density. This is due to the simplicity of fabricating test structures, the ease of making (automatic) measurements,<sup>23</sup> and the large amount of information that can be obtained. Because of the usefulness of this technique its principles are described briefly below.

In the MOS capacitance method, a d-c electric field is applied between the metal electrode\* on the oxide and the silicon. For a given value of this field, a definite charge distribution in the MOS system will arise. The differential capacitance of the silicon space-charge region is then measured by superimposing a small a-c voltage on the d-c bias. The differential capacitance as a function of the applied d-c field (or bias) is now determined. This is conveniently done with an automatic display apparatus. The obtained capacitance versus voltage ( $C-V$ ) curve is then compared with a calculated ideal curve, i.e., one characteristic of a silicon-vacuum interface without surface states. If equilibrium has been established for each value of the d-c field, and the frequency of the small a-c signal is so high that no electronic transfer between various states outside the silicon (e.g., interface, oxide states) and the space-charge region can take place during one period, then this comparison gives the density of charges residing in these states as a function of the silicon surface potential, or, what is equivalent to this, as a function the energy position of these states in the Si forbidden band. The calculation of the ideal  $C-V$  curve under given conditions (type and resistivity of silicon, as well as oxide thickness) and the quantitative comparison with an actually measured one can be performed on a digital computer.<sup>23</sup>

The evaluation of the measurement gives the real or effective density of surface states as a function of silicon surface potential,  $\psi_s$ , or the energy position of the states,  $E$ , if the limitations of the method have been taken into account.<sup>15</sup> The density of states at flat-band condition (i.e.,  $\psi_s = 0$  or  $E = E_F$ ) is designated as  $N_o$  and is frequently

---

\* During our investigations vacuum-deposited aluminum dots or a mercury probe were used.

used to characterize the surface.\* The  $N$  versus  $E$  curves can be differentiated either graphically or by computer, and the obtained values of  $dN/dE$  in units of  $\text{cm}^{-2} (\text{eV})^{-1}$  are plotted against  $E$ . We have found that in the overwhelming majority of cases  $dN/dE$  was constant or zero. In these cases, the value of  $dN/dE$  and  $N_0$  were used to characterize the interface. This method is very sensitive (about  $10^{10}$  electronic charges  $\text{cm}^{-2}$ ) but nonspecific (similar, for example, to resistivity or Hall-effect measurements in semiconductors). Hence, one would like to combine these measurements with more specific ones such as optical absorption, electron spin resonance, radioactive tracer analysis, etc. Although some attempts have been made in this direction, it is very difficult to establish clear-cut correlation between these and charge-distribution measurements. The formidable difficulty is the required sensitivity, about  $10^{10}$  atoms (defect, impurities) per  $\text{cm}^2$ † Our conclusions concerning the behavior of the interface are, therefore, essentially based on indirect inference from investigations of the charge distribution as a function of various experimental variables, such as oxidation conditions, post-oxidation treatments, temperature of measurement, etc.

It is evident from the preceding discussions that the  $Q_{ss}$  determined by this method refers to the sum of all the charges outside the semiconductor except the charge on the metal electrode due to the applied bias. Depending on oxidation conditions, post-oxidation treatment, and deposition of metal electrodes, various charges at the interfaces and in the oxide contribute to  $Q_{ss}$ . A taxonomy of these charges has been given.<sup>24</sup> The separation of  $Q_{ss}$  into its individual components is impossible on the basis of high-frequency MOS capacitance measurements alone, even if the metal-silicon work-function difference can be properly taken into account. From this point of view, one should simply speak of an "effective compensating charge." Because of the widespread use of the term surface state, we have also retained this expression despite the conceptual difficulties connected with it. However, it is used in a completely operational manner.

### 3. PREPARATION OF $\text{SiO}_2$ FILMS

It has already been pointed out that the  $\text{SiO}_2$  film cannot be considered an ideal, charge-free substance. The nature and distribution of traps in the oxide (including the interface proper), i.e., its defect

\* Some authors use the symbol  $N_{FB}$  instead of  $N_0$ .

† In a very recent study of hydrogen and sodium ion transport in  $\text{SiO}_2$  films reported in the *IEEE Trans.*, Vol. ED-14, p. 749, Nov. 1967, S. R. Hofstein showed a substantial correlation between the results obtained by radioactive tracer analysis and those by  $C$ - $V$  measurements.

structure as well as other properties, are determined to a large extent by the preparation method. The problem is somewhat similar to that encountered in the growth of single crystals, where bulk properties are also very much affected by the growth conditions.

Silicon dioxide films can be either grown or deposited on a silicon substrate. In the first case, the silicon atoms in the oxide originate from the substrate, in the second case from an external source. The most commonly used growth method is thermal oxidation in either dry oxygen, wet oxygen, or steam. Anodic oxidation is also a potentially useful method, but it is not as common as thermal oxidation. Various vacuum and chemical deposition methods form the other group of preparation methods. From the standpoint of interface properties, grown films seem to be better and more reproducible than deposited films. Therefore, the emphasis is given here to grown SiO<sub>2</sub> films; deposition methods are only briefly mentioned.

The essential aim in oxide preparation, including subsequent treatments, is to produce films with the following characteristics:

- (1) Ideal interface (i.e., virtually zero surface-state density) or an interface characterized by an ideal  $C-V$  curve shifted along the voltage axis to a predetermined value.
- (2) Good insulating properties, i.e., low pin-hole density, high breakdown strength, high dielectric constant, and low dielectric losses.
- (3) Resistance to ambient stress, especially no ionic motion at elevated temperatures and/or high fields.
- (4) Resistance to attack by the atmosphere.
- (5) Resistance to radiation.

Characteristics (1) and (2) permit the fabrication of good devices. These characteristics require that from the standpoint of electrically active imperfections, the perfection of the oxide and interface must be controlled in a level of about  $10^{10}$  per cm<sup>2</sup> as an integrated density of electronic charges. This corresponds roughly to  $10^{-5}$  adsorbed monolayer, as reflected at the Si-SiO<sub>2</sub> interface.

Not all of the above characteristics have been achieved at the present state of technology, but significant progress has been made with items (1), (2), and (3) by using an SiO<sub>2</sub> film as the insulator.

### 3.1 Growth Methods

Both thermal and anodic oxidation of silicon generally results in non-crystalline oxide films that have a high degree of short-range order (in

the sense that no Si-Si bonds are formed). As a first approximation, the properties of these films are very similar to those of vitreous (fused) silica but the detailed defect structure and even the exact nature of the Si-O bonds depend on the oxidation conditions. These aspects are discussed in a subsequent section.

Since the molar volume of noncrystalline  $\text{SiO}_2$  is about 2.2 times larger than that of silicon, the oxide film is protective. This means that unless imperfections (crystallites, cracks, etc.) develop, the film hinders further oxidation. The mechanism and kinetics of oxide growth are determined by the manner in which the reacting species is supplied to the interface where the growth occurs and/or by the interface reaction. These, in turn, depend on the nature and concentration gradient of the diffusing species, electric field in the film (as due to adsorption of ions, space charges, and applied current or field), imperfections in the oxide, etc. The interaction among these factors determines the structure of the oxide and interface.

### 3.1.1 Thermal Oxidation

From the point of view of kinetics, thermal oxidation of silicon can be divided into three regimes. At room temperature a very thin ( $< 100 \text{ \AA}$ ) oxide film forms at a logarithmic rate,<sup>4</sup> at elevated temperatures (above  $\sim 700^\circ\text{C}$ ) oxidation begins at a linear rate but later becomes parabolic. A special case is oxidation in high-pressure steam ( $p \cong 25 \text{ atm}$  and  $T \cong 500^\circ\text{C}$ ), which always follows a linear law.<sup>25</sup> Most of the oxide films used in the manufacture of silicon devices are grown under conditions that correspond to the linear-parabolic regime. Here, the activation energy associated with the oxidation reaction and/or diffusion is supplied by thermal energy. If the oxidizing gas molecules are in an excited state or are ionized, then less thermal energy is needed, and oxidation may proceed at lower temperatures. This principle was applied by Ligenza<sup>26</sup> and later by Kraitchman<sup>27</sup> who oxidized silicon in a microwave-generated oxygen plasma with or without applying positive voltage to the specimen. The presence of oxygen ions in the oxidizing ambient and the possible high electric field across the oxide make this oxidation method somewhat similar to anodic oxidation. Because of various plasma phenomena the oxidation mechanism is not well understood, but the method is potentially useful, especially when the high temperature required for thermal oxidation is not desirable. This point is discussed later when the various methods are compared.

Thermal oxidation of silicon in dry and wet oxygen, as well as in water vapor, at or below atmospheric pressure has been extensively

studied. Several recent papers summarize the earlier works.<sup>1,28-31</sup> A common feature of these investigations is that the oxidation was done in resistance-heated silica tubes. There are several disadvantages associated with this method, the most important being the easy diffusion of sodium and water through the wall of the silica tube at the oxidation temperature. Thus, for instance, Nakayama and Collins measured non-negligible oxidation rates in dry argon, implying the presence of 0.35 torr water vapor at 1000°C.<sup>29</sup> An additional difficulty is the presence of various adsorbates (usually unknown and not well controlled) on the silicon specimen at the beginning of the oxidation. These difficulties were eliminated by using radio-frequency heating and a gas-cooled reaction vessel, as well as by incorporating an *in situ* pre-oxidation cleaning in hydrogen at high temperature.<sup>8</sup> The results indicate that impurities, especially water and sodium, play an important role in the kinetics and mechanism of oxidation of silicon. This was further demonstrated by purposely introducing sodium into the growing oxide both in oxygen and water-vapor ambients.<sup>32,33</sup>

The results of oxidation in oxygen under very clean conditions, as well as those obtained by doping, were analyzed on the basis of a combined linear-parabolic relationship;

$$\frac{t}{x - x_0} = A + B(x - x_0) \quad (6)$$

where  $t$  is the oxidation time,  $x$  the oxide thickness,  $x_0$  the oxide thickness at the beginning of the oxidation (generally less than 40 Å), and  $A$  and  $B$  are constants.

Such a relationship, occasionally in a somewhat modified form, was used by several workers,<sup>28-31,34,35</sup> and proved to be generally valid. At the beginning of the oxidation, the oxidation rate,  $dx/dt$ , is given as

$$\frac{dx}{dt} = \frac{1}{A} \quad \text{for } x = x_0 \quad (7)$$

which corresponds to a linear rate:

$$x = at + x_0 \quad (8)$$

For large oxide thicknesses, i.e., for  $x \gg A/(2B)$ ,

$$\frac{dx}{dt} = \frac{1}{2Bx} \quad (9)$$

which corresponds to a parabolic rate;

$$x^2 = \frac{t}{B} \equiv kt. \quad (10)$$

It was found that both  $\log a$  and  $\log k$  depend linearly on  $1/T$  with a high level of confidence. Hence,

$$a = a_0 \exp \left\{ -\frac{Q_a}{kT} \right\} \quad (11)$$

and

$$k = k_0 \exp \left\{ -\frac{Q_k}{kT} \right\}. \quad (12)$$

Typical results of Revesz and Evans, as well as of others, for these constants are tabulated in Table I. It is evident from this table that the predominant effect of sodium is an increase of  $k_0$  and  $Q_k$ , especially when water vapor is present. The presence of water vapor increases  $a_0$  and  $Q_a$ , but decreases  $k_0$  and  $Q_k$  somewhat (at least at atmospheric pressure). It is also clear that oxidation in resistance-heated tubes proceeds in an ambient that is contaminated by sodium and/or water in varying degree. The influence of water vapor on the linear-rate constant is illustrated by the great difference between the results with r-f and resistance heating techniques for oxidation in oxygen. This indicates that water diffuses through the wall of the tube if it is hot. As far as the parabolic regime is concerned, the interaction between sodium and water makes the situation quite complicated.

Before formulating a definitive theory of thermal oxidation of silicon, the role of impurities should be investigated in more detail. Nevertheless, the following model can explain most observations. When the oxide thickness is much less than  $A/2B$ , surface reaction is the rate-determining step. This gives rise to a linear-rate law. Since it was found experimentally<sup>25,30,36</sup> that the linear rate depends on the crystallographic orientation of the silicon substrate, the oxidation reaction must occur at the Si-SiO<sub>2</sub> interface. As has been pointed out,<sup>18,35</sup> the migrating species is most probably interstitial oxygen or, if water vapor is present in the oxidizing ambient, interstitial H<sub>2</sub>O and/or OH groups attached to silicon. Since  $a_0$  is larger for water than for oxygen, the density of Si-OH groups or interstitial water is probably larger than that of interstitial oxygen. In the presence of sodium, the en-

Table I—Linear and Parabolic Rate Constants for Thermal Oxidation of Silicon.

Reference	Oxidation ambient	Pressure	Tube	Heating	$u_o$ cm/sec	$Q_n$ eV	$k_o$ cm <sup>2</sup> /sec	$Q_t$ eV
32, 33	O <sub>2</sub>	1 atm	SiO <sub>2</sub>	rf	$1.7 \times 10^{-4}$	1.12	$1.2 \times 10^{-9}$	1.20
33	O <sub>2</sub> + Na	1 atm	SiO <sub>2</sub>	rf	$1.1 \times 10^{-3}$	1.25	$6.0 \times 10^{-9}$	1.32
32, 33	O <sub>2</sub>	1 atm	Al <sub>2</sub> O <sub>3</sub>	res.	$1.7 \times 10^{-1}$	1.88	$1.1 \times 10^{-9}$	1.22
36	O <sub>2</sub>	1 atm	SiO <sub>2</sub>	res.	$2.4 \times 10^{-1}$	2.00	$2.1 \times 10^{-9}$	1.24
36	O <sub>2</sub> - H <sub>2</sub> O	$p_{O_2} = 415$ Torr $p_{H_2O} = 345$ Torr	SiO <sub>2</sub>	res.	2.1	1.96	$5.6 \times 10^{-10}$	0.72
29	Ar - H <sub>2</sub> O	$p_{H_2O} = 4.6$ Torr	SiO <sub>2</sub>	res.	1.7	2.2*	$2.1 \times 10^{-10}$	1.16
30	H <sub>2</sub> O	1 atm	SiO <sub>2</sub>	res.	$7.8 \times 10^2$	2.60	$5.6 \times 10^{-10}$	0.70
33	He - H <sub>2</sub> O + Na	$p_{H_2O} = 24$ Torr	Al <sub>2</sub> O <sub>3</sub>	res.	$4 \times 10^{-3}$	1.27	$4.4 \times 10^{-4}$	2.3

\* Determined from Figure 4 of Reference (29). The authors' value for  $Q_n$ , 3.3 eV, is probably in error.

trance of water, or rather hydroxyl, is facilitated through autocatalytic reactions known from the corrosion of glasses.<sup>37</sup>

As previously mentioned, the linear rate constants depend on the silicon orientation. The order of increasing oxidation rates in the range of 500° to 1200°C is (100), (110), and (111). Ligenza attributed this orientation dependence to the mechanism by which the silicon bonds form an activated complex with the reacting water molecule.<sup>25b</sup> A somewhat similar explanation can be given by considering the strain

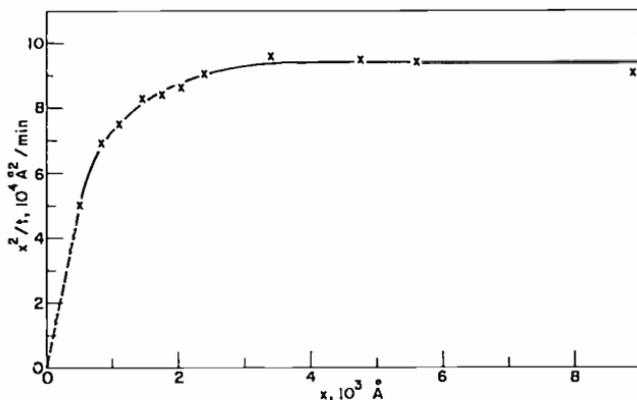


Fig. 3—Transition from linear to parabolic oxide growth. The ratio  $x^2/t$  is shown as a function of oxide thickness,  $x$ ;  $t$  is the oxidation time. The slope of the curve extrapolated to zero thickness gives the linear growth rate at  $x = 0$ . With increasing oxide thickness  $x^2/t$  approaches a constant value characteristic of the parabolic regime of growth. The oxidation was done in dry  $O_2$  at 1250°C in a resistance-heated alumina tube.

energies of various Si-O complexes formed at the surface. For the analog Ge-O system these values also increase in the order (100), (110), and (111).<sup>38</sup> From studies on interface effects we have concluded that the Si-SiO<sub>2</sub> interface has indeed a greater tendency toward order if it is formed on (100) rather than on (111) surface. This is discussed in the next Section.

The transition from linear to parabolic rate generally extends to 1000 to 2000 Å, as shown in Figure 3. This unusually large thickness range may be due to the relatively easy migration of oxygen along structural channels in the oxide; this is discussed in Section 5.

Parabolic growth in dry  $O_2$  shows a remarkable similarity to oxygen permeation through fused silica, as pointed out by Motzfeldt.<sup>39</sup> Both processes depend linearly on pressure, indicating that  $O_2$  is the migrating species. On the other hand, both can be influenced by an externally

imposed electric current,<sup>40,41</sup> showing that a significant portion of the flux is due to ionic migration. It was proposed that permeation and oxide growth take place via an interstitially diffusion mechanism; the total flux is the sum of molecular and ionic fluxes.<sup>39</sup> The oxygen interstitial molecules are in a constant exchange process with the oxide, resulting in atomic oxygen interstitials that can dissociate into negative ions and holes. Assuming that at a given temperature the concentration of holes originating from this process is much larger than the intrinsic hole concentration, and that the diffusion coefficient is much larger for holes than for ions, the ratio of the concentrations of ionic and molecular oxygen at 1000°C was estimated as about  $2 \times 10^6$  (using  $4.3 \times 10^6 \exp\{-2.44/kT\}$  cm<sup>2</sup>/sec as the coefficient of oxygen self-diffusion<sup>42</sup>). Because of this high ratio the ionic flux predominates, but it is the molecular form that reacts with silicon at the oxide-silicon interface.

Incorporation of sodium results in an increased density of broken Si-O-Si bridges and, hence,  $k_n$  increases; but the presence of sodium hinders the motion of oxygen and, hence  $Q_k$  increases.

If water vapor is present, a large variety of interactions can occur between H<sub>2</sub>O and SiO<sub>2</sub>, leading to Si-OH, interstitial H<sub>2</sub>, and H<sub>2</sub>O, etc.,<sup>43</sup> and the transport mechanism is quite complex. Thus, it was found for fused SiO<sub>2</sub> that the diffusion coefficient increases with concentration, and the activation energy varies from 0.44 to 0.87 eV in the temperature range of 800° to 1050°C.<sup>44</sup> Qualitatively, with increasing water vapor pressure,  $k_o$  increases and  $Q_k$  decreases due to the increased density of broken Si-O-Si bridges. In the presence of sodium the incorporation of OH is easier, hence the increase in  $k_o$ ; however, the motion of OH is hindered, and hence the increase in  $Q_k$ .

Another aspect of impurity effects in thermal oxidation of silicon is the redistribution of group III and V elements used for doping of silicon. Thus, phosphorus is rejected by the growing oxide and accumulates in the silicon, whereas boron is accumulated in the oxide and depleted in the silicon.<sup>45</sup> If the doping level in silicon is very high ( $> 10^{20}$  cm<sup>-3</sup>) the oxidation kinetics may also be affected.<sup>25,46</sup>

During oxidation, oxygen may be dissolved in silicon. The solubility at 1000°C is about  $2 \times 10^{17}$  atoms/cm<sup>3</sup>. The dissolved oxygen can result in the formation of donor levels, especially after a heat treatment at 300° to 500°C. However, if gold is present in the silicon, charge relations at the interface become complicated.<sup>47</sup>

Interaction between water vapor in the oxidizing ambient and mechanical damage in silicon can lead to line defects (probably shallow stacking faults) in silicon.<sup>48</sup>

### 3.1.2 Anodic Oxidation

Oxidation of silicon can also be performed electrolytically by applying a sufficient overvoltage in an electrolyte that does not dissolve the oxide. The migration of ionic defects in the oxide and, hence, the oxide growth is due to the high electric field across the film.

Anodization of silicon in a nonaqueous electrolyte was investigated by Schmidt and Michel.<sup>49</sup> The same method was used by Revesz, who studied the properties of the Si-SiO<sub>2</sub> interface.<sup>10</sup> In the course of his studies, the following observations on the anodization mechanism have been made.

During constant-current oxidation, the field across the oxide is constant,  $1.2 \times 10^7$  to  $1.9 \times 10^7$  volts/cm, depending on current density and the orientation of silicon; thus the oxidation rate is about 6 Å/V. The most interesting feature of anodic oxidation of silicon is the very low oxidation efficiency, 0.4 to 1.6%, corresponding to a current that is overwhelmingly electronic rather than ionic. This is quite in contrast to the anodization behavior of, say, tantalum and aluminum. The oxidation rate and efficiency show definite and parallel trends; they increase with increasing oxide thickness and current density.

Under constant-voltage anodization, the oxide thickness and current density vary with the logarithm and reciprocal of time, respectively. The oxidation rate and efficiency decrease with time and thickness. This decrease could be due to decreasing current density and/or structural changes in the oxide. It has been proposed that the current is essentially due to injection by tunneling of electrons from the electrolyte into the oxide conduction band.<sup>50</sup> Since the anodization current is predominantly electronic, its decrease cannot be caused by the decreasing field alone (as is the case with, e.g., aluminum<sup>51</sup>); trapping of electrons (holes) and/or rearrangement of the oxide structure probably play important roles. Indeed, it was found that electronic conductance of SiO<sub>2</sub> films can be greatly reduced by constant-voltage anodization without increasing the oxide thickness.<sup>52,53</sup>

Oxidation in oxygen plasma with d-c bias is similar to anodization in an electrolyte but, as previously mentioned, various complications arise.

The main advantage of anodic as opposed to thermal oxidation is that it can be performed at much lower temperatures. Hence processes that deteriorate the minority carrier lifetime, can be avoided. Also, unwanted changes in the doping profile within silicon can be minimized. The disadvantages of anodic oxidation are limited thickness

(2000-3000 Å) because of breakdown effects, possible incorporation of impurities of the electrolyte, and usually somewhat higher density of interface states than that obtained by the best thermal oxidation.

### 3.2 Deposition Methods

Silicon dioxide can be deposited by evaporation in vacuum, by sputtering, and by various vapor-phase chemical reactions. Thus, for instance, SiO<sub>2</sub> films were prepared by heating an SiO<sub>2</sub> source with an electron beam in vacuum.<sup>54</sup> An example of sputtering is in the use of a silicon cathode in a d-c plasma of about  $2 \times 10^{-2}$  Torr of oxygen.<sup>55</sup> In these processes the silicon substrate is at low temperature, say, below 100°C.

The simplest method of vapor-phase deposition is thermal decomposition of a silicon-organic compound, e.g., ethyltriethoxy-silane, C<sub>2</sub>H<sub>5</sub>Si(C<sub>2</sub>H<sub>5</sub>)<sub>3</sub>, at 650° to 750°C.<sup>56</sup> By using an rf-induced argon plasma, even unheated substrates could be coated.<sup>57</sup> Oxidation of silane, SiH<sub>4</sub>, by oxygen at about 300°C<sup>58</sup> or reaction of silicon halogen compounds, e.g., SiCl<sub>4</sub>, with water vapor at about 1100-1350°C<sup>59</sup> also result in an SiO<sub>2</sub> film. In the latter case the water vapor is obtained by a reaction between CO<sub>2</sub> and H<sub>2</sub>.

The advantages of the deposition methods are low temperature (except for the pyro-hydrolytic reaction) and usually high deposition rates. However, unless the deposition temperature is high or post-deposition heat treatment is applied, these films are generally porous and the interface properties are inferior to those obtained by the growth methods. The structure of evaporated and sputtered oxides is generally not characterized by a high degree of short-range order, and in many cases their composition can be given as SiO<sub>x</sub>, where, depending on the deposition conditions,  $x$  varies from one to two.

## 4. Si-SiO<sub>2</sub> INTERFACE PROPERTIES

In this section the influence of oxidation conditions, especially the oxidation rate at the end of the oxidation, on the effective surface state density is described. The description is followed by a presentation of the results of high-temperature annealing in helium and low-temperature heat treatment in hydrogen. A correlation between the density of interface states and interface disorder is pointed out. Ion migration in the oxide at elevated temperature under high field as well as electron (hole) redistribution between silicon and the interface region in the oxide may affect the silicon surface potential and, hence,

the MOS capacitance. These effects are also discussed. Finally, the various interface phenomena are summarized and some conclusions are drawn.

#### 4.1 Influence of Oxidation Conditions and Post-Oxidation Treatments

It is well known that the properties of single crystals are very much affected by the growth conditions, such as growth rate, impurities in the raw material, growth ambient, etc. It was expected that the properties of grown  $\text{SiO}_2$  films and the Si-SiO<sub>2</sub> interface would also be influenced by similar factors. It was also expected that various defect gradients and non-equilibrium distribution of electric charges associated with the oxide growth as well as the structure of the interface could be modified by high-temperature treatment in a neutral ambient (annealing).

It was recognized early in the study of MOS devices that the properties of the Si-SiO<sub>2</sub> interface are affected by oxidation conditions (e.g., use of dry oxygen or wet oxygen as the oxidizing ambient) as well as post-oxidation treatments (e.g., heating in hydrogen or wet nitrogen). The influence of oxidation rate and temperature in dry-oxygen oxidation as well as various parameters (temperature, time, contamination level) of post-oxidation annealing in helium was systematically studied by the authors;<sup>7,8,60</sup> the following discussion of interface properties is based on this work.

##### 4.1.1 Oxidation

The studies of the influence of the oxidation rate on the interface properties were done predominantly on thermally oxidized specimens. The oxidation was stopped suddenly (quenched) at a predetermined time (corresponding to a given value of thickness at a given temperature) by removing the specimen from the hot zone of the resistance-heated furnace or, in the case of oxidation by r-f heating, by turning off the r-f supply. The final oxidation rate, that is, the rate at which the Si-SiO<sub>2</sub> interface moved at the end of the oxidation, was determined from the appropriate thickness versus oxidation-time graphs.

We have found that thermal oxidation results in predominantly donor-type surface states; these states fall into two classes depending on whether they are or are not affected by the oxidation rate. For these states, both  $N_o$  and  $dN/dE$  increase with increasing final oxidation rate, and their values increase as the temperature decreases.

Qualitatively similar results were obtained for anodic oxidation in

0.04 M solution of KNO<sub>3</sub> in N-methylacetamide.<sup>10</sup> Constant-current anodization, corresponding to oxidation rates of 30 to 70 Å/minute, resulted in a high density of interface states. With constant-voltage anodization, the interface density usually decreased with the anodization time, i.e., with final oxidation rate.

Since faster oxidation would be expected to result in a more disordered interface structure, the oxidation-rate-dependent states are most probably associated with the phase boundary and, thus, belong to the group of interface states. The annealing experiments to be described support this conclusion and indicate, in addition, that some of the rate-independent states are also interface states, but can be eliminated by proper methods. Although two radically different methods of oxidation were used, the influence of the oxidation rate is qualitatively very similar. The crystallographic orientation of the silicon substrate was found to have the following influence. For identical oxidation conditions, the (100) orientation resulted in considerably lower values of both  $N_o$  and  $dN/dE$ ; the latter generally approached zero.

All these results indicate that a more disordered interface, which corresponds to higher oxidation rate at a given temperature or to lower oxidation temperature for a given oxidation rate, is associated with a higher density of imperfections. This is manifested in a higher density of surface states. This relationship between growth rate, temperature, and disorder is well known from the field of crystal growth.

The crystallographic orientation of the silicon substrate influences the effect of oxidation rate dependence on surface-state density; this influence can be attributed to a greater tendency to order in the interface region of the (100) silicon than of the (111). This can be seen from the geometry of the surfaces, and from the values of strain energies calculated by Green and Lieberman for the similar Ge-O system: 0.1 or 0.8 eV for (100) and 1.4 or 3.6 eV for (111) orientation, depending on the exact nature of the configuration of the surface complex.<sup>38</sup> Also, it is interesting to note that the generally observed orientation dependence of the surface-state density<sup>61-64</sup> is the same as that of the liner oxidation rate (in the surface-controlled regime of growth), i.e., density and oxidation rate increase in the order (100), (110), (111). The influence of orientation on the oxidation rate was attributed to steric effects at the Si-SiO<sub>2</sub> interface as determined by the density of Si bonds available for the reaction there. This was discussed in Section 3.

### 4.1.2 Annealing

We have mostly studied the effects of high-temperature ( $> 800^\circ\text{C}$ ) heat treatment in helium, the so-called annealing. This was performed as follows. Specimens oxidized in the  $\text{Al}_2\text{O}_3$  tube furnace were either removed from the hot zone (without opening the system) until the oxygen was essentially replaced by helium and then were re-inserted (i.e., their temperature was low during the transition from  $\text{O}_2$  to He) or they were left in the hot zone during the transition from  $\text{O}_2$  to He. Occasionally, they were exposed to room ambient between oxidation and annealing. Specimens that were processed in the rf-heated reactor were annealed by changing the ambient from  $\text{O}_2$  to He and keeping the temperature at the same value as for oxidation for 15 minutes followed by cooling to  $700^\circ\text{C}$  at a rate of approximately  $50^\circ\text{C}/\text{minute}$ ; at this temperature the heat was turned off. Anodically oxidized specimens were annealed in the rf-heated tube after careful rinsing.

Annealing in helium reduces the surface-state density. The initial annealing phase is characterized by a logarithmic decrease of  $N_s$  and  $dN/dE$  with time. Similar time dependence is frequently observed with chemisorption, and is attributed to a rearrangement of the surface structure as caused by chemisorption,<sup>65</sup> or to diffusion of electrons through a surface barrier.<sup>66</sup> From this similarity between the effect of annealing and the effect of chemisorption, one would assume that the basic phenomena involved in annealing occur in the vicinity of the interface. Indeed, we have found, by stepwise dissolution of the oxide, that the charge is essentially within a  $100 \text{ \AA}$  region in the oxide at the Si-SiO<sub>2</sub> interface. Studies on slow-trapping type instability (see below) indicated that the states involved in this process are within approximately  $2.5 \text{ \AA}$  of the interface in the oxide.<sup>67</sup> The logarithmic regime is followed by a complex behavior that is attributed to water and other contaminants present in the ambient and/or at the oxide surface at the beginning of annealing.  $t_a$ ,  $t_\beta$  and the corresponding values of  $N_s$ , as well as the slopes in the three regions indicated in Figure 4 can differ substantially, depending on the contamination level, especially that due to water, and the temperature during the transition from oxidation to annealing. Again, changes in  $N_s$  are, in general, coupled with similar changes in  $dN/dE$ . Both  $N_s$  and  $dN/dE$  decrease faster with annealing time at a higher temperature. As a matter of fact, at higher temperatures  $dN/dE$  is zero at the end of the exponential decrease. The annealing behavior of the anodically oxidized specimens was quite similar to that described above.

By combining the pre-oxidation cleaning in hydrogen, thermal oxi-

dation in dry oxygen, and annealing in helium into an *in situ* process and using r-f heating, essentially zero surface-state density could be achieved for both (111) and (100) orientations.\* However, the (111) specimens were more sensitive to contamination and annealing conditions than the (100) specimens. The increased sensitivity of a more disordered interface points out the possible importance of cooperative

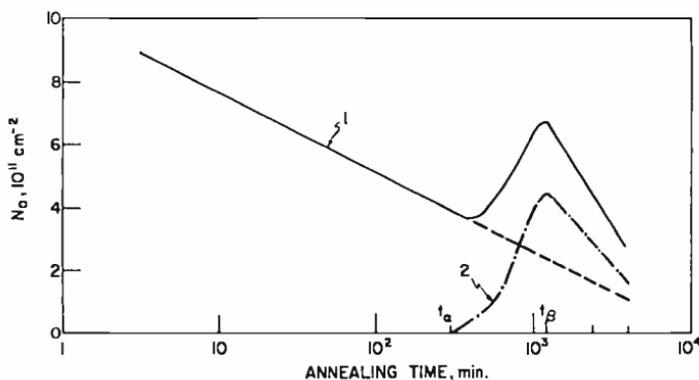


Fig. 4—Density of ionized surface states at zero surface potential versus annealing time. Curve (1) represents typical experimental results obtained by thermal oxidation in a resistance heated Al<sub>2</sub>O<sub>3</sub> tube, followed by annealing in helium in the same furnace. The rise of curve (1) between  $t_a$  and  $t_\beta$  indicates that some contamination occurred during annealing. Under very clean conditions, e.g., using the *in situ* method with r-f heating, curve (1) would be a straight line that reaches a  $N_s$  value of about  $10^{10}\text{cm}^{-2}$  in about 15 minutes at  $1100^\circ\text{C}$ . Curve (1) in this figure can be broken up into the extrapolation of the first portion ( $t < t_a$ ) of curve (1) and curve (2). The latter represents contamination effects, probably migration of hydroxyl and/or sodium in the oxide. Both  $t_a$  and  $t_\beta$  represent transition periods rather than a well-defined time. The extent of these periods is a function of the experimental conditions.

phenomena at the interface. The high perfection of a properly processed interface is also revealed by the observations that the  $C$ - $V$  curves were independent of frequency between 10 kHz and 10 MHz. The temperature dependence of the MOS capacitance was equivalent to that of an ideal space-charge region in silicon, and the interface and slow trapping instability (see Section 4.2) was practically nil.

\* Comparable results were obtained by Heiman by oxidizing (111) and (100) silicon films grown on sapphire substrates.<sup>68</sup> On the other hand, Tarui et al reported lower values of  $dN/dE$  for oxides deposited by hydrolysis of SiCl<sub>4</sub> at  $1100^\circ\text{C}$  than for thermally grown oxides; they attributed the difference to the oxide preparation.<sup>69</sup> However, thermal oxidation was done on chemically cleaned specimens, whereas the oxide deposition was performed *in situ* after epitaxial growth or high temperature etch in HCl. These observations emphasize the importance of cleaning the silicon prior to oxidation.

### 4.1.3 Post-Oxidation Heat Treatment in $H_2$

Since others have found that post-oxidation treatment in  $H_2$  at not too high a temperature ( $< 600^\circ C$ ) can also result in a reduction of the surface-state density,<sup>70</sup> we made some experiments with this treatment in order to compare it with high-temperature annealing. The  $H_2$  treatment was done at  $500^\circ C$  for 15 minutes in a conventional resistance-heated silica tube or in the rf-heated reaction vessel.

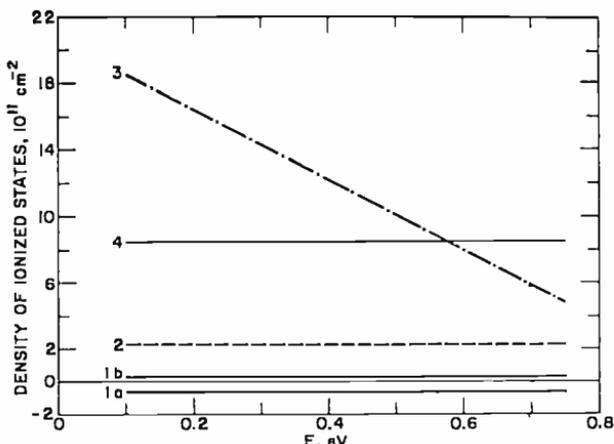


Fig. 5—Summary of influence of oxidation, annealing in He, and  $H_2$ -treatment on surface-state density. The energy,  $E$ , is given with respect to the Si valence band. The specimens were  $\sim 10$  ohm-cm, p-type. They were oxidized by the r-f method, mostly after *in situ* pre-oxidation  $H_2$  cleaning. Curves (1a) and (1b) are for (111) orientation, annealed; they indicate a spread could be reduced by about  $10^{11} \text{ cm}^{-2}$  (for 1000 Å thick oxide) due to without taking into account the metal-silicon work-function difference, this spread could be reduced by about  $10^{11} \text{ cm}^{-2}$  (for 1000 Å thick oxide) due to the difference in the work functions of aluminum and mercury electrodes used in these investigations. Similar curves were obtained for (100) orientation. These curves remained the same after post-oxidation  $H_2$  treatment. Curve (2) is for (100) orientation with quenching; this specimen was not affected by post-oxidation  $H_2$  treatment. Curve (3) is for (111) orientation with quenching; a similar curve was obtained from annealed (111) but without pre-oxidation cleaning in  $H_2$ . For curve (4), the conditions are as for Curve (3), but after post-oxidation  $H_2$  treatment; the (111) annealed specimen without pre-oxidation  $H_2$  cleaning behaved similarly.

If oxidized specimens are heated in hydrogen at  $300\text{-}500^\circ C$  the value of  $dN/dE$  is reduced to essentially zero. Well-annealed specimens are hardly affected by such a treatment. Since this reaction is quite fast and occurs at a relatively low temperature, it is reasonable to assume that it is an interface reaction (e.g., saturation of unsaturated bonds<sup>70</sup>) that leads to the disappearance of states within the forbidden band. If the oxidation is performed in the presence of water vapor, then the reacting hydrogen can originate within the oxide itself by

several reactions between the various forms of hydrogen in the Si-O network.<sup>43</sup> This might be one of the causes of the usually observed changes in surface state density after heating steam-oxidized silicon in neutral or other ambients.

The post-oxidation H<sub>2</sub> treatment, as described above, should be distinguished from an H<sub>2</sub> treatment that introduces donor surface states,<sup>71</sup> probably by a modification of the defect structure of the oxide. It is known that at elevated temperatures, hydrogen reacts with SiO<sub>2</sub> to form, among others, Si-OH groups and trivalent Si.<sup>72</sup> This trivalent Si may act as a donor center (see Section 5).

The effects of annealing and low-temperature H<sub>2</sub> treatment on the surface-state density are compared in Figure 5 for r-f thermal oxidation. It is evident that high-temperature He annealing is superior to the H<sub>2</sub> treatment. Further indication of this superiority is that the H<sub>2</sub> treatment does not remove the slow-trapping-type instability while He annealing does.

## 4.2 Stability Behavior

From both theoretical and practical viewpoints it is very important to investigate the stability of the charge distribution, as established by the oxidation and post-oxidation treatments, as well as the deposition of the metal electrode. In the theoretical realm, these studies reveal very interesting ionic and electronic transport processes in the oxide and/or at the silicon-oxide interface. Practically, the devices whose characteristics are determined or influenced by the Si-SiO<sub>2</sub> interface must be as stable as possible under extreme working conditions, that is, elevated temperature (up to about 300°C) and high electric field across the oxide of the order of 10<sup>6</sup> volts/cm. Under these conditions the following phenomena can occur: electron (hole) redistribution between silicon and the interface region in the oxide (including interface states), migration of ions in the oxide, and interaction between migrating ions and the interface region. The migrating ions may originate from the oxide itself or from contamination that occurred during various post-oxidation steps, especially the deposition of the metal electrode. These so-called instability phenomena affect the silicon surface potential and, hence, can be revealed by changes in the MOS capacitance curves. In this sense, the space-charge region in silicon serves as a detector to study transport processes in the oxide and at the interface with a sensitivity that was hitherto impossible to achieve (about 10<sup>10</sup> electronic charges/cm<sup>2</sup>).

The various instability phenomena observed in MOS structures whose oxide films and/or metal electrodes were produced under medium

or high contamination conditions can be summarized as follows. Application of positive voltage at elevated temperatures (150 to 300°C) to the metal electrode\* generally leads to a shift of the  $C-V$  curve, as measured at room temperature, toward more negative electrode (gate) voltages.<sup>73</sup> This behavior is attributed to the migration of sodium ions from the  $M-SiO_2$  to the  $SiO_2-Si$  interface<sup>74</sup> and/or to the migration of protons.<sup>75</sup> Since the trapping and/or generation of ions occurs at the interface, both processes give rise to a voltage-asymmetric behavior, i.e., at not too high a temperature (e.g., 150°C) positive gate bias has a large effect whereas a negative bias has hardly any. It was generally found that the extent of this instability can be reduced by etching off a thin layer of the oxide before electrode evaporation. This indicates that the sodium that was incorporated into the oxide during its growth accumulated at the outer surface. The behavior of this sodium is similar to that of the sodium introduced artificially into the oxide before electrode evaporation.<sup>74</sup> If the instability is due to the simple migration of, say, sodium ions within the oxide without trapping and/or generation phenomena at the interfaces, then the effect of bias is symmetric. Such behavior was observed if glass was substituted for the  $SiO_2$  film.<sup>76</sup> It is interesting to note that there is a great difference between the asymmetric sodium-based instability and the symmetric one, despite the fact that the migrating species is the same. In the case of the voltage-asymmetric instability, the situation is further complicated because both sodium and hydrogen ions can move in a slow and fast form.<sup>67</sup> Evidently, from the standpoint of its interactions with the oxide and/or interfaces, the behavior of sodium (or an other similar impurity) is quite complex. This behavior may be related to the manner in which the sodium was introduced into the oxide (cf. distinction between permanent and metastable hydroxyl in  $SiO_2$  glass, the former incorporated during melting, the latter introduced later, e.g., by a reaction with  $H_2$ ,<sup>43</sup>). In the usual MOS structures, the voltage-symmetric instability is generally not observed. Under negative bias at relatively high temperature (300°C), the  $C-V$  curve shifts again to more negative rather than positive gate voltages.<sup>67,68</sup> This is attributed to electronic charge transfer between silicon and interface states, and is called the slow-trapping instability.<sup>67†</sup>

In addition to the ion drift and slow-trapping instabilities, a third form of instability may be present. This instability is revealed, by

---

\* Generally, the application of a voltage of either polarity at elevated temperature to the metal electrode and subsequent  $C-V$  measurement at room temperature is called the BT (bias-temperature) test.

† In some previous papers this has been called the interface instability.

positive bias temperature treatment, as a distortion of the  $C$ - $V$  curve and a change in the ratio of minimum to maximum capacitance owing to the introduction of fast states that follow the a-c signal. This is called the surface-state instability.

Our results showed that, even though the ion-drift instability is caused primarily by the lack of sufficient cleanliness during the vacuum

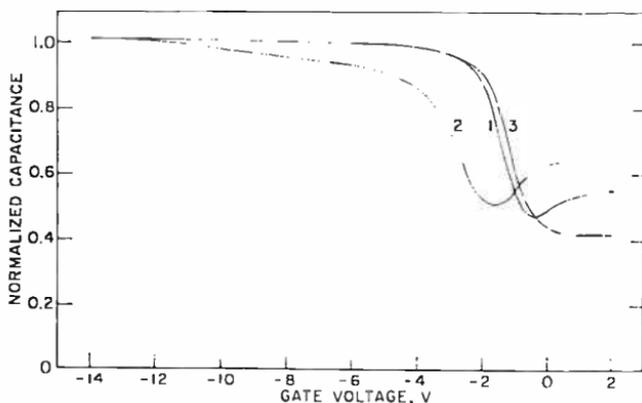


Fig. 6—Effects of bias-temperature tests on MOS capacitance. P-type silicon of 10 ohm-cm resistivity was oxidized under medium contamination level and annealed in helium at 1100°C. The oxide thickness is 1900 Å. The electrode is vacuum deposited aluminum. The measurements were done at room temperature. Curve (1) was taken after electrode deposition. It corresponds to  $dN/dE = 0$  and  $N_s = 1.7 \times 10^{11} \text{ cm}^{-2}$  or, if corrected for work-function difference,  $N_{s, \text{corr}} = 0.8 \times 10^{11} \text{ cm}^{-2}$ . Curve (2) shows that significant shift and distortion of the  $C$ - $V$  curve occur when a +10 volts bias is applied at 300°C for 1 minute. Curve (3) demonstrates that after -40 volts is applied at 300°C until saturation is reached, the density of the effective ionized donor states is decreased by about  $2 \times 10^{10} \text{ cm}^{-2}$ .

deposition of contacts to the oxide, contamination (most probably sodium and/or hydrogen (water)) during oxidation may lead to ion-drift and/or surface instabilities (see Figure 6). No slow-trapping instability was found, but the  $C$ - $V$  curve shifted slightly towards a more positive voltage after a negative bias was applied. Specimens oxidized anodically with constant current showed similar behavior.

Reduction of the contamination led to great decrease of the positive-bias instability. This was achieved mainly by the *in situ* process mentioned above, and by extreme care in deposition of contacts. Only a small ( $< 0.2$  volt for a 1000 Å oxide) and parallel shift of the  $C$ - $V$  curve was observed. This means that some migration of positive ions occurred during the bias-temperature test but they did not interact with the Si-SiO<sub>2</sub> interface. This shift is equivalent to  $4 \times 10^{10}$  elec-

tronic charges/cm<sup>2</sup> or less as reflected to the interface. The lack of interaction between migrating ions and the interface again points out the importance of cooperative processes at the interface: if the interface is nearly perfect it is difficult to introduce disorder. Since the extent of this shift is approximately the same as that for negative bias, it is possible that the instability is due to ion migration of the symmetric type. This conclusion is further substantiated by the observation that etching the oxide before electrode evaporation did not affect the positive bias instability. Evidently the impurity responsible for this instability (most likely sodium, but protons may be involved too) is associated with the oxide in a manner that resembles sodium in glass. The reported asymmetric distribution of sodium in SiO<sub>2</sub> films grown in resistance-heated silica tubes<sup>78</sup> does not provide an explanation for these observations, unless we assume that this sodium is in a virtually immobile form. Indeed, the presence in SiO<sub>2</sub> of inactive sodium was shown by radioactive tracer methods.<sup>78</sup> Another possibility is that the positive shift of the *C-V* curve under negative bias-temperature test is due to negative ions or polarization of the oxide. In the latter case it is conceivable that inactive sodium and/or hydrogen ions change their positions under the electric field even though they do not migrate. This process is similar to dielectric losses in glasses that are caused by small displacements of alkali ions.<sup>79</sup> It is interesting to note that the concept of bound and unbound sodium in glass was used by several workers in describing d-c and a-c conduction processes in glasses. It is also known from studies on fused silica that its dielectric properties depend strongly on water (proton, hydroxyl) as an impurity, whereas protons generally do not contribute to the d-c conductance unless other impurities (e.g., aluminum) are present. The nature of this small positive shift is not yet completely understood.

Very similar stability behavior was observed with SiO<sub>2</sub> films grown with constant-voltage anodization. The accumulation of "unbound" sodium (or similar) ions at the outer surface of the oxide film is prevented during anodic oxidation because the high field ( $\sim 2 \times 10^7$  V/cm) across the oxide drives them out. However, even the high field cannot prevent the incorporation of some impurities into the oxide, but these are now in a "bound" form. Again, the migration of hydrogen ions cannot be completely disregarded.

It should be pointed out that the positive shift under negative bias-temperature test could not be observed before the successful elimination of the slow trapping instability. The crucial step is careful annealing at high temperature. The importance of this step can be seen from the fact that in its absence the slow trapping instability is gen-

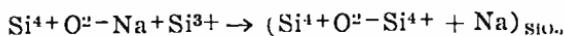
erally large even though (1) the oxidation resulted in essentially zero  $dN/dE$ , (2) some annealing in various ambients (N<sub>2</sub>, Ar, O<sub>2</sub>, wet O<sub>2</sub>) was performed, or (3) the specimen was removed from the oxidizing furnace with various pull rates.<sup>21</sup> In Reference (21) the extent of this instability corresponds to a change in  $N_o$  of about  $6 \times 10^{11} \text{ cm}^{-2}$  after a bias-temperature test at 400°C and  $-2 \times 10^6 \text{ V/cm}$  for 2 min. In addition to the large shift, the  $C$ - $V$  curve has also deformed, corresponding to the introduction of approximately  $5 \times 10^{11} \text{ cm}^{-2}$  interface states. Note that the original  $C$ - $V$  curves showed that  $N_o$  is of the order of  $10^{11} \text{ cm}^{-2}$ . Deal et al<sup>21</sup> pointed out that the total change in  $C$ - $V$  characteristics at a given field was proportional to the initial  $Q_{ss}$  value. Miura and Matukura reported similar observations.<sup>9</sup> In their case, the original value of  $N_o$  was also a few times  $10^{11} \text{ cm}^{-2}$ . In contrast to Deal et al, they emphasized that the extent of slow trapping instability was considerably smaller for (100) than (111) orientation.

### 4.3 Discussion of Interface Properties

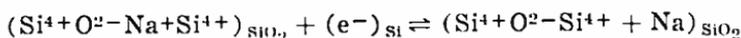
The above results demonstrate that under proper preparation conditions, especially by a combination of cleaning, thermal oxidation, and annealing into an *in situ* process, and by minimizing contamination from and/or through the reaction vessel, an Si-SiO<sub>2</sub> interface with a high degree of perfection can be obtained. This perfection is primarily revealed by the lack of surface states as well as interface and slow trapping type instabilities. Since the silicon-oxygen interface has a great tendency toward disorder as observed in chemisorption studies<sup>60</sup> and the variety of defects in the noncrystalline silicon dioxide is very large (see Section 5), this high perfection is quite remarkable. A high degree of interface perfection is not very common, even in those systems where both components are well-characterized crystalline materials, e.g., semiconductor heterojunctions. The perfection of the Si-SiO<sub>2</sub> interface is most probably due to the noncrystalline nature of the oxide. The relatively large freedom in the atomic positions, as compared with a single crystal, may provide an accommodation of the two components without the appearance of dislocations, vacancies, etc. Thus, all the "dangling" silicon bonds are saturated with oxygen.

From the results presented here, it is clear that the oxidation mechanism and the nature of the interface are intimately interrelated. The influence of oxidation rate and temperature on the interface state density was also observed by others, but their interpretations are somewhat different. Thus, Kooi<sup>81</sup> attributes the larger values of  $N_o$  measured when larger amounts of sodium were present in the oxide to the formation of  $\text{Si}^{4+} \text{O}^{2-} \text{Na}^+ \text{Si}^{3+}$  complex in the vicinity of the Si-SiO<sub>2</sub>

interface. The complex behaves as a donor by giving off an electron to silicon according to the reaction



Kooi further states that the density of these complexes is determined by the transport of  $\text{Na}^+$  toward the  $\text{Si-SiO}_2$  interface during oxidation and the motion of neutral Na into the silicon or back to the oxide-gas interface. According to Kooi, this process could explain why the concentration of sodium at the  $\text{Si-SiO}_2$  interface and the "oxide charge" decreases with increasing oxide thickness (i.e., with decreasing final oxidation rate). The neutral form of sodium arises by the reaction



Even though sodium and/or hydrogen contamination during oxidation may contribute to the interface states or "oxide charge," it is not primarily responsible for them. Our oxide films that resulted in virtually perfect interface, as described above, have an estimated Na content in the oxide of  $10^{16} - 10^{17}$  atoms/cm<sup>3</sup>. Oxides grown under sodium doping (see Section 3) contained  $10^{20} - 10^{21}$  Na atoms/cm<sup>3</sup>. The values of  $N_o$  and  $dN_o/dE$  were  $7.6 \times 10^{12}$  cm<sup>-2</sup> and  $2.8 \times 10^{12}$  (eV)<sup>-1</sup> cm<sup>-1</sup>, respectively. However, after He annealing  $dN_o/dE$  became virtually zero and  $N_o$  was  $8 \times 10^{11}$  cm<sup>-2</sup> (for p-type Si, (111) orientation, 670 Å oxide) without observable change in the oxide Na content. This value of  $N_o$  is only twice as large as that given by Kooi for a comparable specimen with an Na content of less than  $2 \times 10^{16}$  atoms/cm<sup>3</sup>. Apparently, most of the sodium introduced during oxide growth is in an inactive form from the standpoint of both interface states and ion-drift instability.

Deal et al.<sup>21</sup> attribute the "oxide charge" to the presence of trivalent silicon or oxygen vacancies in a region about 200 Å thick in the oxide at the  $\text{Si-SiO}_2$  interface. This explanation is very similar to ours, i.e., that disorder at the interface is responsible for the interface states, and those states that are outside the silicon forbidden band can be considered as the "oxide charge" (see below).<sup>\*</sup> However, their hypothesis that the slow trapping instability is a fundamental property of the  $\text{Si-SiO}_2$  interface caused by the injection of excess silicon into the oxide under negative field is not justified, since, as we have discussed, this instability can be eliminated by proper annealing.

<sup>\*</sup> It should be noted that the possible role of trivalent silicon and other defects in the oxide in interface phenomena was pointed out quite early.<sup>18</sup>

Kubo and Ichinohe attempted to correlate the density of interface states, expressed as  $N_{it}$ , with the amount of "dissolved" oxygen present at the Si-SiO<sub>2</sub> interface at the end of the oxidation.<sup>22</sup> They have calculated the concentration of this oxygen by using equations given by Deal and Grove.<sup>24</sup> However, it can be shown<sup>23</sup> that these equations are not quite correct. Furthermore, it is difficult to visualize how dissolved oxygen either in molecular or ionic form could act as donor type interface state. This difficulty was recognized by the authors.

From the effects of oxidation conditions on the density of interface states and oxidation mechanism, as well as from the annealing behavior, the following model emerges. Various impurities, mostly sodium and water (hydroxyl, proton) participate in the thermal oxidation of silicon. Some of the impurities contribute to the disorder at the interface region. They can also contribute to the ion migration and interface instability. Part of this disorder is a result of the oxidation process itself in the sense that (1) some bonds remain unsaturated and/or (2) defects (e.g., trivalent silicon, oxygen vacancies, etc.) are present in the oxide close to the interface. The density of interface states is determined by this disorder. It is less for lower final oxidation rates and higher temperatures. These states can be divided into the following groups:

(1) Donor states close to or above the conduction band of silicon. This group is permanently ionized, and represents a fixed positive charge, the so-called "oxide charge". These states can be eliminated by high-temperature annealing, unless some impurity (e.g., H<sub>2</sub>O, Na) interferes.

(2) Acceptor and/or donor states within the forbidden band of silicon. The density of these states can be reduced by high-temperature annealing or hydrogen treatment at 300° to 500°C. Impurities in the oxide may contribute under certain conditions to these states, leading to interface instability.

(3) Hole traps close to the valence band edge of silicon. These states can trap holes from the silicon valence band causing the slow-trapping instability. They can be eliminated by proper high-temperature annealing but not by H<sub>2</sub> treatment. Sodium in the oxide does not contribute to these states, since even Na-doped oxides did not show the slow-trapping instability.

The importance of cooperative processes at the interface, as manifested by the dependence of surface-state density on orientation, annealing behavior, and interaction between migrating ions and interface states, has already been mentioned. Another aspect of such processes is shown by the effects of electron bombardment on the Si-SiO<sub>2</sub> inter-

face.<sup>63</sup> It was observed that part of the damage introduced, as revealed by an increased  $dN/dE$ , is qualitatively proportional to the original value of  $dN/dE$ , and is zero if no interface states were originally present. Thus, the extent of destruction of order at the interface by electron bombardment or by introduction of impurities is determined by the perfection of the interface structure. Another manifestation of cooperative effects is the relationship between original surface charge and the extent of slow-trapping instability as observed by Deal et al<sup>21</sup> and mentioned above. The role of cooperative phenomena in order-disorder transitions of two-dimensional surface structures and in chemisorption was also emphasized by Lander.<sup>64</sup>

The fact that, regardless of orientation, surface states can be eliminated by proper cleaning, oxidation, and annealing indicates that the state of the Si-SiO<sub>2</sub> system is essentially determined by the interface, most probably by the degree of order there. This conclusion concerning the importance of the interface is at variance with one advanced recently,<sup>65</sup> namely that the Si surface potential is predominantly determined by electronic interactions between Si and SiO<sub>2</sub> due to their work-function difference. It is, however, in agreement with results obtained by others.<sup>21</sup> From the above discussion, one can deduce that the net charge in surface states cannot be thought of as a combination of a fixed positive charge in the oxide and negative charge in acceptor interface states whose density is higher for (100) than (111) orientation, as has been suggested.<sup>64</sup>

## 5. "BULK" PROPERTIES OF SiO<sub>2</sub> FILMS

We have seen in the preceding section that a carefully prepared Si-SiO<sub>2</sub> interface is quite perfect. The very small extent of instability phenomena shows that the oxide film is also quite perfect. The high *schubweg*\* value of electrons injected into the oxide (of the order of 10  $\mu\text{m}$  for a field of 10<sup>6</sup> V/cm)<sup>66</sup> and the low concentration of traps ( $\sim 3 \times 10^{14}/\text{cm}^3$ )<sup>19</sup> also demonstrate a perfection that was not expected for a noncrystalline material. This great perfection is most probably due to the high degree of short-range order in these films. If, however, the complete lack of long-range order does not exist, i.e., crystallization occurred, even in a localized manner, then the protective properties of the oxide film deteriorate and the density of interface states increases. However, even when no gross structural changes can be detected, optical and other properties reveal significant differ-

---

\* *Schubweg* is the mean distance that a carrier drifts under the influence of an applied electric field before it is immobilized by deep trapping.

ences among SiO<sub>2</sub> films prepared by various methods. The various aspects of noncrystallinity and the nature of the Si-O bond and its influence on oxide properties are discussed below.

### 5.1 Noncrystallinity

It was shown by several workers that thermal or anodic oxidation of silicon usually results in noncrystalline oxide films.<sup>87</sup> In contrast to most oxidation processes, the oxidation occurs without localized nucleation, that is, the silicon surface is uniformly covered with a noncrystalline oxide. Even chemisorption of oxygen on silicon leads to an unusual noncrystalline surface structure.<sup>80</sup> The grown SiO<sub>2</sub> films, as well as those prepared by chemical deposition methods, stay noncrystalline even after prolonged heat treatments at about 1200°C in various ambients. If, however, the silicon surface is not smooth enough (e.g., sand-blasted), then protrusions, cracks, etc. may act as nucleation sites for the growth of crystalline oxide. Small crystallites formed at imperfections in the early stage of growth are overtaken by the growing oxide (recall that oxygen diffuses through the oxide and growth occurs at the Si-SiO<sub>2</sub> interface) and they remain at the outer surface. It is significant that these crystallites do not act as crystallization centers during the later stage of growth, at least below a few micrometers thickness. These results, as well as the following discussion of the role of  $\pi$ -bonds in the stability of noncrystalline structure and electronic conduction, were recently published.<sup>30</sup>

### 5.2 Properties of the Si-O Bond

A comparison of silicon dioxide with the oxides of other metals and semiconductors reveals that there is no other oxide that shows such a great tendency to form noncrystalline structure and to remain in this form even during high-temperature treatments. Another unique property of SiO<sub>2</sub>, as previously mentioned, is the very low ratio of ionic to electronic current during anodic oxide growth coupled with high field across the oxide. These and other properties are most probably related to the quantum chemistry of the Si-O bond and to the spatial arrangement of Si and O atoms in the oxide. Although the fundamental nature of the Si-O bond is the same in the various crystalline and noncrystalline forms of SiO<sub>2</sub>, relatively small changes in the spatial arrangement, defect structure, etc. may have profound effects on transport and interface properties.

#### 5.2.1 Quantum Chemistry of Si-O Bond

One of the most striking features of the Si-O bond is that the Si-O distance is much less than the expected value, as can be seen from the

following comparison. In all the modifications of  $\text{SiO}_2$  the measured Si-O distance is 1.60 Å, whereas the sum of covalent radii is 1.91 Å for a single bond or 1.79 Å for a double bond, and the sum of ionic radii is 1.73 Å. Even if the Schomaker-Stevenson correction (for the partial ionic character of the bond) is taken into account, the calculated single bond length is only 1.76 Å. Pauling proposed that the  $p$ -electrons of the oxygen that do not participate in the (partially ionic)  $\sigma$ -bond may be involved in forming bonds between Si and O by occupying the otherwise empty  $d$ -levels of silicon.<sup>88</sup> Thus, the total bond is a result of resonance between ionic and covalent bonds, the latter having  $\sigma$  and  $d\pi$ — $p\pi$  components. He calculated that the ionic character of  $\sigma$ -bond is 41%, the amount of  $\pi$ -bond character is 32%, and the  $\sigma$ -bond character is 27%. The net formal charge on Si is +0.36 electronic unit, and the calculated Si-O distance is 1.63 Å. Using the above values for the  $\sigma$ - and  $\pi$ -bond character of the Si-O bond, the Si-O-Si bond angle can be roughly estimated as follows. For a pure  $\sigma$ -bond the angle would be  $90^\circ$ , and for a pure  $\pi$ -bond it would be  $180^\circ$ , hence the estimated value is given as  $90[0.27/(0.27 + 0.32)] + 180[0.32/(0.27 + 0.32)] = 138^\circ$ . This value can be compared with  $143^\circ \pm 17^\circ$  as reported for vitreous  $\text{SiO}_2$ .<sup>89</sup>

Cruickshank described the bonds in the  $\text{SiO}_4^{4-}$  group on the basis of molecular orbital theory.<sup>90</sup> His arguments are used here in a somewhat modified form. In  $\text{SiO}_2$  the silicon atom is surrounded tetrahedrally by four oxygen atoms. The hypothetical electron distribution in an  $\text{SiO}_{4/2}$  tetrahedron before bonds have been established is as follows:  $(\text{Si}3s^2)^4$ ,  $(\text{Si}3d)^0$ , and  $4/2(2p)^4$ ; altogether there are 12 electrons, neglecting the  $1s$ , and  $2s$  and  $2p$  electrons of silicon, as well as the  $1s$  and  $2s$  electrons of oxygen. The bonds available to form molecular orbitals (MO) are as follows:  $4(\text{Si}3s^2)$ ,  $5(\text{Si}3d)$ ,  $4(2p_y)$ , and  $4/2(o2p_x)$ ; altogether there are 15 orbitals. (It was assumed that the paired electrons of oxygen are on the  $(2p_x)$  orbital, and that the  $(2p_z)$  orbitals are involved in bonds with silicon on the opposite sides of the oxygen atoms.) The  $(\text{Si}3s^2)$  orbitals form four  $\sigma$ -bonds with the four  $(2p_y)$  orbitals; these are fully occupied with 8 electrons, i.e., in the MO description they correspond to  $(\sigma)^8$ . The five originally empty  $3d$  orbitals of the silicon split in the tetrahedral environment into a lower doubly degenerate set of  $e$  orbitals and an upper triply degenerate set of  $t_2$  orbitals. If we take the coordinate axes parallel to sides of the cube enclosing the tetrahedron (the Si atom being at the center of the cube), then the  $e$  and  $t_2$  sets comprise the  $d_{x^2-y^2}$  and  $d_{z^2}$  orbitals and the  $d_{xy}$ ,  $d_{xz}$ , and  $d_{yz}$  orbitals, respectively. As Cruickshank pointed out, the overlap with the  $(2p)$  orbitals is better for the  $e$  than for the  $t_2$

set. Hence the  $e$  orbitals are more effective in bonding. Depending upon whether low- or high-spin configuration is favored, the corresponding electron distribution is  $(e)^4(t_{2g})^0$  or  $(e)^2(t_{2g})^2$ , respectively. For transition-metal ions, due to the small splitting energy in tetrahedral coordination, the high-spin configuration is the general case; however, with silicon, the situation may be different because the occupancy of these orbitals is a result of bond formation. Taking the low-spin case, the four electrons form two strong  $\pi$ -bonds, and the three  $t_{2g}$  orbitals remain empty (nonbonding orbitals). Two more nonbonding (empty) orbitals originate from the  $(02p_x)$  orbitals that originally contained the paired electrons. Finally, we have four empty  $\sigma^*$  (anti-bonding) orbitals (derived from the  $(02p_y)$  orbitals). In summary, we have 6 strongly bonding orbitals ( $4\sigma + 2\pi$ ) fully occupied with the available 12 electrons, as well as 5 nonbonding and 4 anti-bonding orbitals that are empty. This gives rise to a bond order of 1.5. If the Si-O-Si angle were  $180^\circ$ , the oxygen would have an  $sp$  hybrid orbital; this would form a  $\pi$ -bond (of bond order 1) with the silicon and, hence, the total bond would be a  $\sigma$ - $\pi$  double bond of order 2.

### 5.2.2 Role of $\pi$ -Bond in the Noncrystallinity and Electronic Conductance

The important role of covalent bonds in the formation of vitreous oxides is well known.<sup>91</sup> Also, there are indications that the nature of these bonds may influence the high-field electronic conduction processes in insulating oxide films. This can be illustrated by comparing the anodization behavior of silicon, germanium, titanium, and zirconium. The electronegativity of these elements decreases in the above order, except for germanium which has the largest value. The coordination of the metal and oxygen atoms in the oxides increases from 4:2 for SiO<sub>2</sub> to 8:4 for ZrO<sub>2</sub>. The general trends in both electronegativity and coordination indicate a decreasing tendency to form noncrystalline oxides in the above order. This was indeed found to be the case, especially for anodic oxides; Si, as previously discussed, always forms noncrystalline oxide films, whereas Zr never does. (Thermal oxidation of Ge, Ti, and Zr consistently results in (poly)crystalline oxides.) Furthermore, the ratio of ionic to electronic current during anodization, as reflected in the oxidation efficiency, increases from about 1% for Si to essentially 100% for Zr.<sup>92</sup> These trends clearly indicate a connection between covalency, the ability to form noncrystalline oxides, and electronic conduction under high field conditions.

The large electronegativity of germanium, as well as the facts that the stable crystalline form of its oxide has a silica structure and that

both silicon and germanium have similar semiconducting properties, makes a closer comparison of the oxidation behavior of silicon and germanium very interesting. First of all, unlike the case with silicon, thermal oxidation of germanium begins with localized nucleation and results in a (poly)crystalline film.<sup>93</sup> Anodic oxidation of germanium in nonaqueous electrolytes, however, produces noncrystalline films.<sup>94,95</sup> The oxidation efficiency is higher than for silicon, but the ionic resistivity, as determined from the apparent resistivity during constant current anodization and oxidation efficiency, is about the same as for silicon. This shows that anodic  $\text{GeO}_2$  has a lower apparent electronic conductivity than  $\text{SiO}_2$  film. These results are unexpected in the light of the generally observed behavior of silicon and germanium compounds which, in accordance with the electronegativity values (1.74 for Si, 2.02 for Ge), clearly indicates that Ge has a greater tendency to form covalent bonds than does Si. The discrepancy can, however, be resolved by considering the contribution of  $d\pi$ - $p\pi$  bonds to the total covalent character of the bond. Urch has pointed out<sup>96</sup> that, because of a difference in radial atomic wave functions, the  $4d$  orbitals are much less efficient at  $\pi$ -bonding with  $2p$  orbitals than are  $3d$  orbitals in comparable situations.

In addition to the large electronic conduction during anodic oxidation, appreciable electronic conduction also takes place in the Si-SiO<sub>2</sub>-electrolyte system when Si is the anode but the voltage is so low that no oxide growth occurs.<sup>50</sup> The magnitude of this electronic conduction is larger by several orders of magnitude than that observed with photoemission of electrons (holes).<sup>19,86</sup> The dependence on oxide thickness of this conductance indicated that the likelihood of Schottky emission being the controlling mechanism decreases with increasing thickness. Also, with increasing thickness the activation distance increases if conduction is due to hopping, or the energy spread of traps in the oxide increases if space-charge-limited current is the conduction mechanism.

Impedance measurements on MOS capacitors occasionally showed significant electronic conduction. Since (1) the temperature of measurements was relatively low (about 20°C), (2) no polarization phenomena were observed, and (3) the extent of possible ionic conduction (as inferred from the negligible effects of an ion migration under high field at elevated temperature) is practically nil, the conduction is electronic rather than ionic. This conduction occurs in localized spots rather than being a "bulk" property of the oxide. We have qualitatively observed that anodic oxides had a higher density of these spots than thermally grown oxides; also, the spots were more conducting in the former. This

was very surprising since, due to the very nature of the growth process, anodic oxide films are expected to be free of pinholes. Indeed, we have found by electron-microscope investigations that anodic oxides are more perfect than thermally grown oxides. This is in agreement with the observations of others.<sup>37</sup> Thus, it is not probable that the conduction is caused by simple metal-to-silicon contacts through pinholes in the oxide. It is more likely that the oxide contains regions ("channels") that are relatively highly conducting. These channels should not be considered as large-scale distortions of the Si-O network; rather, they represent easy paths for the electron flow. Among the thermally grown oxides, those produced in dry oxygen by the very clean *in situ* method using r-f heating contained more conducting spots than those prepared in resistance-heated silica tube furnaces, especially in steam. Also, oxides grown on (100) silicon had a greater density of these spots than those grown on (111) substrates. This observation indicates that electronic conduction is an intrinsic property of these channels and that impurities block the electronic current.

In the light of the above discussion, it is not surprising that, due to the complexity of the Si-O bond and the microheterogeneity of the noncrystalline oxide, electronic conduction in SiO<sub>2</sub> films is far from being understood. Nevertheless, the nature of the Si-O bond indicates that the energy-band diagram shown in Figure 1 is somewhat oversimplified. The true energy-band diagram probably would show similarities with that of transition-metal oxides. Thus, we can expect that, depending on the injection and/or excitation processes, conduction may take place by (1) electrons in the antibonding  $\sigma^*$ -orbital (conduction band), (2) holes in the  $\sigma$ -orbital (valence band), (3) electrons or holes in the  $e^-$  and/or  $t_2^-$ -orbitals ( $\pi$ -band), (4) hopping of electrons in  $e^-$  and/or  $t_2^-$ -orbitals, and (5) impurity effects, i.e., band conduction or hopping, depending on whether the impurity orbitals (e.g., trivalent Si) do or do not overlap.

Note that the  $\pi$ -electrons are in a higher energy level than the  $\sigma$ -electrons and their orbitals are also more delocalized. Therefore, their presence favors electronic conduction. They enhance the formation of noncrystalline structures since they hinder rotation along the bond.

### 5.2.3 Dependence of Si-O Bond on Oxidation Method

Although the Si-O bond is essentially the same in vitreous silica and SiO<sub>2</sub> films grown or deposited by various methods, there are indications that, depending on the preparation method (including after treatments), small changes in the exact nature of the bond can occur.

We have observed, for instance, that the relative polarizabilities of  $\text{SiO}_2$  films may depend on the preparation method.<sup>6</sup> If we take the polarizability of vitreous silica as unity, then films grown thermally in dry  $\text{O}_2$  have a value of 0.964; whereas anodic oxides show higher polarizabilities, about 1.035.

The crystalline modifications of  $\text{SiO}_2$  also reveal nonnegligible variations in the properties of the Si-O bond and the following trend can be established. With decreasing anisotropy the Si-O-Si angle, mole refraction, molar volume, and bond-force constant increase from  $\alpha$ -quartz to  $\beta$ -cristobalite.<sup>98</sup> This can be attributed to an increase in the  $\pi$ -character of the bond. In noncrystalline  $\text{SiO}_2$ , the Si-O-Si bond angle does not have a fixed value but is characterized by a distribution about a mean. Relatively small variations in the bond angles can have profound influence on the nature of  $\pi$ -bonds as was illustrated by the example of  $\alpha$  and  $\beta$  quartz.<sup>99</sup> A nonrandom distribution of bond angles may lead to microheterogeneities on the atomic scale.

### 5.3 Microheterogeneous Structure\*

#### 5.3.1 Experimental Evidence

Gas permeability studies indicated the presence of "microchannels" in thermally grown  $\text{SiO}_2$  films; the diameter of these channels was estimated to be less than 50Å.<sup>99</sup> Etching of  $\text{SiO}_2$  films also proceeds in an inhomogeneous manner and the fast etching regions have an increased tendency toward dielectric breakdown.<sup>100</sup> The localized nature of electronic conduction in MOS structures was described above.

It is very instructive to compare infrared absorption, mole refraction, etching behavior, and electronic conduction of  $\text{SiO}_2$  films prepared by various methods. This comparison is given in Table II. As the frequency of the Si-O-Si stretching vibration decreases from about 1090 to 1040  $\text{cm}^{-1}$  and its half bandwidth increases from 75 to 110  $\text{cm}^{-1}$ , the etching rate, electronic conduction, and mole refraction increase in the following order: (1) steam-grown and (2) dry-oxygen-grown thermal oxide, and (3) constant-current-grown and (4) constant-voltage-grown anodic oxide. The heat treatment behavior of  $\text{SiO}_2$  films reveals further interesting features. After a very short treatment of anodic oxide at relatively low temperatures (5 minutes, 900°C), the etch rate decreased by almost a factor of four, the infrared absorption shifted to higher frequency, and the mole refraction decreased to a value characteristic of thermal oxides and vitreous

\* Many of the ideas presented in this section have not been previously published. They were, however, expressed in A. G. Revesz's Doctoral Dissertation, University of Technical Sciences, Budapest, 1967.

Table II—Infrared Absorption, Mole Refraction, Etching Behavior, and Electronic Conduction of Noncrystalline SiO<sub>2</sub> Films.

Material	Stretching vibration absorption		Absorption at 800 cm <sup>-1</sup>	Relative mole refraction	Etch rate		Electronic conduction
	position, cm <sup>-1</sup>	half width, cm <sup>-1</sup>			A/min	relative	
steam-grown	1079-1080 <sup>a</sup>	75-78 <sup>a</sup>	—	—	20 <sup>a</sup>	1.0 <sup>a</sup>	low <sup>c,d</sup>
steam-grown	1093 <sup>b</sup>	~90 <sup>b</sup>	stronger <sup>b</sup>	1.04 <sup>b</sup>	2.0 <sup>b</sup>	1.0 <sup>b</sup>	
dry O <sub>2</sub> -grown	1074 <sup>b</sup>	73 <sup>b</sup>	weaker <sup>b</sup>	1.00-1.04 <sup>b,e</sup>	—	—	medium <sup>c,d</sup>
anodic, const. current	1052 <sup>a</sup>	112 <sup>a</sup>	—	1.06-1.08 <sup>b,e</sup>	190 <sup>a</sup>	8.5 <sup>a</sup>	high <sup>d</sup>
anodic, const. voltage	1039 <sup>a</sup>	93 <sup>a</sup>	—		—	—	
anodic, heat treated	1075 <sup>a</sup>	85 <sup>a</sup>	—	1.04 <sup>b</sup>	20-54 <sup>a</sup>	1.0-3.5	—
steam, anodized (no incr. in thickness)	1054-1074 <sup>a</sup>	82-105 <sup>a</sup>	—	—	120 <sup>a</sup>	6.0 <sup>a</sup>	—
pyrolytic	~1082 <sup>b</sup>	92 <sup>b</sup>	strong <sup>b</sup>	1.02-1.04 <sup>b</sup>	6-20 <sup>b</sup>	3-10 <sup>b</sup>	—
pyrolytic, heat treated	1092 <sup>b</sup>	88 <sup>b</sup>	strong <sup>b</sup>	1.02-1.04 <sup>b</sup>	2.0 <sup>b</sup>	1.0 <sup>b</sup>	—

Superior letters refer to the following references: a, Ref. (54); b, Ref. (53); d, Ref. (50); and e, Ref. (6). In this Table, the relative etch rates are also given because the etchants reported in (a) and (b) were not the same. The molar refractivities were calculated on the basis of refractive index and density values given in the indicated references. The molar refractivity of quartz (7.19 cm<sup>-3</sup>) was taken as unity in calculating the relatives values.

fused silica.<sup>53</sup> The only reasonable explanation is that relatively slight rearrangements of the bond angles during heat treatment caused the changes. It is very unlikely that Si-O bonds could have been broken and that atomic diffusion could have occurred at these temperatures and during such short times.

The phenomena described above are very complex, and more careful and well-correlated experiments are needed before they can be explained in a completely consistent manner. This is probably partially due to the fact that the nature of defects is much more complicated in noncrystalline than in crystalline solids. So far as we know, no rigorous treatment of this problem is available. Nevertheless, a model is proposed here to explain a least some of these observations. It is essentially based on the noncrystalline structure of the SiO<sub>2</sub> films. In principle, the model can be applied to other noncrystalline materials as well.

### 5.3.2 Defects in Noncrystalline Solids

The following discussion is restricted to that class of noncrystalline materials that is characterized by a high degree of short-range order and chemical homogeneity, i.e., to glasses or vitreous solids (SiO and Si<sub>3</sub>N<sub>4</sub> are, for instance, excluded). As we have already mentioned, grown or deposited SiO<sub>2</sub> films belong to this class because the Si atoms are, in general, surrounded tetrahedrally by four oxygen atoms, but the Si-O-Si angles vary around a mean value. This situation can be contrasted with that existing in a perfect single crystal, where the bond angles have well-defined fixed values, and with an amorphous solid (e.g., SiO), where no short-range order exists (Si-Si bonds also occur).

In the light of the large variations possible in the bond angle without the bonds breaking, a perfect glass can be defined as a noncrystalline material that has perfect short-range order, but has an atomic network that is completely random. The second criterion means that the mean value and distribution of the bond angles are independent of the spatial position and size of the volume element under consideration. A nonrandom distribution of the bond angles leads to microheterogeneities on the atomic scale that may be distributed in either a random or a correlated manner. If the randomly distributed microheterogeneities show a larger degree of order than their surrounding, then they correspond to the so-called "crystallites" of the well-known crystallite theory of glasses.<sup>102</sup> A correlated distribution of bond angles along preferred directions leads to the appearance of channel-type microheterogeneities in which the nature of the Si-O bond, especially the extent of the  $d\pi$ - $p\pi$  overlap, is different.

There are indications that these channels in SiO<sub>2</sub> films have a greater degree of order than their environment. In this sense the channel in a glassy solid can be roughly considered as an inverse analogy of the edge dislocation in a single crystal, the latter also being a correlated defect but characterized by less order than its surrounding.

Deviations in the Si-O distance may also be partially responsible for channel formation. If these deviations lead to a rupture of the Si-O bonds, that is, to localized disruption of the short-range order, or if impurities have been introduced, then defects that are very similar to point defects in single crystals (vacancy, interstitial) have been created. An increase of the size of microheterogeneities, perhaps with the cooperation of point defects, can result in microcrystallites. This situation corresponds roughly to the degradation of a single crystal to a polycrystal through the appearance of small and large angle boundaries.

The flexibility of the noncrystalline structure (as manifested in the large spread in bond angles) makes a clear distinction between the various types of disorder almost impossible. This means that the transition from relatively small disturbances in the electron distribution to gross defects (voids, pinholes, etc.) may be continuous. Therefore, a discussion, for instance, of transport processes in noncrystalline SiO<sub>2</sub> films can be very difficult and ambiguous if concepts developed in connection with single crystals are used but no attention is paid to the peculiarities of the defect structures.

Although localized variations in bond strength (and to some extent in bond angles) have been considered in connection with the two pentagonal dodecahedral models of vitreous silica,<sup>102,103</sup> these models differ to a great extent from that proposed here. Both models consider the bonds as essentially ionic and are based on Si-O-Si angles of 180°. These requirements are, however, not fulfilled in reality. Nevertheless, a concept similar to that of a channel emerged from this model as it was stated that "undistorted pentagonal dodecahedrons (pd) by themselves cannot fill space except for a volume occupied by a single chain of pd."<sup>103</sup> The chains are surrounded by weak regions. The length of such a chain was stipulated to be perhaps 100 Å. It should be noted that many physical properties, such as radial distribution of atoms, mechanical properties, etc., could be explained by the pd model, but the properties we are most concerned with were not discussed. The microheterogeneities discussed here are also different from those described by Vogel,<sup>101</sup> since those are chemical inhomogeneities in multi-component glasses, (e.g., SiO<sub>2</sub>-Na<sub>2</sub>O) whereas we are concerned with a one-component system.

### 5.3.3 Discussion of Observed Phenomena

In the light of this model, the observations described above indicate the following trends. Anodic oxidation greatly enhances the formation of channels. This is understandable if we consider that the electric field during anodic growth ( $2 \times 10^7$  V/cm) is about 10% of the internal field between atoms in the oxide. It can be expected that such a high field exerts an influence on the oxide structure. Furthermore, as is the case with dislocations, if a channel forms it will have a tendency to grow. Oxidation in steam results in a minimum density of channels. Infrared absorption spectra indicate that this oxide film contains more ring structures than the others.<sup>105</sup> The incorporation of hydrogen apparently enhances the formation of these rings and decreases the chance of creating channels. The overall order is larger in the oxides that have ring structures than those with well-developed channels. Seemingly, the formation of channels is coupled with an increase of disorder in their environment, so that the overall disorder increases. These channels are probably similar to the structural tunnels in quartz along the *c*-axis.

During thermal or anodic growth, there is a tendency to lose the correlation; hence, the density of undistorted channels decreases with thickness, and the electronic conduction becomes increasingly controlled by the oxide instead of the interface. Incorporation of impurities or trapping of electrons (holes) may block the channels and thus decrease the electronic conduction.

The influence of the substrate orientation is evident from the atomic arrangements of various planes of the silicon lattice. As previously mentioned, the (100) orientation enhances the formation of a more ordered oxide structure at the interface and, hence, the development of channels.

The presence of quartz-like regions in vitreous fused silica has been noted; these grow in a preferred direction during the solidification of the melt.<sup>106</sup> In this case the migration of sodium was found to be very rapid because it occurred along structural tunnels that were left in the vitreous fused silica as relics of the quartz crystal. Similarly, Hofstein found, during an investigation of instability phenomena in MOS structures, that one mode of transport of sodium ions in the oxide under certain conditions is so fast that it probably occurs within channels rather than through the random network.<sup>67</sup>

### 5.4 Point Defects in SiO<sub>2</sub>

Point defects in vitreous SiO<sub>2</sub> have been enumerated by Stevels and Kats.<sup>107</sup> A description of defects that probably play important

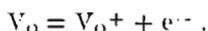
roles in transport and interface phenomena associated with SiO<sub>2</sub> films is given in this section.

The occurrence of silicon vacancies is not very likely because of the high energy of the Si-O bond and the fourfold coordination of Si, but a somewhat similar defect may exist under certain conditions. This is a reduced form of silicon that can be formally written as Si<sup>3+</sup>. This trivalent Si can arise during oxidation, or it can be introduced by high-temperature treatment in hydrogen (see below). According to the following reaction, it behaves as donor:



If this defect is in the vicinity of the Si-SiO<sub>2</sub> interface, it may donate the electron to silicon; in this sense, it can be considered as a donor-type oxide (or interface) state. It may also participate in electronic conduction (see previous section).

Oxygen vacancy may be created during oxidation, especially in the vicinity of the Si-SiO<sub>2</sub> interface, but because the oxidation rate-constant is not pressure independent, it cannot be the major diffusing species. Nevertheless, it may be important as a donor-type state,



if the electron is transferred to silicon.

Since the thermal growth of SiO<sub>2</sub> proceeds by the diffusion of some ionized oxygen-bearing species, and the parabolic rate-constant increases with increasing oxygen pressure, it is very probable that oxygen interstitial is the species mainly responsible for the growth. However, as discussed in Section 3, the oxygen interstitial is generally associated with extrinsic defects, e.g., Na, H.

From the discussions of Sections 3 and 4 it is evident that sodium or similar metals and various forms of hydrogen-containing species are the most important extrinsic defects in grown SiO<sub>2</sub> films. Introduction of sodium into the oxide leads to breaking of the Si-O-Si bridge by forming the configuration Si-ONa NaO-Si or Si-ONa O<sub>i,n</sub>-Si. Hydrogen in vitreous silica can be present in a large variety of forms—molecular and atomic interstitial, interstitial water, and as OH groups attached to Si.<sup>48</sup> In the latter case an important distinction arises depending on how the H was introduced. Reaction of the oxide with water at elevated temperature results in the following configuration: Si-OH HO-Si, whereas reaction with hydrogen leads to the trivalent form of silicon<sup>49</sup>: Si-OH Si<sup>3+</sup>. The donor characteristics of this reduced silicon was previously pointed out,<sup>48</sup> and recently it was confirmed by electron-spin-resonance studies.<sup>108</sup>

## 6. CONCLUSIONS

A nearly perfect Si-SiO<sub>2</sub> interface can be obtained by careful pre-oxidation cleaning, oxidation, and annealing, especially if these processes are combined into an *in situ* process. The degree of perfection corresponds to about 10<sup>9</sup> to 10<sup>10</sup> electronic charges/cm<sup>2</sup>. Another characteristic of such an interface is the lack of trapping phenomena. The significance of this achievement from the standpoint of surface physics and chemistry is comparable to that of obtaining nearly perfect silicon single crystals from the viewpoint of solid-state physics and chemistry. It establishes the Si-SiO<sub>2</sub> interface as the best understood solid-solid interface system at the present. The high degree of perfection of the oxide film, especially as obtained by thermal oxidation of silicon and the possibility of measuring very small changes in the silicon space-charge capacitance made it possible to study ionic and electronic transport processes in the oxide at a level lower than ever before reached. One important result of these studies is the demonstration of impurity effects in the thermal oxidation of silicon. Another is the insight obtained into the structure of noncrystalline SiO<sub>2</sub> and the recognition of the importance of  $\pi$ -bonds in stabilizing noncrystalline structures and in various phenomena such as electronic conduction and optical properties. Considerations regarding the nonrandom distribution of these  $\pi$ -bonds led to a new concept in the area of defects in noncrystalline solids, namely, microheterogeneities on the atomic scale (channels), by which various transport processes in SiO<sub>2</sub> films can be explained.

The concepts and measurement methods developed during the past five years in connection with the Si-SiO<sub>2</sub> interface investigations open up new avenues for studying other semiconductor-insulator systems. Indeed, several such systems are already being investigated.

From a technological point of view, the advances made with the Si-SiO<sub>2</sub> system, especially the high degree of control of its properties, are important in the manufacture of silicon devices in general and MOS devices in particular. These advances also provide great promise for large-scale integration if localized breakdown phenomena can be reduced and if a high degree of uniformity of the desired properties across the whole Si wafer can be achieved.

## ACKNOWLEDGMENTS

The authors should like to thank J. T. Wallmark for a critical reading of the manuscript. The constant interest and encouragement of P. Rappaport is appreciated.

## BIBLIOGRAPHY

1. For example, J. T. Wallmark and H. Johnson, eds., *Field Effect Transistors*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966.
2. A. Many, Y. Goldstein, and N. B. Grover, *Semiconductor Surfaces*, North-Holland Publishing Company, Amsterdam, 1965.
3. E. S. Schlegel, "A Bibliography of Metal-Insulator-Semiconductor Studies," *IEEE Trans. Electron Devices*, Vol. ED-14, p. 728, Nov. 1967.
4. R. J. Archer, "Optical Measurement of Film Growth on Silicon and Germanium Surfaces in Room Air," *J. Electrochem. Soc.*, Vol. 104, p. 619, Oct. 1957.
5. R. J. Archer, "Determination of the Properties of Films on Silicon by the Method of Ellipsometry," *J. Opt. Soc. Amer.*, Vol. 52, p. 970, Sept. 1962.
6. A. G. Revesz and K. H. Zaininger, "Some Properties of SiO<sub>2</sub> Films on Silicon," *J. Phys. (Paris)*, Vol. 25, p. 66, Jan.-Feb. 1964.
7. A. G. Revesz, K. H. Zaininger, and R. J. Evans, "Interface States and Interface Disorder in the Si-SiO<sub>2</sub> System," *J. Phys. Chem. Solids*, Vol. 28, p. 197, Feb. 1967.
8. A. G. Revesz, "Thermal Oxidation of Silicon: Growth Mechanism and Interface Properties," *Phys. Stat. Sol.*, Vol. 19, p. 193, Jan. 1967.
9. Y. Miura and Y. Matukura, "Instabilities of MOS Structure," *Jap. J. Appl. Phys.*, Vol. 6, p. 582, May 1967.
10. A. G. Revesz, "Anodic Oxidation of Silicon in KNO<sub>3</sub>-N-Methylacetamide Solution: Interface Properties," *J. Electrochem. Soc.*, Vol. 114, p. 629, June 1967.
11. A. R. Plummer, "The Semiconductor-Gas and Semiconductor-Metal Systems," in *The Electrochemistry of Semiconductors*, edited by P. J. Holmes, Academic Press, London and New York, 1962.
12. A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, Inc., New York, 1967.
13. L. M. Terman, "An Investigation of Surface States at a Silicon/Silicon Oxide Interface Employing Metal-Oxide-Silicon Diodes," *Solid State Electronics*, Vol. 5, p. 285, Sept.-Oct. 1962.
14. K. Lehovec, A. Slobodskoy, and J. L. Sprague, "Field-Effect-Capacitance Analysis of Surface States on Silicon," *Phys. Stat. Sol.*, Vol. 3, p. 447, No. 3, 1963.

15. K. H. Zaininger and G. Warfield, "Limitations of the MOS Capacitance Method for the Determination of Semiconductor Surface Properties," *IEEE Trans. Electron Devices*, Vol. ED-12, p. 179, April 1965.
16. W. Shockley and G. L. Pearson, "Modulation of Conductance of Thin Films of Semiconductors by Surface Charges," *Phys. Rev.*, Vol. 74, p. 232, 1948.
17. M. M. Atalla, E. Tannenbaum, and E. J. Scheibner, "Stabilization of Silicon Surfaces by Thermally Grown Oxides," *Bell Syst. Tech. J.*, Vol. 30, p. 749, May 1959.
18. A. G. Revesz, "The Defect Structure of Grown Silicon Dioxide Films," *IEEE Trans. Electron Devices*, Vol. ED-12, p. 97, March 1965.
19. R. Williams, "Photoemission of Electrons from Silicon into Silicon Dioxide," *Phys. Rev.*, Vol. 140A, p. 569, Oct. 1965.
20. B. E. Deal, E. H. Snow, and C. A. Mead, "Barrier Energies in Metal-Silicon Dioxide-Silicon Structures," *J. Phys. Chem. Solids*, Vol. 27, p. 1873, Dec. 1966.
21. B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, "Characteristics of the Surface-State Charge ( $Q_{ss}$ ) of Thermally Oxidized Silicon," *J. Electrochem. Soc.*, Vol. 114, p. 266, March 1967.
22. E. H. Nicollian and A. Goetzberger, "The Si-SiO<sub>2</sub> Interface-Electrical Properties as Determined by the Metal-Insulator-Silicon Conductance Technique," *Bell Syst. Tech. J.*, Vol. 46, p. 1055, July-Aug. 1967.
23. K. H. Zaininger, "Automatic Display of MIS Capacitance versus Bias Characteristics," *RCA Rev.*, Vol. 27, p. 341, No. 3, Sept. 1966.
24. A. G. Revesz, "Taxonomy of Electric Charges in Metal-Insulator-Semiconductor Systems," *Proc. IEEE*, Vol. 54, p. 1002, July 1966.
25. J. R. Ligenza; (a), "Oxidation of Silicon by High-Pressure Steam," *J. Electrochem. Soc.*, Vol. 109, p. 73, Feb. 1962, and (b), "Effect of Crystal Orientation on Oxidation Rates of Silicon in High Pressure Steam," *J. Phys. Chem.*, Vol. 65, p. 2011, Nov. 1961.
26. J. R. Ligenza, "Silicon Oxidation in an Oxygen Plasma Excited by Microwaves," *J. Appl. Phys.*, Vol. 36, p. 2703, Sept. 1965.

27. J. Kraitchman, "Silicon Oxide Films Grown in a Microwave Discharge," *J. Appl. Phys.*, Vol. 38, p. 4323, Oct. 1967.
28. B. E. Deal and A. S. Grove, "General Relationship for the Thermal Oxidation of Silicon," *J. Appl. Phys.*, Vol. 36, p. 3770, Dec. 1965.
29. T. Nakayama and F. C. Collins, "Kinetics of Thermal Growth of Silicon Dioxide Films in Water Vapor-Oxygen-Argon Mixtures," *J. Electrochem. Soc.*, Vol. 113, p. 706, July 1966.
30. W. A. Pliskin, "Separation of the Linear and Parabolic Terms in the Steam Oxidation of Silicon," *IBM J. Res. and Dev.*, Vol. 10, p. 198, May 1966.
31. P. J. Burkhardt and L. V. Gregor, "Kinetics of the Thermal Oxidation of Silicon in Dry Oxygen," *Trans. Metallurgical Soc. of AIME*, Vol. 236, p. 299, March 1966.
32. A. G. Revesz and R. J. Evans, "Thermal Oxidation of Silicon as an Impurity-Controlled Process," *IEEE Trans. Electron Devices*, Vol. ED-14, p. 789, Nov. 1967.
33. A. G. Revesz and R. J. Evans, to be published.
34. J. T. Law, "The High Temperature Oxidation of Silicon," *J. Phys. Chem.*, Vol. 61, p. 1200, 1957.
35. A. G. Revesz, K. H. Zaininger, and R. J. Evans, "Deviations from Parabolic Growth in the Thermal Oxidation of Silicon," *Appl. Phys. Letters*, Vol. 8, p. 57, No. 3, Feb. 1966.
36. B. E. Deal, "The Oxidation of Silicon in Dry Oxygen, Wet Oxygen, and Steam," *J. Electrochem. Soc.*, Vol. 110, p. 527, June 1963.
37. R. J. Charles, "Static Fatigue of Glass," *J. Appl. Phys.*, Vol. 29, p. 1549, Nov. 1958.
38. M. Green and A. Liberman, "Kinetics and Mechanism of Oxygen Adsorption on Single Crystals of Germanium," *J. Phys. Chem. Solids*, Vol. 23, p. 1407, 1962.
39. K. Motzfeldt, "On the Rates of Oxidation of Silicon and Silicon Carbide in Oxygen, and Correlation with Permeability of Silica Glass," *Acta Chemica Scandinavica*, Vol. 18, p. 1596, No. 7, 1964.
40. P. J. Jorgensen, "Effect of an Electric Field on Silicon Oxidation," *J. Chem. Phys.*, Vol. 37, p. 874, Aug. 1962.
41. P. J. Jorgensen and F. J. Norton, "Oxygen Permeation through Fused SiO<sub>2</sub>," *VIIIth International Congress on Glass*, p. 310, Gordon and Breach, New York, 1966.

42. R. Haul and G. Dümbgen, "Untersuchung der Sauerstoffbeweglichkeit in Titandioxyd, Quarz und Quarzglas mit Hilfe des heterogenen Isotophenaustausches," *Zt. f. Elektrochemie*, Vol. 66, p. 636, 1962.
43. R. W. Lee, "On the role of hydroxyl in the diffusion of hydrogen in fused silica," *J. Phys. Chem. Glasses*, Vol. 5, p. 35, April 1964.
44. T. Drury, G. J. Roberts, and J. P. Roberts, "Diffusion of 'Water' in Silica Glass," *Advances in Glass Technology* (VI International Congress on Glass) Plenum Press, 1962, p. 249.
45. M. M. Atalla and E. Tannenbaum, "Impurity Redistribution and Junction Formation in Silicon by Thermal Oxidation," *Bell Syst. Tech. J.*, Vol. 39, p. 933, July 1960.
46. B. E. Deal and M. Sklar, "Thermal Oxidation of Heavily Doped Silicon," *J. Electrochem Soc.*, Vol. 112, p. 430, 1965.
47. A. G. Nassibian, "Effect of Diffused Oxygen and Gold on Surface Properties of Oxidized Silicon," *Solid-State Electronics*, Vol. 10, p. 879, 1967.
48. D. J. Thomas, "Surface Damage and Copper Precipitation in Silicon," *Phys. Stat. Sol.*, Vol. 3, p. 2261, 1963.
49. P. F. Schmidt and W. Michel, "Anodic Formation of Oxide Films on Silicon," *J. Electrochem Soc.*, Vol. 104, p. 230, April 1957.
50. A. G. Revesz, "Noncrystalline Structure and Electronic Conduction of Silicon Dioxide Films," *Phys. Stat. Sol.*, Vol. 24, p. 115, 1967.
51. A. Charlesby, "Ionic Current and Film Growth of Thin Oxide Layers on Aluminum," *Proc. Phys. Soc., Sec., B*, Vol. 66, p. 317, 1953.
52. W. Haas, "Silicon Dioxide as Dielectric in Solid Electrolyte Capacitors," *J. Electrochem. Soc.*, Vol. 109, p. 1192, 1962.
53. P. F. Schmidt and M. J. Rand, "Mutual Transformation of Thermal and Anodic  $\text{SiO}_2$  Films," *Solid State Communications*, Vol. 4, p. 169, 1966.
54. W. A. Pliskin and H. S. Lehman, "Structural Evaluation of Silicon Oxide Films," *J. Electrochem. Soc.*, Vol. 112, p. 1013, Oct. 1965.
55. W. R. Sinclair and F. G. Peters, "Preparation of Oxide Glass Films by Reaction Sputtering," *J. Am. Cer. Soc.*, Vol. 46, p. 20, 1963.

56. J. Klerer, "A Method for the Deposition of SiO<sub>2</sub> at Low Temperatures," *J. Electrochem Soc.*, Vol. 108, p. 1070, Nov. 1961.
57. S. W. Ing, Jr., and W. Davern, "Glow Discharge Formation of Silicon Oxide and The Deposition of Silicon Oxide Thin Film Capacitors by Glow Discharge Techniques," *J. Electrochem. Soc.*, Vol. 112, p. 284, March 1965.
58. N. Goldsmith and W. Kern, "The Deposition of Vitreous Silicon Dioxide Films from Silane," *RCA Rev.*, Vol. 28, p. 153, 1967.
59. W. Steinmaier and J. Bloem, "Successive Growth of Si and SiO<sub>2</sub> in Epitaxial Apparatus," *J. Electrochem. Soc.*, Vol. 111, p. 206, Feb. 1964.
60. A. G. Revesz and K. H. Zaininger, "The Influence of Oxidation Rate and Heat Treatment on the Si Surface Potential in The Si-SiO<sub>2</sub> System," *IEEE Trans. Electron Devices*, Vol. ED-13, p. 246, Feb. 1966.
61. J. F. Delord, D. G. Hoffman, and G. Stringer, "Use of MOS Capacitors in Determining the Properties of Surface States at the Si-SiO<sub>2</sub> Interface," *Bull. Am. Phys. Soc.*, Vol. 10, p. 546, 1965.
62. P. Balk, P. J. Burkhardt, and L. V. Gregor, "Orientation Dependence of Built-In Surface Charge on Thermally Oxidized Silicon," *Proc. IEEE*, Vol. 53, p. 2133, 1965.
63. P. V. Gray and D. M. Brown, "Density of SiO<sub>2</sub>-Si Interface States," *Appl. Phys. Letters*, Vol. 8, p. 31, 1966.
64. Y. Miura, "Effect of Orientation on Surface Charge Density of Si-SiO<sub>2</sub> Interface," *Jap. J. Appl. Phys.*, Vol. 4, p. 958, 1965.
65. P. T. Landsberg, "On the Logarithmic Rate Law in Chemisorption and Oxidation," *J. Chem. Phys.*, Vol. 23, p. 1079, June 1955.
66. P. Mark, "The Role of Surface States in the Photoelectronic Properties of Insulating Cadmium Sulfide Crystals," *RCA Rev.*, Vol. 26, p. 461, Dec. 1965.
67. S. R. Hofstein, "Stabilization of MOS Devices," *Solid-State Electronics*, Vol. 10, p. 657, 1967.
68. F. P. Heiman, "Thin-Film Silicon-On-Sapphire Deep-Depletion MOS Transistors," *IEEE Trans. Electron Devices*, Vol. ED-13, p. 855, Dec. 1966.
69. Y. Tarui, Y. Komiya, H. Teshima, and K. Nagai, "Surface States of Silicon-Silicon Dioxide Interface Grown by Vapor Deposition," *Jap. J. Appl. Phys.*, Vol. 5, p. 275, April 1966.

70. For example, E. Kooi, "Effects of Low-Temperature Heat Treatments on the Surface Properties of Oxidized Silicon," *Philips Res. Repts.*, Vol. 20, p. 578, Oct. 1965.
71. J. Olmstead, J. Scott, and P. Kuznetzoff, "Hydrogen-Induced Surface Space-Charge Regions in Oxide-Protected Silicon," *IEEE Trans. Electron Devices*, Vol. ED-12, p. 104, 1965.
72. G. Hetherington and K. H. Jack, "Water in vitreous silica Part I. Influence of 'water' content on the properties of vitreous silica," *Phys. Chem. Glass*, Vol. 3, p. 129, Aug. 1962.
73. D. R. Kerr, J. S. Logan, P. J. Burkhardt, and W. A. Pliskin, "Stabilization of  $\text{SiO}_2$  Passivation Layers with  $\text{P}_2\text{O}_5$ ," *IBM J. Res. Dev.*, Vol. 8, 376, Sept. 1964.
74. E. H. Snow, A. S. Grove, B. E. Eal, and C. T. Sah, "Ion Transport Phenomena in Insulating Films," *J. Appl. Phys.*, Vol. 36, p. 1664, 1965.
75. S. R. Hofstein, "A Model for the Charge Motion and Instability in the MOS Structure," *IEEE Trans. Electron Devices*, Vol. ED-13, p. 227, 1966.
76. D. R. Kerr, "Effect of Temperature and Bias on Glass-Silicon Interfaces," *IBM J. Res. Dev.*, Vol. 8, p. 385, Sept. 1964.
77. Y. Miura and Y. Matukura, "Instabilities of MOS Structure," *Jap. J. Appl. Phys.*, Vol. 6, p. 582, May 1967.
78. E. Yon, W. H. Ko, and A. B. Kuper, "Sodium Distribution in Thermal Oxide by Radio Chemical and MOS Analysis," *IEEE Trans. Electron Devices*, Vol. ED-13, p. 276, 1966.
79. For a review, see A. E. Owen, "Electric Conduction and Dielectric Relaxation in Glass," in *Progress in Ceramic Science*, Vol. 3, J. E. Burke, ed., pp. 77-179, The MacMillan Co., N.Y., 1963.
80. J. J. Lander and J. Morrison, "Low Energy Electron Diffraction Study of Silicon Surface Structures," *Annals of the New York Academy of Sciences*, Volume 101, Art. 3, p. 605, Jan. 1963.
81. E. Kooi, "The Surface Charge in Oxidized Silicon," *Philips Res. Repts.*, Vol. 21, p. 477, 1966.
82. S. Kubo and E. Ichinohe, "The Influence of Dissolved Oxygen in  $\text{SiO}_2$  on C-V Characteristics," *Jap. J. Appl. Phys.*, Vol. 6, p. 1072, Sept. 1967.
83. K. H. Zaininger, "Irradiation of MIS Capacitors with High-Energy Electrons," *IEEE Trans. Nuclear Science*, Vol. NS-13, p. 237, Dec. 1966.

84. J. J. Lander, "Chemisorption and Ordered Surface Structures," *Surface Science*, Vol. 1, p. 125, 1964.
85. J. Lindmayer, "Field Effect Studies of the Oxidized Silicon Surface," *Sol. State Electronics*, Vol. 9, p. 275, 1966.
86. A. M. Goodman, "Photoemission of Electrons from Silicon and Gold into Silicon Dioxide," *Phys. Rev.*, Vol. 144, p. 588, April 1966.
87. For example, A. N. Knopp and R. Stickler, "The Structure and Perfection of Thermally Grown Oxide Films on Silicon," *Electrochemical Technology*, Vol. 5, p. 37, Jan.-Feb. 1967.
88. L. Pauling, "Interatomic Distances and Bond Character in the Oxygen Acids and Related Substances," *J. Phys. Chem.*, Vol. 56, p. 361, March 1952.
89. J. Zarzycki, *Verres et Refractaires*, Vol. 11, p. 3, 1957.
90. D. W. J. Cruickshank, "The Role of 3d-Orbitals in  $\pi$ -Bonds between (a) Silicon, Phosphorus, Sulphur, or Chlorine and (b) Oxygen or Nitrogen." *J. Chem. Soc.*, p. 5486, 1961.
91. R. L. Myuller, "Solid State Chemistry and the Vitreous State," in *Solid State Chemistry*, ed. by Z. U. Borisova, Consultants Bureau, N.Y., 1966, pp. 1-33.
92. L. Young, *Anodic Oxide Films*, Academic Press, N.Y., 1961.
93. J. W. Faust, "Dislocations and the Nucleation of Oxidation of Ge," *Acta Metallurgica*, Vol. 11, p. 1077, Sept. 1963.
94. S. Zwerdling and S. Sheff, "The Growth of Anodic Oxide Films on Germanium," *J. Electrochem. Soc.*, Vol. 107, p. 338, April 1960.
95. R. W. Wales, "The Anodic Formation of Germanium Oxidic Films," *J. Electrochem. Soc.*, Vol. 110, p. 914, Aug. 1963.
96. D. S. Urch, "The Perbromate Problem," *J. Inorg. Nucl. Chem.*, Vol. 25, p. 771, 1963.
97. A. Politycki and E. Fuchs. "Elektronenmikroskopische Untersuchung elektrolytisch erzeugter Siliziumoxydschichten," *Zt. Naturforschung*, Vol. 142, p. 271, 1959.
98. A. G. Revesz, to be published.
99. S. W. Ing, Jr., R. E. Morrison, and J. E. Sandor, "Gas Permeation Study and Imperfection Detection of Thermally Grown and Deposited Thin Silicon Dioxide Films," *J. Electrochem. Soc.*, Vol. 109, p. 221, March 1962.
100. A. D. Lopez, "Fast Etching Imperfections in Silicon Dioxide Films," *J. Electrochem. Soc.*, Vol. 113, p. 89, Jan. 1966.

101. For a review, see, e.g., J. M. Stevels, in *Encyclopedia of Physics*, (S. Fluegge, ed.) Vol. 13, p. 510, Springer, Berlin, 1962.
102. L. W. Tilton, "Noncrystal Ionic Model for Silica Glass," *J. Res. National Bureau of Standards*, Vol. 59, p. 139, Aug. 1957.
103. H. A. Robinson, "On the Structure of Vitreous  $\text{SiO}_2$ —I. A New Pentagonal Dodecahedral Model," *J. Phys. Chem. Solids*, Vol. 26, p. 209, 1965.
104. W. Vogel, "Neue Strukturvorstellungen vom Glas und ihr Zusammenhang mit einigen physikalischen und chemischen Eigenschaften," *Phys. Stat. Sol.*, Vol. 14, p. 255, 1966.
105. B. D. Saksena, K. C. Agarwal, and G. S. Jauhri, "The Ring Band of Cyclo-Silicates," *Trans. Far. Soc.*, Vol. 59, p. 276, 1963.
106. T. Dunn, G. Hetherington, and K. H. Jack, "The High-Temperature Electrolysis of Vitreous Silica," *Phys. Chem. of Glasses*, Vol. 6, p. 16, Feb. 1965.
107. J. M. Stevels and A. Kats, "The Systematics of Imperfections in Silicon-Oxygen Networks," *Philips Res. Rep.*, Vol. 11, p. 103, 1956.
108. Y. Nishi, "Electron Spin Resonance in  $\text{SiO}_2$  Grown on Silicon," *Jap. J. Appl. Phys.*, Vol. 5, p. 333, 1966.

# MULTIPATH PROBLEMS IN COMMUNICATIONS BETWEEN LOW-ALTITUDE SPACECRAFT AND STATIONARY SATELLITES\*

BY

S. H. DURRANI† AND H. STARAS\*\*

*Summary*—The growing demand for continuous, wide-band communications between ground stations and low-altitude earth-orbiting spacecraft can be met through the use of synchronous-satellite repeaters, but the link between repeater and spacecraft suffers from earth-scattered multipath. This paper examines the occurrence of multipath in a typical system consisting of two synchronous satellites and a "target" spacecraft at orbit altitudes of 100 to 1,000 miles. The effects of target height and position, and of target and repeater antenna directivity are discussed. Special attention is paid to a retrodirective array repeater antenna, which focuses a narrow beam towards the target by sensing a pilot transmitted by the target.

Expressions are derived for the multipath parameters of interest, namely the average scattered power, fading margin, fading rate, and coherence bandwidth. The earth's sphericity and surface roughness are taken into account, and the resulting integrals are evaluated approximately by the method of steepest descent. Curves showing the multipath parameters at various spacecraft positions are obtained for spacecraft heights of 100 and 400 miles over land or sea, with vertical and horizontal polarizations. Methods of reducing the multipath interference by employing circular polarization, high antenna directivity at the spacecraft, or a narrow-band pilot filter in the repeater are discussed.

## INTRODUCTION

A LARGE number of space missions, both manned and unmanned, utilize low-altitude spacecraft orbiting at heights between 100 and 1000 miles above earth. At these low altitudes the spacecraft are unable to maintain continuous contact with the Mission Control Center because of line-of-sight constraints. It is conceivable that future systems will use high-altitude satellite repeaters to relay messages between the "target" spacecraft and ground when the target is not directly visible at the ground station. Although this ensures continuity of coverage, the link between target and repeater is subject to multipath interference produced by reflections from the earth.

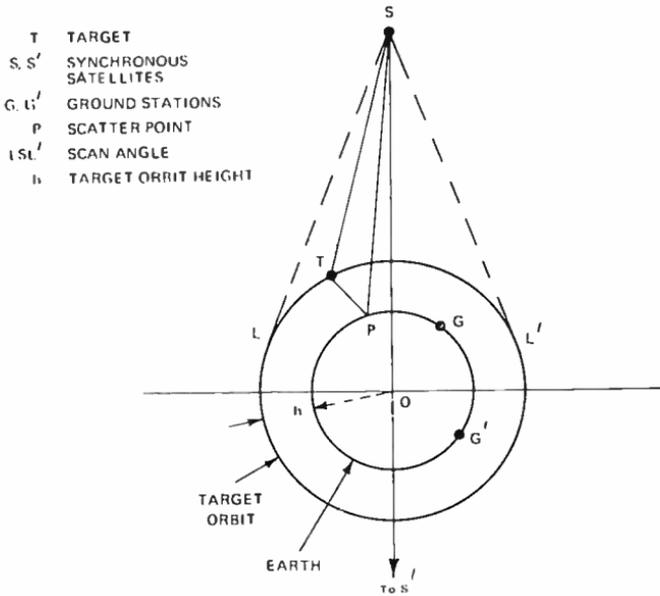
Figure 1 shows the typical situation, where a target T orbiting the

\* The work was done under a NASA contract (NASW-1447).

† RCA Astro Electronics Division, Princeton, N. J.

\*\* RCA Laboratories, Princeton, N. J.

earth at height  $h$  communicates with a ground station  $G$  through a repeater satellite  $S$ . (A second repeater  $S'$  and ground station  $G'$  enable complete coverage if the repeaters are at synchronous altitude and both repeater and target are in equatorial orbit.) The signal from  $T$  to  $S$  arrives along the direct path  $TS$  and the indirect path  $TPS$ ,



where  $P$  is the point of reflection or scatter on earth. The ratio of powers along the two paths is defined as  $K$ :

$$K = \langle P_s \rangle / P_d \quad (1)$$

where  $\langle P_s \rangle$  is the average scattered power received along the indirect path and  $P_d$  is the power received along the direct path.

Note that the scattered power  $P_s$  is a random variable and is described in terms of the statistical nature of the rough-earth scattering surface.

The magnitude of  $K$ , and hence the severity of multipath interference, depends on three factors:

1. The amount of power incident on the earth along  $TP$ , expressed as a function of the power  $P_d$ . This clearly is a function of the system geometry and the antenna pattern of the target  $T$ .

2. The amount of (average) power scattered in the direction PS per unit power incident along TP. For a smooth earth, P is the point of specular reflection, and the power ratio is the familiar "reflection coefficient." By analogy, the ratio for a rough earth may be termed the "average scatter coefficient." Its value depends on the frequency, polarization, and angle of incidence of the incoming wave, and of the electrical and statistical (roughness) properties of the earth surface.
3. The antenna pattern at the repeater, which determines the gain along direction SP relative to the gain along ST.

It is convenient to classify the first and third items as system effects, since they depend on the system geometry and antenna patterns, and to consider the second item as a propagation effect, depending primarily on the scatter phenomenon at a rough earth.

#### SYSTEM EFFECTS

In principle, the multipath problem can be eliminated by using extremely narrow antenna beams at both target and repeater and ensuring that they are properly aimed at each other, so that the scattered signal is highly suppressed. The order of suppression achieved in practice is investigated in this section.

#### *Target Antenna*

Multipath suppression is negligible if the target antenna has low gain or is omnidirectional, because the power radiated towards the earth is relatively large in these cases. A similar situation exists even with a high-gain antenna if the target loses attitude control and starts tumbling (which happened, e.g., in the Gemini 8 spacecraft); the antenna is effectively omnidirectional in such a case. Under these conditions the amount of multipath interference is a function of the earth's scattering properties and is considered in detail in later sections.

The multipath interference is reduced significantly with a moderate antenna gain. For any position of the target T in Figure 1, and assuming that the antenna points along TS, the suppression along TP depends on the antenna pattern corresponding to the angle STP, which may be called the directivity angle at T. This angle varies as the target moves in its orbit; it is  $0^\circ$  for the grazing position (when the line ST is tangential to earth) and  $180^\circ$  for the sub-orbital position (when the target is vertically below the repeater satellite, i.e., STO is a straight line). However, it is unlikely that the target will communicate with S from a grazing position, because another satellite (e.g., S') would have

taken over before that. Thus the smallest value of the directivity angle occurring for a given orbit—and hence the least suppression—depends on the fraction of orbit assigned to each repeater, and hence on the number of such repeaters. It also depends on the target antenna pattern, as indicated earlier. A typical case is analyzed by assuming two repeaters, and a target with a parabolic antenna of  $6^\circ$  beam width (27 dB gain).

### *Multipath Suppression by a Directional Antenna*

The ratio of power radiated in any direction  $\delta$  to the power along the axis may be defined as a suppression factor,  $F$ ,

$$F = 20 \log (E/E_0), \quad (2)$$

where  $E_0$  and  $E$  are the electric field intensities along the axis and at angle  $\delta$  from it. For a parabolic target antenna with a  $6^\circ$  beam, the standard pattern shows that the first side-lobe maximum occurs at  $\delta = \pm 12^\circ$  and is 24 dB below the maximum field. Actually, the suppression is found to be more than 24 dB over a part of the main lobe also. Thus  $F$  is less than  $-24$  dB for  $|\delta| \geq 6.6^\circ$  and  $F$  lies between 0 and  $-3$  dB for  $|\delta| \leq 3^\circ$ .

The next step is to determine the actual values of  $\delta$  occurring during the target orbit, recognizing  $\delta$  as the directivity angle discussed earlier.

The multipath geometry of Figure 1 is shown in more detail in Figure 2. The angle  $\delta$ , the target position angle  $\gamma$  (angle TOS), and the target aspect angle  $V$  (defined as the angle between TS and the target local vertical OT) are also shown in the figure. For a given height  $h$ , the angles  $V$  and  $\delta$  are related to the target's position angle  $\gamma$ ; the relation is sketched in Figure 3 for three selected heights.

Since the system has two repeaters, each repeater must cover half the orbit of T, i.e., the angle  $\gamma$  is limited to the range from 0 to  $\pm 90^\circ$ . From Figure 3, the range of  $\delta$  for this range of  $\gamma$  is found to be

$$\delta > 18^\circ \text{ for } h > 500 \text{ miles}$$

$$\delta \geq 9.5^\circ \text{ for } h = 200 \text{ miles}$$

$$\delta \geq 5^\circ \text{ for } h = 100 \text{ miles}$$

By interpolation of the last two curves,  $\delta \geq 6.6^\circ$  for  $h = 135$  miles. Combining this with the previous result that  $F < -24$  dB for  $|\delta| \geq 6.6^\circ$ , it is clear that the suppression due to antenna directivity is always more than 24 dB ( $F < -24$  dB) for target orbits of 135 miles or higher.

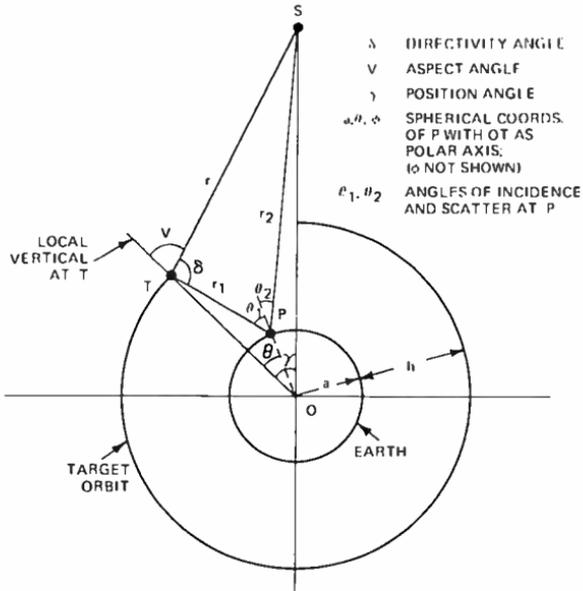


Fig. 2—Multipath geometry.

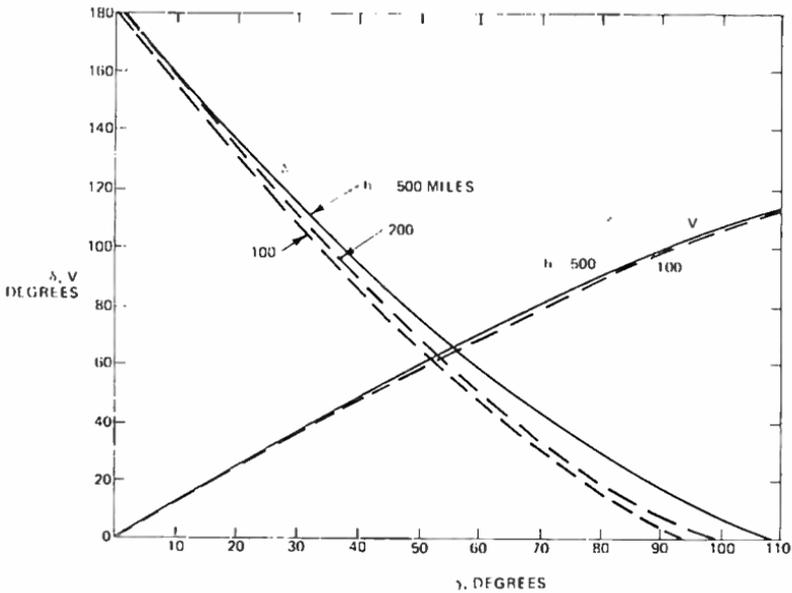


Fig. 3—Angular relations in multipath geometry.

Similar calculations for  $h = 100$  miles show that the suppression is between 9.4 and 24 dB for the range  $\gamma = 88.5^\circ$  to  $\gamma = 90^\circ$  (which is  $1.5 \cdot 90 = 1.7\%$  of the contact period) and more than 24 dB for the remaining period.

To summarize, a medium-gain antenna on the target is usually sufficient to suppress the power radiated towards earth to a very small value, so that the power arriving at the repeater along the indirect path is negligible. In particular, for a  $6^\circ$  beam width (27 dB gain) antenna, this holds for all target heights greater than 135 miles, and for a large fraction of the contact period for heights down to 100 miles, assuming two symmetrically located repeaters in both cases.

### **Repeater Antenna**

Although, in principle, a repeater antenna with reasonable directivity can also be effective in suppressing multipath, the values of gain available in practice are limited. This is seen by considering three representative types of antenna, namely fixed-reflector, steerable-reflector, and retrodirective array antennas.

Assuming that a fixed reflector (say a parabolic dish) is used at the repeater S, its beam width must be large enough to "see" the target in all positions of the target orbit (except when eclipsed by the earth). Thus, from Figure 1, the beam width must equal the scan angle  $LSL'$ . With such a wide beam, the antenna cannot discriminate between signals arriving along TS and PS; in fact, both lie within the scan angle and are received equally well.

A steerable reflector can be made to have a beam width narrower than the scan angle, but the antenna accepts interference when tracking the target over a region such that the line ST intersects the earth. For a low-altitude target, this region forms a major fraction of the total coverage zone. It can also be shown that for  $h \leq 1000$  miles, a satellite dish with beam width  $\geq 4.5^\circ$  will receive both direct and indirect signals for every visible position of the target; for  $h \leq 500$  miles, all beams wider than  $2.2^\circ$  suffer from the same drawback.

In practice, therefore, both fixed and steered-reflector antennas are subject to multipath interference irrespective of their gain, because they accept the indirect signal for at least a part of the target orbit.

The multipath effects in a retrodirective array are considered in more detail because of certain unusual features.

### **Multipath Effects Associated with a Retrodirective Array**

The retrodirective array (also called self-focusing or self-phasing array) operates by automatically forming a narrow beam directed to-

wards the target. The target initiates the action by radiating a pilot signal (usually from an omnidirectional antenna). Each individual element of the repeater receives the pilot with a certain phase and generates a signal whose phase is a conjugate of the incoming phase. The combined action of all elements results in a high gain in both receiving and transmitting modes. The received signals are all added in phase, and the transmitted signals are beamed towards the target with the full array gain. Moreover, once a high-gain beam has been formed at the repeater with the help of an essentially omnidirectional pilot transmitted by the target, the target can use a medium or high-gain antenna (say a monopulse type) that locks on to the beam; this enables the link to support wide-band traffic under normal conditions.

With certain restrictions, the total array gain  $G$  is approximately equal to the product of the individual element gain  $g$  and the number of elements  $N$ . Thus, a large  $G$  can be obtained by using a large  $N$ . However, the gain  $g$  is limited by the requirement that the individual element must be able to "see" the target throughout the scan angle (angle SLS' in Figure 1). This is the same requirement as for a fixed reflector, and the array is, therefore, equally subject to multipath.

The occurrence of multipath degrades the performance of a retrodirective array. Thus, the arrival of interfering signals along TS and TPS results in an improper phase distribution between elements, and the individual outputs do not add in phase as they would otherwise do. This reduces the effective array gain in the receiving mode. Other effects relate to the array's transmitting mode. Thus, the elements treat the two signals arriving along TS and PS as independent signals and automatically form two retrodirective beams pointing back to T and P (i.e., to the image of T, since the image lies on the extension of line SP). This has three disadvantages. First, the effective gain of the array in direction ST is reduced because a part of the satellite transmitter power is wasted in direction SP. Second, the two beams from S recombine at T to produce fading. Third, the target's monopulse antenna may not be able to lock on the proper beam (ST), with the result that wide-band traffic cannot be supported by the link between S and T.

To summarize, the multipath effects primarily depend on the target antenna directivity and the orbit geometry. The repeater antenna may be considered as being essentially omnidirectional in this context, in spite of its potentially high gain, because the steerable reflector or a retrodirective array is just as vulnerable to multipath as a wide-beam fixed reflector.

EVALUATION OF  $K$ 

The ratio  $K$  was defined in Equation (1) as  $\langle P_s \rangle / P_d$ . It is evaluated in this section by ignoring antenna directivities and considering only the propagation effects.

The average scattered power  $\langle P_s \rangle$  is found by integrating the average power  $P_n$  scattered from each elementary area of the surface. Thus

$$\langle P_s \rangle = \int P_n dS = a^2 \int \int P_n \sin \theta d\theta d\phi, \quad (3)$$

where  $a$  is the radius of earth and  $(a, \theta, \phi)$  are the spherical coordinates of an arbitrary scattering element on the earth's surface, as shown in Figure 2.

$P_n$  is a function of the frequency, polarization, and angle of incidence of the incoming wave, and of the electrical and statistical properties of the earth. It has been studied extensively in the literature; from a result by Beckmann (Reference (1), Equation 10, Section 12.4 and Equation 9, Section 11.2) with minor change of notation, it is given by

$$\frac{P_n}{P_d} = \frac{D^2 |R|^2}{4\pi} \frac{r^2 \cot^2 \beta_n}{r_1^2 r_2^2 \cos^4 \beta} \exp \left\{ -\frac{\tan^2 \beta}{\tan^2 \beta_n} \right\}, \quad (4)$$

The angles and distances involved are shown in Figure 4; the symbols are defined as follows:

$D$  = divergence factor to account for scattering from a curved instead of a plane surface

$$= \left\{ 1 + \frac{2r_1 r_2}{a(r_1 + r_2) \cos \theta} \right\}^{-1/2} \left\{ 1 + \frac{2r_1 r_2}{a(r_1 + r_2)} \right\}^{-1/2},$$

$R$  = Fresnel reflection coefficient for a plane surface,

$\tan \beta_n = 2\sigma T$  ( $\sim$  rms slope of surface undulations),

$$\tan^2 \beta = \frac{\sin^2 \theta_1 + \sin^2 \theta_2 - 2\sin \theta_1 \sin \theta_2 \cos \phi}{(\cos \theta_1 + \cos \theta_2)^2}.$$

It is to be noted that  $\sigma$  is the rms "bump" height and  $T$  is the corre-

<sup>1</sup> P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, The Macmillan Co., New York, 1963.

lation length of the surface fluctuations, so that  $\sigma/T$  is a measure of the rms slope of the undulations. The reflection coefficient  $R$  is to be taken as  $R_v$  or  $R_h$ , corresponding to vertical or horizontal polarization of the incoming wave.

Equation (4) holds under the assumptions that (1) the surface undulations can be described by a two-dimensional Gaussian distribution; (2)  $\sigma/T$  is fairly small; (3)  $\sigma/\lambda$  is fairly large (where  $\lambda$  is the

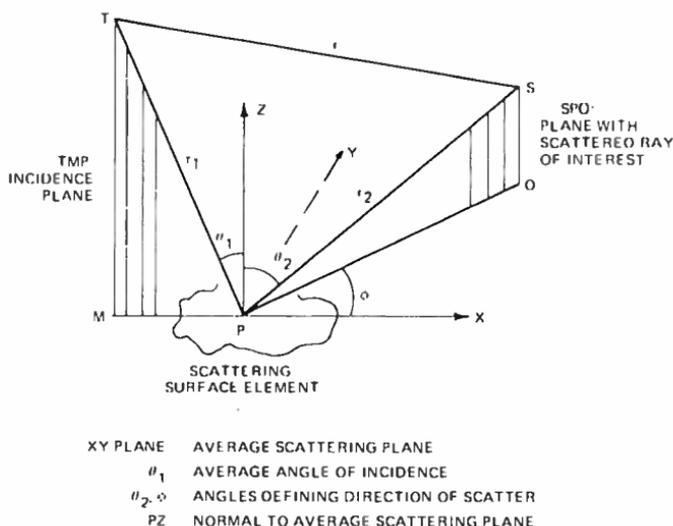


Fig. 4—Scattering geometry.

wavelength); and (4) the autocorrelation of the surface fluctuations is an analytic function. From assumption (2) it follows that the Kirchhoff-Huygen approximation is valid and that shadowing and multiple scattering effects are unimportant.

The evaluation of  $\int P_n dS$  in a closed form is not possible because of the complicated form of  $P_n$  in Equation (4). However, the method of steepest descent can be used (Reference (2)) to obtain an approximate solution if the parameter  $\tan \beta_n$  ( $\sim$  rms slope of rough surface) is sufficiently small. This method has been applied to the case of flat earth (Reference (3)); the present case of a spherical earth is similar except for the presence of the divergence factor  $D$ . Details of the evaluation are given in the Appendix.

<sup>2</sup> H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, p. 503; Cambridge University Press; 1950.

<sup>3</sup> H. Staras, "Rough Surface Scattering on a Communication Link," *Radio Science*, (to be published in June 1968).

The procedure requires the locating of the stationary points of  $\tan^2 \beta$ , i.e., the simultaneous solutions of

$$\frac{\partial \tan^2 \beta}{\partial \phi} = 0, \quad (5a)$$

$$\frac{\partial \tan^2 \beta}{\partial \theta} = 0. \quad (5b)$$

It can be shown that if  $\sigma T$  is small and the target height  $h$  is small compared to the earth radius  $a$ , then the solution of Equation (5) is

$$\phi = 0, \theta_1 = \theta_2, \text{ if } h/a \ll 1. \quad (6)$$

As seen from Figure 2, this condition implies that the stationary point is the point of specular reflection. Thus, the average scattered power comes primarily from a region surrounding the specular point, provided the surface undulations are gentle and the target height is small. The size of this region is determined by the rms slope  $\sigma T$  rather than by Fresnel zone considerations (as it is for a smooth reflector).

According to Equation (4), the value of  $P_s$  near the stationary point depends on the value of  $\tan^2 \beta$  in that region. It can be shown that, with Equation (6),

$$\tan^2 \beta = (\phi^2 + 4) \tan^2 B + M^2 (\theta - \theta_s)^2, \quad (7)$$

where  $B = V - \theta_s$ ,

$V =$  target aspect angle,

$\theta_s =$  value of  $\theta$  at the stationary point,

$a, \theta, \phi =$  coordinates of P,

$$M = 1 + \left( \frac{1}{2} \right) \frac{H^2}{(H^2 + H \theta_s^2)^{1/2} (H^2 + (1 + H) \theta_s^2)},$$

$H = h/a$ .

Equations (7), (4), and (3) can be combined to evaluate  $\langle P_s \rangle$ . The result is

$$K = \langle P_s \rangle / P_d = D^2 |R|^2 Q, \quad (8)$$

where  $D$  and  $R$  have been defined previously and  $Q$  is defined by

$$Q = \frac{\tan A}{\tan B} \left[ 1 - H \tan^2 A + \frac{4\theta_s}{\sin 2A} \right]^{-1} \quad (9)$$

$$A = V - 2\theta_s, \quad B = V - \theta_s.$$

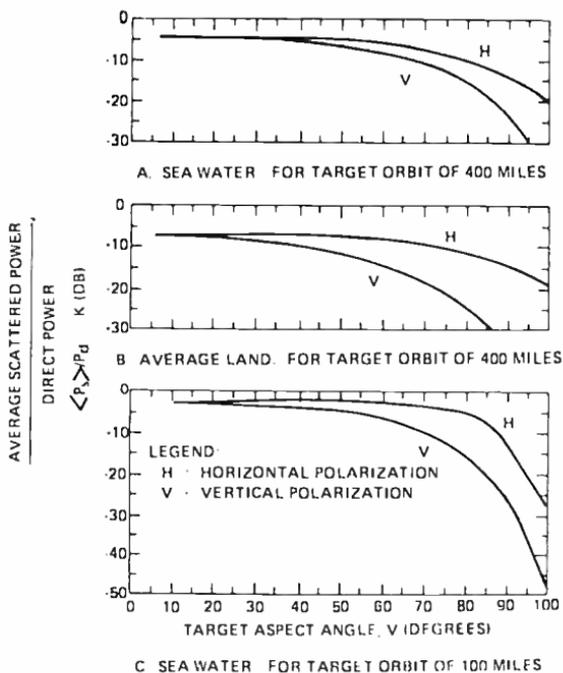


Fig. 5—Average scattered power (normalized) received at the repeater.

Note that  $D$ ,  $R$  and  $Q$  must be evaluated at the specular point ( $\theta = \theta_s$ ) defined by Equation (6).

The values of  $K$  computed from Equation (8) are plotted against  $V$  in Figure 5 for different terrains (scatter over sea water and average land) and different target heights (100 and 400 miles). From Figure 2 it should be noted that  $V$  can be larger than  $90^\circ$ ; in fact  $V = 115.5^\circ$  for the grazing position at  $h = 400$  miles and  $V = 103^\circ$  for grazing at  $h = 100$  miles.

Figure 5 shows that for a wide range of the angle  $V$ , the average scattered power is comparable to the power in the direct signal. It

also shows that the scattered signal power decreases as the altitude of the target increases.

Another interesting result seen from Figure 5 is the decrease in  $K$  as  $V$  approaches its largest values (near grazing). This is fortunate, because the target antenna directivity cannot easily be used to discriminate against the ground scattered signal for positions near grazing (when paths TS and TPS tend to merge). Thus the scattering phenomenon itself discriminates against the ground-scattered signal for angles near grazing, but at other angles the scattered signal would normally be quite large and special measures would be needed to discriminate against it.

#### FADING MARGIN

The fading margin is the additional power needed at the transmitter to overcome multipath fading at the receiver. The value can be readily determined when multipath occurs due to reflection from a smooth surface. For this case, the direct and indirect signals are

$$s_1(t) = \sin(\omega_1 t + \phi_1), \quad (10)$$

$$s_2(t) = A \sin(\omega_2 t + \phi_2). \quad (11)$$

Here the levels have been normalized with respect to the direct signal, the frequencies  $\omega_1$  and  $\omega_2$  differ because of the different Doppler shifts along the two paths, and the phase difference  $\phi_1 - \phi_2$  includes the difference in the path lengths as well as the phase shift due to reflection. The resultant signal has an amplitude varying between  $1 - A$  and  $1 + A$ . The fading depth then corresponds to the minimum threshold level  $T_L = 1 - A$ , and the fading margin, in dB, is

$$M_f = -20 \log_{10} T_L. \quad (12)$$

When the reflection occurs from a rough earth, the indirect signal of Equation (11) must be replaced by the sum of a number of scattered components:

$$s_2(t) = \sum_m A_m \sin(\omega_{2m} t + \phi_{2m}), \quad m = 1, 2, \dots \quad (13)$$

It is well known that the scattered components have a Rayleigh distribution of amplitudes and uniform distribution of phases.

To determine the fading depth and fading margin, it is necessary to find the minimum level  $T_L$  to which the resultant of  $s_1(t)$  and  $s_2(t)$  fades. Because of the statistical nature of  $s_2(t)$ ,  $T_L$  can only be de-

scribed in statistical terms; it is usual to define the fading margin in terms of the level  $T_L$  that is exceeded for 99% of the time (i.e., with a 99% probability).

The desired values of  $T_L$  in terms of the power ratio  $K$  are available from a table by Norton et al<sup>4</sup>, and are reproduced in Table I for the case of 99% probability.

Table I—Power Level  $T_L$  Exceeded by the Sum of Steady and Random Signals with 99% Probability.

$K^*$ (dB)	$T_L$ (dB)	$K$ (dB)	$T_L$ (dB)
-40	-0.14	0	-15.66
-35	-0.26	2	-15.24
-30	-0.46	4	-14.25
-25	-0.83	6	-12.89
-20	-1.53	8	-11.29
-18	-1.97	10	-9.54
-16	-2.54	12	-7.70
-14	-3.30	14	-5.80
-12	-4.34	16	-3.87
-10	-5.77	18	-1.91
-8	-7.81	20	-0.06
-6	-10.57		
-4	-13.44		
-2	-15.18		

\*  $K$  = Power in Rayleigh distributed random signal. Both  $K$  and  $T_L$  are expressed in dB relative to power in steady signal.

Recalling from Figure 5 that  $K$  is a function of target position (among other factors), it follows that the fading margin varies as the target moves in its orbit. The highest margin is needed for the highest  $K$ , which occurs for  $V = 0$ , i.e., when the target is directly beneath the repeater. The worst case for typical terrains is  $K = -4$  dB; from Table I, a fading margin of 13.44 dB is adequate for 99% of the time.

#### FADING BANDWIDTH

Reference to Equations (10) and (11) for a smooth earth shows

<sup>4</sup> K. A. Norton et al, "Probability Distribution of the Amplitude of a Constant Vector Plus a Rayleigh Distributed Vector," *Proc. IRE*, p. 1354; Oct. 1955.

that the resultant signal fades at a rate  $\omega_1 - \omega_2$  (beat frequency). For a rough earth, Equation (11) is replaced by Equation (13) and there are two types of beat frequencies, first, between the direct signal and individual components of the scattered signal and, second, between different components of the scattered signal itself. The fading, therefore, occurs at various rates (instead of a single rate) and it is necessary to define a fading bandwidth  $B_f$  that ranges from zero to a frequency equal to the highest beat frequency of the two types described above.

The highest beat frequency of the second type (between different components of the scattered signal) is termed the scattered signal band spread,  $B_s$ , and is found to be quite large (several kHz). The highest beat frequency of the first type is much smaller than  $B_s$ , because the (doppler-shifted) frequency of the direct signal lies nearly in the center of the (doppler shifted) band  $B_s$ . Thus the fading bandwidth  $B_f$  is nearly equal to the band spread  $B_s$  of the scattered signal.

To calculate  $B_s$ , it is noted that the power spectrum and the time autocorrelation functions are Fourier transforms, and that  $B_s \approx (2\tau_n)^{-1}$  where  $\tau_n$  is the time  $\tau$  for which the autocorrelation function  $C(\tau)$  falls to  $1/e$  of its maximum value  $C(0)$ .  $C(\tau)$  is found from the definition

$$C(\tau) = \langle E(t)E^*(t + \tau) \rangle \quad (14)$$

where  $E(t)$  is the scattered signal at time  $t$  and  $E^*(t + \tau)$  is its conjugate at time  $(t + \tau)$ . Changing the notation of Equation (13) and replacing it by its exponential form,

$$E(t) = \sum_i E_i \exp(j\phi_i) \quad (15a)$$

where  $\phi_i$  are random relative phases. It follows that

$$\langle E(t)E^*(t) \rangle = \sum_i E_i^2 \quad (15b)$$

which represents the average power  $P_n$  scattered by an elementary surface. Again,

$$E(t + \tau) = \sum_i E_i \exp j(\phi_i t + \Delta\phi_i) \quad (16)$$

where  $\Delta\phi_i$  is a systematic (not random) phase change introduced by

the motion of the target vehicle. Substituting Equations (15) and (16) in (14),

$$C(\tau) = \sum_i E_i^2 \exp(-j\Delta\phi_i). \quad (17)$$

Using continuous functions instead of discrete variables, this becomes

$$C(\tau) = \int P_n \exp\{-jk(R_n - R_n)\} dS. \quad (18)$$

where  $k = 2\pi/\lambda$  is the wave number and  $R_n - R_n$  is the change (in time  $\tau$ ) in the distance between the target vehicle and an element of the scattering surface. It is assumed that the synchronous satellite's position relative to the scattering surface does not change in the period  $\tau$ .

Equation (18) shows that the evaluation of  $C$  is very similar to the evaluation of  $K$  in the previous section, and can be carried out by the method of steepest descent. Since multiplicative constants cancel out in evaluating  $C(\tau)/C(0)$ , it is enough to replace  $P_n$  by

$$P_n \sim \exp\{-\tan^2 \beta / \tan^2 \beta_0\} \\ \sim \exp\left\{-\frac{T^2}{4\sigma^2} [(\phi^2/4) \tan^2 B + M^2 (\theta - \theta_s)^2]\right\}, \quad (19)$$

where use has been made of Equations (4) and (7).

The other quantity needed for Equation (18) is the differential distance  $R_n - R_n$ , which depends on the speed and direction of motion of the target vehicle. Figure 6 shows two cases: Case I, where the target velocity vector  $TT'$  lies in the great circle plane (plane  $XZ$ ) containing the target vehicle and the synchronous satellite, and Case II, where the target velocity is perpendicular to the great-circle plane in question. Then it can be shown that

$$R_n - R_n \sim b \sin(\theta_1 - \theta) \cos \phi, \quad (\text{Case I}) \quad (20a)$$

or

$$R_n - R_n \sim b \sin(\theta_1 - \theta) \sin \phi, \quad (\text{Case II}) \quad (20b)$$

where  $b = v\tau$  is the distance  $TT'$  traveled by the target in time  $\tau$ , and  $v$  ( $\approx 8$  km/sec for low orbits) is the speed of the target vehicle relative to earth.

From Equation (19), it is observed that  $P_n$  is substantially different from zero only when  $\phi$  is near zero (perhaps on the order of  $\pm 10^\circ$ );

therefore  $\cos \phi \approx 1$  and  $\sin \phi \approx \phi$ . With these approximations, inserting Equations (19) and (20) into Equation (18) gives the desired  $C(\tau)$ . As shown in the Appendix, the results for the two cases using angles  $A, B$  defined in Equation (9) are

$$\text{Case I: } |C(\tau)| \sim \exp - \left\{ \frac{2kv\tau\sigma}{T} \frac{(1 - H \tan^2 A) \cos A}{1 - H \tan^2 A + (4 \theta_x / \sin 2A)} \right\}^2 \quad (21)$$

$$\text{Case II: } |C(\tau)| \sim \exp - \left\{ \frac{2kv\tau\sigma \sin A}{T \tan B} \right\}^2. \quad (22)$$

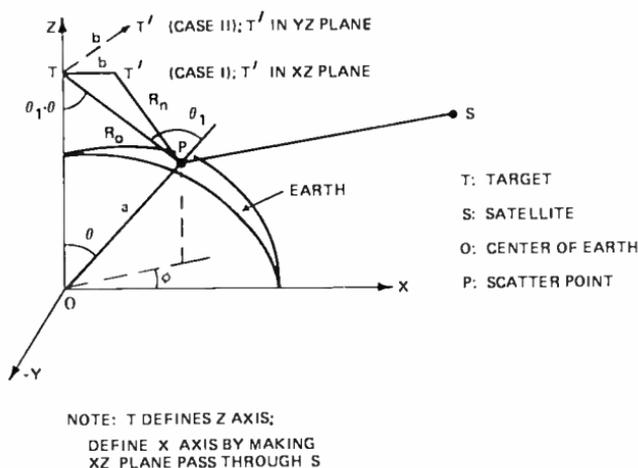


Fig. 6—Geometry for calculation of differential path length.

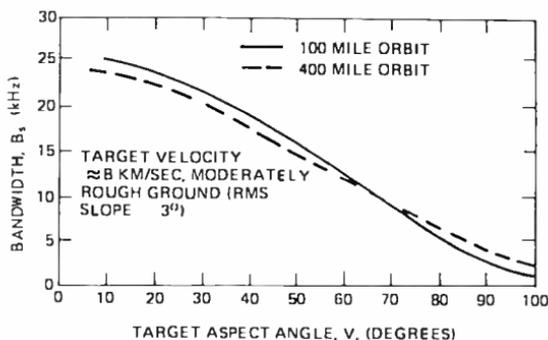
The value of  $\tau = \tau_0$  for which  $|C(\tau)|$  falls to  $1/e$  of its maximum value is readily found in each case. Then using  $B_x = (2\tau_0)^{-1}$ , the results are

$$\text{Case I: } B_x = kv \left( \frac{\sqrt{2}\sigma}{T} \right) \frac{(1 - H \tan^2 A) \cos A}{1 - H \tan^2 A + (4 \theta_x / \sin 2A)} \quad (23)$$

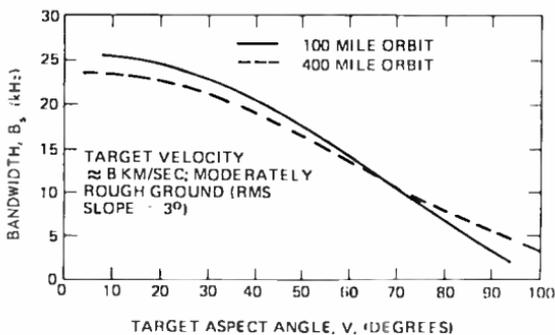
$$\text{Case II: } B_x = kv \left( \frac{\sqrt{2}\sigma}{T} \right) \frac{\sin A}{\tan B}. \quad (24)$$

Figure 7 shows the values of  $B_x$  found from these equations using  $kv = 4 \times 10^3$  sec,  $\sigma/T = 1, 20$  and target heights of 100 and 400 miles.

Since  $v \approx 8$  km/sec for these heights, the curves hold for  $k = 0.5/cm$ , i.e., a frequency of 2.4 GHz. Note that  $\sigma/T = 1/20$  implies an rms slope of  $3^\circ$  at the earth's surface. The graphs show that the scattered signal has a band spread varying from nearly 25 kHz to 3 kHz for these



A. CASE I. TARGET ORBIT IN THE PLANE OF THE TARGET, SATELLITE, AND EARTH CENTER.



B. CASE II. TARGET ORBIT PERPENDICULAR TO THE PLANE OF THE TARGET, SATELLITE, AND EARTH CENTER

Fig. 7—Bandwidth of scattered signal.

parameters. The differential doppler between the direct signal and the center of the scattered band is found to vary between 0 and a few kHz, depending on the target altitude. Thus the fading bandwidth  $B_f$  is nearly equal to the band spread  $B_s$  (as stated earlier) and is given by the curves of Figure 7.

#### COHERENCE BANDWIDTH

The coherence bandwidth in a multipath environment is defined as the frequency separation between two transmitted signals for which

the correlation between the received signal envelopes falls to a specified fraction of the maximum (usually 1 *e* or 1/2). For two signals  $E(\omega)$  and  $E(\omega + \Omega)$  with frequency separation  $\Omega$ , the correlation is defined as

$$C(\Omega) = \langle E(\omega) E^*(\omega + \Omega) \rangle. \quad (25)$$

Thus the coherence bandwidth  $B_c$  is

$$B_c = \frac{\Omega_c}{2\pi} \quad (26)$$

where  $\Omega_c$  is such that

$$|C(\Omega_c) - C(0)| = 0.5.$$

The evaluation of  $C(\Omega)$  is similar to that of  $C(\tau)$  in the previous section. Thus, assuming that

$$E(\omega) = \sum_i E_i(\omega) \exp(j\phi_i), \quad (27a)$$

it follows that

$$\begin{aligned} E(\omega + \Omega) &= \sum_i E_i(\omega + \Omega) \exp\{j(\phi_i + \Delta\phi_i)\} \\ &\approx \sum_i E_i(\omega) \exp\{j(\phi_i + \Delta\phi_i)\} \end{aligned} \quad (27b)$$

where the phase shift  $\Delta\phi_i$  is due to the frequency separation  $\Omega$  and the path length  $L_i$  along the *i*th path. It has been assumed that  $E_i$  does not vary appreciably with frequency; this is true if  $\Omega$  is small compared to  $\omega$ , but breaks down if  $\Omega$  tends to infinity. Combining Equations (25) and (27) gives

$$C(\Omega) = \sum_i E_i^2(\omega) \exp(j\Delta\phi_i), \quad (28)$$

which tends in the limit to

$$C(\Omega) = \int P_n \exp(-j\Delta\phi_i) dS. \quad (29)$$

The evaluation is similar to the evaluation of  $K$  and  $C(\tau)$ . Thus  $P_n$  is

replaced by Equation (19), and  $\Delta\phi_i$  is expressed in terms of its variations close to the stationary point  $P_1$ :

$$\Delta\phi_i = (\Omega/c) L_i, \quad (30a)$$

where  $c$  is the velocity of light and

$$\begin{aligned} L_i &= TP + PS \text{ for the } i\text{th path} \\ &= TP_1 + P_1S + g(\theta, \phi) \end{aligned} \quad (30b)$$

where  $g$  is a function of the coordinates  $\theta, \phi$  of the  $i$ th scattering element near the stationary point  $P_1$ . It is shown in the Appendix that

$$g(\theta, \phi) = \frac{a^2 \sin^2 \theta_s \phi^2 + a^2 (\theta - \theta_s)^2}{2 TP_1}, \quad (31)$$

and the evaluation of Equation (29) with the help of Equations (19), (30), and (31), and ignoring the constant of multiplication, leads to the result

$$C(\Omega) = I_1 I_2 \quad (32)$$

where

$$\begin{aligned} I_1 &= \left\{ -j \frac{\Omega}{c} \frac{a^2 \sin^2 \theta_s}{2 TP_1} - \frac{T^2}{16\sigma^2} \tan^2 B \right\}^{-1/2} \\ I_2 &= \left\{ -j \frac{\Omega}{c} \frac{a^2}{2 TP_1} - \frac{T^2}{4\sigma^2} M^2 \right\}^{-1/2}. \end{aligned}$$

The expression for  $I_1$  holds in general, but must be replaced by a constant in two special cases, namely, when  $\theta_s = 0$  (i.e., the target is below the repeater, or in the sub-orbital position) and when  $B = 90^\circ$  (i.e., the target is at the grazing position, so that the lines TS and TPS coincide). For these cases  $C(\Omega)$  is proportional to  $I_2$  alone, and the value of  $\Omega_n$  for which  $|C(\Omega_n)/C(0)| = 1/2$  is found relatively easily.

The highest coherence bandwidth occurs for  $\theta_s = 0$ , i.e., the sub-orbital position, and is found to be

$$B_c = 7.2 \left( \frac{T}{\sigma} \right)^2 \frac{(1 + 2H)^2}{2H}. \quad (33)$$

With  $T/\sigma = 20$  (i.e.,  $3^\circ$  rms slope), the results are  $B_r = 63.4$  kHz for a target height of 100 miles and 20.7 kHz for 400 miles. The corresponding values for the grazing position are found to be on the order of 2 kHz.

The coherence bandwidth is a measure of the traffic bandwidth that can be supported in the presence of multipath; the use of a wider band will result in the occurrence of uncorrelated fades over parts of the bandwidth. The preceding results show that the bandwidth is restricted to a few kHz for the target heights and terrain parameters assumed.

#### METHODS OF SUPPRESSING MULTIPATH

The occurrence of multipath is undesirable because it requires extra power in the form of a fading margin, degrades reception, and imposes coherence bandwidth limitations. Three methods of suppressing multipath interference are considered—utilizing the effects of antenna directivity, polarization, and narrow-band filtering.

##### *Antenna Directivity*

As discussed earlier, directivity of the target antenna plays an important role in reducing multipath. For a  $6^\circ$  beam width, the suppression is more than 24 dB (i.e.,  $F < -24$  dB) for target orbits above 135 miles, assuming a system with two repeater satellites. From Figure 5, the worst-case value of  $K$  is  $-4$  dB; this occurs for small values of the angle  $V$ , when  $F$  is found to be  $< -30$  dB. For those  $V$  (nearly  $90^\circ$ ) that have the least suppression ( $F = -24$  dB), the value of  $K$  is seen to be  $-10$  dB or less. Thus, the total suppression is more than 34 dB and the multipath is reduced to a negligible level.

From Table I, the fading margin for  $K = -4$  dB is 13.4 dB but for  $K = -34$  dB it is only 0.3 dB. Thus target antenna directivity enables a 13.1 dB saving in power requirements.

##### *Circular Polarization*

The reversal of sense of rotation occurring on reflection from the earth can be exploited to reduce multipath interference, because the direct signal suffers no reversal while the indirect signal does. The amount of suppression of the indirect signal depends on the reflection (or scatter) coefficient  $K$ . Thus, it is assumed that the target and repeater antennas are circularly polarized and the signal falling on earth has equal components of vertical and horizontal polarization. Then the

reflected signal will have components of magnitudes determined by the respective values of  $K$ , and the sense of rotation will be opposite to that of the direct signal. If  $K$  for the two polarizations differs by 0.5 dB or less, the receiving antenna rejects the indirect signal by 25 dB or more compared to the direct signal; the rejection is 3 dB if the vertical polarization is suppressed (at Brewster's angle), because then the reflected signal is linearly polarized while the direct signal is circularly polarized.

The suppression achieved by this technique is shown in Figure 8, which holds for a target orbit altitude of 400 miles above a moderately rough sea. The curves of  $K$  for vertical and horizontal polarization are reproduced from Figure 5 and are used to determine the additional suppression at each target position. The third curve gives the effective value of  $K$  after including the rejection due to polarization; it peaks at  $K = -13.5$  dB instead of  $-4$  dB as before. The fading margins for these values of  $K$  are 3.6 dB and 13.4 dB, respectively—a saving of 9.8 dB.

### *Narrow-Band Filtering*

The success of the two preceding methods depends on the proper orientation of the target and repeater antennas. This may be impractical if the target is not under proper control (for instance, in case of equipment malfunction). The only recourse in such cases is to use narrow-band filters in each receiver, and thereby reject a part of the scattered signal power. The rejection can be appreciable because the typical band spread is quite large; as seen from Figure 7,  $B_s$  varies from 3 kHz to 25 kHz. With a filter bandwidth  $b$ , the ratio of indirect to direct signal powers is effectively reduced from  $K$  to  $Kb/B$ . The reduced fading margin needed for this reduced interference can be determined from Table I. The results are sketched in Figure 9 for various values of  $b/B$ . The saving in fading margin achieved with a given filter is read as the difference between the ordinates for the top curve ( $b \geq B$ ) and the curve for the appropriate  $b$ . More details are available in Reference (5).

This method is particularly useful in reducing the multipath interference in a system using a retrodirective array at the repeater. As described earlier, a retrodirective array operates by forming a very narrow beam in the direction of a pilot signal received from the target.

---

<sup>5</sup> S. H. Durrani, "Rejection of Multipath Interference in Satellite Communications by Use of Narrowband Filters," *IEEE Trans. on AES*, Jan. 1968.

Since the pilot channel can be made extremely narrow, the earth-scattered signal received by the array can be virtually eliminated, and the degradation in array operation associated with multipath can be avoided. The saving in fading margin is important if the target has a limited amount of power that must be shared by the pilot and communication channels. The restoration of array operation is more significant; thus when the pilot has been "cleaned" of its interference, the array operates with its full gain directed towards the target and suppresses radiation towards the earth. The target's monopulse antenna can then lock on the main beam, thus establishing high antenna gains at both ends of the link and permitting the flow of high-bandwidth traffic on the link.

### CONCLUSION

Multipath interference due to earth-scattered signals can have serious effects on the communications link between a low-altitude earth-orbiting target vehicle and a synchronous-altitude repeater satellite. For the typical case of a target height of 100 miles above a moderately rough sea, the fading margin of 13.4 dB is required to compensate fading with 99% probability. The fading does not occur at a single rate but contains a range of beat frequencies, because of the presence of several components in the scattered signal. The highest beat frequency is called the fading bandwidth and is found to vary between 3 kHz and 25 kHz for different positions of the target in its orbit. The coherence bandwidth, which determines the useful traffic bandwidth, varies from about 2 kHz to 63 kHz (for a 400 miles target the maximum coherence bandwidth is reduced to about 20 kHz).

The effects of multipath cannot be appreciably reduced by using narrow-beam antennas at the repeater, because both the direct and indirect (earth-scattered) signals from a low-altitude target appear to arrive at the repeater along the same direction. This holds for both reflector and retrodirective types of antennas. The effects can be reduced by using target-antenna directivity, circular polarization, and narrow-band filtering techniques. A typical saving of 13 dB in fading margin can be achieved by the first method; typical values for the second and third methods are shown in Figures 8 and 9. The first two methods require control of antenna orientation on the target, which may be impractical in an emergency. The last method is applicable to this case; it has special value for a system using a retrodirective array at the repeater because it enables the full array gain to be achieved (and wide-band traffic to be carried) by suppressing the multipath effects in a narrow-band pilot channel.

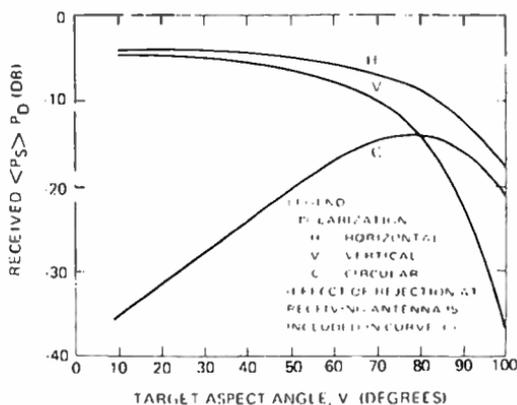


Fig. 8—Multipath rejection by using circular polarization.

#### APPENDIX—EVALUATION OF INTEGRALS

##### Evaluation of $K$

This appendix supplies some details for determining the stationary point of the expression for  $P_o$  (Equation (4)) and evaluating  $K$  (Equation (8)).

It is seen from Equation (4) that  $P_o$  is largest near  $\beta = 0$ , and its variation there depends primarily on the exponential term involving  $\tan^2 \beta$ , while the term  $\cos^4 \beta$  varies more slowly. Thus the stationary point is defined by those values of  $\theta$  and  $\phi$  for which the partial deriva-

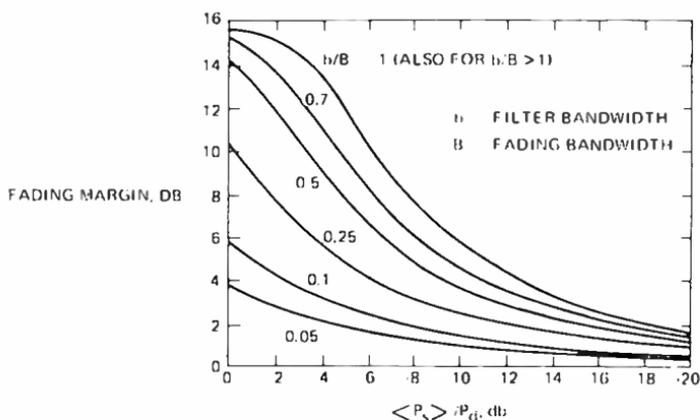


Fig. 9—Fading margin versus  $K$  for various filter bandwidths.

tives of  $\tan^2 \beta$  with respect to  $\theta$  and  $\phi$  vanish; i.e., we must find the simultaneous solutions of

$$\frac{\partial \tan^2 \beta}{\partial \phi} = 0 \quad \text{and} \quad \frac{\partial \tan^2 \beta}{\partial \theta} = 0 \quad (34)$$

where

$$\tan^2 \beta = \frac{\sin^2 \theta_1 + \sin^2 \theta_2 - 2 \sin \theta_1 \sin \theta_2 \cos \phi}{(\cos \theta_1 + \cos \theta_2)^2}.$$

The solution of the first of the above equations is  $\phi = 0$  and  $\phi = \pi$ . With  $\phi = \pi$ , it is not possible to find a solution to the second equation; however, a solution can be found with  $\phi = 0$ . Specifically, the second equation becomes

$$\frac{\partial \tan^2 \beta}{\partial \theta} = \frac{\partial}{\partial \theta} \left( \frac{\sin \theta_1 - \sin \theta_2}{\cos \theta_1 + \cos \theta_2} \right)^2 = 0, \quad (35)$$

or

$$\frac{\partial}{\partial \theta_1} \left( \frac{\sin \theta_1 - \sin \theta_2}{\cos \theta_1 + \cos \theta_2} \right)^2 \frac{d\theta_1}{d\theta} + \frac{\partial}{\partial \theta_2} \left( \frac{\sin \theta_1 - \sin \theta_2}{\cos \theta_1 + \cos \theta_2} \right)^2 \frac{d\theta_2}{d\theta} = 0.$$

Since  $\theta$  is always small if the target altitude  $h$  in Figure 2 is much less than the earth's radius  $a$ , the following geometric relations apply.

$$\begin{aligned} r_1^2 &= a^2 + (a + h)^2 - 2a(a + h)\cos\theta \\ &\approx h^2 + (1 + h/a)a^2\theta^2 = a^2 \{H^2 + (1 + H)\theta^2\}. \end{aligned}$$

Note that  $h/a$  has been defined as  $H$ . Also, since angle  $\text{PTO} = \theta_1 - \theta$  is small,

$$\sin(\theta_1 - \theta) \approx a\theta/r_1 = \theta / \sqrt{H^2 + (1 + H)\theta^2} \quad (36)$$

and

$$\theta_2 \approx V - \theta.$$

Equation (36) defines  $\theta_1$  and  $\theta_2$  in terms of  $\theta$ . It can be shown that  $d\theta_2/d\theta = -1$  and

$$\frac{d\theta_1}{d\theta} = 1 + \frac{1}{\sqrt{H^2 + H\theta^2}} \left[ \frac{H^2}{H^2 + (1 + H)\theta^2} \right].$$

Using these relations, Equation (35) leads to

$$\begin{aligned} \frac{\partial \tan^2 \beta}{\partial \beta} &= 2 \left( \frac{\sin \theta_1 - \sin \theta_2}{\cos \theta_1 + \cos \theta_2} \right) \left\{ \left[ \frac{(\sin \theta_1 - \sin \theta_2) \sin \theta_1 + \cos \theta_1 (\cos \theta_1 + \cos \theta_2)}{(\cos \theta_1 + \cos \theta_2)^2} \right] \right. \\ &\quad \left. \frac{d\theta_1}{d\theta} - \left[ \frac{(\sin \theta_1 - \sin \theta_2) \sin \theta_2 - \cos \theta_2 (\cos \theta_1 + \cos \theta_2)}{(\cos \theta_1 + \cos \theta_2)^2} \right] \right\} \\ &= 2 \left( \frac{\sin \theta_1 - \sin \theta_2}{\cos \theta_1 + \cos \theta_2} \right) \left\{ \left( \frac{\sin \theta_1 - \sin \theta_2}{\cos \theta_1 + \cos \theta_2} \right)^2 + 1 \right. \\ &\quad \left. + \left[ \frac{1 + \cos(\theta_1 + \theta_2)}{(\cos \theta_1 + \cos \theta_2)^2} \right] \left[ \frac{H^2}{\sqrt{H^2 + H\theta^2} [H^2 + (1 + H)\theta^2]} \right] \right\} = 0 \end{aligned} \quad (37)$$

The quantity in the braces in Equation (37) is always positive, so that the solution to  $\partial \tan^2 \beta / \partial \theta = 0$  must occur when  $\theta_1 = \theta_2$ .

The physical implication of these results is that the major contribution to the average scattered power comes from the neighborhood of the specular point  $\phi = 0$ ,  $\theta_1 = \theta_2$ , although the size of the effective scattering region is not determined by the first Fresnel zone but rather by the rough surface parameter  $\tan \beta_0 = 2\sigma/T$ . Mathematically speaking, this implies that  $P_o$  is highly peaked at the specular point. It is, therefore, sufficient to represent  $P_o$  fairly accurately only in the neighborhood of the stationary point. In the remainder of this Appendix, the subscript "s" means that the function is evaluated at the stationary point. Following the procedure required for the method of steepest descent<sup>2</sup>,

$$\begin{aligned} \tan^2 \beta &\approx \frac{1}{2} \left( \frac{\partial^2 \tan^2 \beta}{\partial \phi^2} \right)_s \phi^2 + \frac{1}{2} \left( \frac{\partial^2 \tan^2 \beta}{\partial \theta^2} \right)_s (\theta - \theta_s)^2 \\ &\approx \frac{1}{4} \left( \tan^2 B \right) \phi^2 + M^2 (\theta - \theta_s)^2 \end{aligned} \quad (38)$$

which is the result cited in Equation (7).

Finally, it is necessary to verify Equation (8) for  $K$ . Evaluating the defining Equations (3) and (4) near the stationary point, where  $\cos \beta_0 \approx 1$ , it is found that (assuming that  $r \approx r_2$ )

$$\begin{aligned}
 K &= \langle P_s \rangle / P_d \\
 &\approx (\alpha^2 \sin \theta_s / P_d) \iint P_o \, d\theta d\phi \\
 &\approx \left[ \frac{D^2 |R|^2}{4\pi (r_1^2)} \right]_s \alpha^2 \sin \theta_s \cot^2 \beta_o I
 \end{aligned} \tag{39}$$

where

$$I = \iint \exp(-\tan^2 \beta / \tan^2 \beta_o) \, d\theta d\phi.$$

Note that  $D$  and  $R$  are understood to denote the values of these parameters near the stationary point. The integration is carried out by using Equation (38) and allowing the limits to go to  $\pm \infty$  in accordance with the method of steepest descent:

$$\begin{aligned}
 I &\equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[ -\frac{(\phi^2/4) \tan^2 B + M^2 (\theta - \theta_s)^2}{\tan^2 \beta_o} \right] d\theta d\phi \\
 &= \frac{\sqrt{\pi} 2 \tan \beta_o \sqrt{\pi} \tan \beta_o}{\tan B \quad M}
 \end{aligned} \tag{40}$$

where repeated use has been made of the formula

$$\int_{-\infty}^{\infty} \exp(-p^2 x^2) \, dx = \frac{\sqrt{\pi}}{p}.$$

Substituting Equation (40) into Equation (39), using the values of  $\alpha^2 (r_1^2)_s$  and  $\sin \theta_s$  found from Equation (36) and noting that  $\theta_1 - \theta_s \approx \theta_2 - \theta_s = 2\theta_x$  (which is defined as the angle  $A$ ), the result is found to lead to Equation (8).

### Evaluation of $C(\tau)$

It was shown that the autocorrelation function of the scattered signal,  $C(\tau)$ , is given by Equation (18), with  $P_o$  and  $(R_o - R_s)$  defined by Equations (19) and (20). It will now be shown that this leads to Equations (21) and (22) for  $C(\tau)$ .

It has already been shown that  $P_n$  is highly peaked near the stationary point  $\phi = 0$  and  $\theta = \theta_x$ . With the approximation  $\cos\phi \approx 1$  and approximating  $\sin(\theta_1 - \theta)$  by a Taylor series near  $\theta = \theta_x$ , Equation (20a) acquires the form

$$R_n - R_n \approx v\tau \{p_1 + p_2 (\theta - \theta_x)\}, \quad \text{Case I} \quad (41)$$

where  $p_1 = \theta_x \sqrt{H^2 + (1+H)\theta_x^2}$  and  $p_2 = H^2 / [H^2 + (1+H)\theta_x^2]^{3/2}$ . For Equation (20b), using  $\sin\phi \approx \phi$  and setting  $\sin(\theta_1 - \theta) \approx \sin(\theta_1 - \theta_x) \equiv \sin A$ , one gets

$$R_n - R_n \approx v\tau (\sin A) \phi, \quad \text{Case II.} \quad (42)$$

Substitution of these values and Equation (19) into Equation (18) gives, for Case I,

$$C(\tau) \sim \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[-M_1^2 \phi^2 - M_2^2 (\theta - \theta_x)^2] \exp[-jk v\tau \{p_1 + p_2 (\theta - \theta_x)\}] d\theta d\phi, \quad (43)$$

where  $M_1$  and  $M_2$  are constants introduced to simplify Equation (19); similarly for Case II,

$$C(\tau) \sim \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[-M_1^2 \phi^2 - M_2^2 (\theta - \theta_x)^2] \exp[-jk v\tau (\sin A) \phi] d\theta d\phi. \quad (44)$$

The integral with respect to  $\theta$  is evaluated by noting that, with  $a$ ,  $b$ , and  $c$  as constants,

$$\int_{-\infty}^{\infty} \exp[-ja x] \exp[-b^2(x-c)^2] dx = \frac{\sqrt{\pi}}{b} \exp\left(-\frac{a^2}{4b^2} - jac\right). \quad (45)$$

The same formula with  $c = 0$  allows calculation of the integral relative to  $\phi$ . Since only the magnitude is of interest, the term  $\exp(-jac)$  can be omitted in each case. Again, the scale factor  $\sqrt{\pi}/b$  can be omitted because it cancels out in  $|C(\tau)/C(0)|$ , which is the quantity

of ultimate interest. After some manipulation, it is found that

$$|C(\tau)| \sim \exp[-\{kv\tau p_2/2M_2\}^2], \quad \text{Case I} \quad (46)$$

$$\sim \exp[-\{kv\tau \sin A/2M_1\}^2]. \quad \text{Case II} \quad (47)$$

These lead to Equations (21) and (22) on substituting the values of  $p_2$ ,  $M_1$ , and  $M_2$ .

### Evaluation of $C(\Omega)$

$C(\Omega)$  is the correlation between received signal envelopes for a transmitted frequency separation of  $\Omega$ , and it is defined by Equation (29) of the text. Some of the steps leading to Equations (31) and (32) are discussed here.

The path length  $L_i$  for the  $i$ th path is defined in Equation (30b) as  $L_i = TP + PS$  where  $P$  is a point in the neighborhood of the stationary point  $P_1$ . To demonstrate that Equation (31) holds, construct a plane tangential to the mean earth surface at  $P_1$  in Figure 2. Recall that the spherical coordinates of  $P_1$  are  $(a, \theta_s, 0)$ , with angle  $\theta$  measured from the polar axis  $OT$  and the angle  $\phi$  measured from the plane  $TOP_1$ . Let the tangential plane at  $P_1$  intersect the axis  $OT$  in the point  $P_2$ . Set up a rectangular coordinate system  $(x', y')$  in the tangential plane with  $P_1$  as the origin,  $P_1P_2$  as the  $y'$  axis, and  $P_1P_3$  (in the direction of positive  $\phi$ ) as the  $x'$  axis. Then it can be shown that a point  $Q(x', y')$  close to  $P_1$  in the tangential plane is described by

$$x' \approx a \sin \theta_s \tan \phi \approx a \phi \sin \theta_s \quad (48)$$

$$y' \approx a(\cos \theta - \cos \theta_s)/\sin \theta_s \approx -a(\theta - \theta_s) \quad (49)$$

because  $\phi \approx 0$ ,  $\theta \approx \theta_s$ .

Now, for the point  $Q$ , Equation (30b) gives

$$L_q = TQ + QS \equiv TP_1 + P_1S + g, \quad (50)$$

where

$$g \equiv (TQ - TP_1) + (QS - P_1S). \quad (51)$$

The differential distances  $TQ - TP_1$  and  $QS - P_1S$  can be expressed in terms of  $x'$ ,  $y'$  and translated to  $\theta$ ,  $\phi$  by using Equations (48) and (49). After some simplification this leads to Equation (31).

To obtain Equation (32), it is necessary to substitute for  $P_o$  and  $g$  in Equation (29). Ignoring the constant multipliers, this gives

$$C(\Omega) \sim \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp [ - M_1^2 \phi^2 - M_2^2 (\theta - \theta_s)^2 ] \exp \left[ - j \frac{\Omega}{c} \{ N_o + N_1^2 \phi^2 + N_2^2 (\theta - \theta_s)^2 \} \right] d\theta d\phi \quad (52)$$

with  $N_o \equiv (TP_1 + P_1S)$ ,  $N_1^2 \equiv a^2 \sin^2 \theta_s / 2TP_1$ ,  $N_2^2 \equiv a^2 / 2TP_1$ . The integral is evaluated by collecting the terms in  $\phi^2$  and those in  $(\theta - \theta_s)^2$ .

The phase term  $\exp(-j\Omega N_o/c)$  is ignored because only the magnitude of  $C(\Omega)$  will be needed ultimately; moreover; the scale factor  $\pi$  may be ignored because it cancels out in  $|C(\Omega)/C(0)|$ . These simplifications lead to Equation (32).

# PROBABILITY DISTRIBUTION OF TIME TO PHASE LOCK FOR A SECOND-ORDER PHASE-LOCKED LOOP

By

F. S. KEBLAWI

RCA Astro-Electronics Division  
Hightstown, N. J.

*Summary*—This paper presents the results of an investigation of the time required for a second-order phase-locked loop to achieve phase lock following a step-function perturbation. Two sets of loop equations are presented, one for the stable region and the other for the unstable region. The probability density distribution of the time to lock is determined. In addition, the relationship between the noise-loop bandwidth and the mean time to lock is derived.

## INTRODUCTION

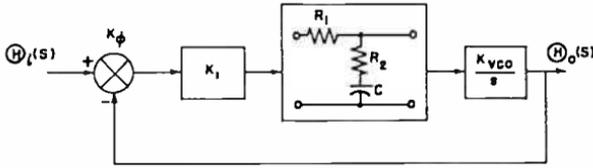
THE total time required for a phase-locked loop to achieve full lock can be treated as a sum of the time required for frequency lock and the time required for phase lock. The time required for a second-order phase-locked loop to achieve frequency lock has been studied<sup>1</sup> for the condition of a frequency offset between the loop voltage-controlled oscillator (VCO) and the signal frequency. In that study, the time required for phase lock was neglected on the assumption that it is much smaller than the time required for frequency lock. However, this assumption is not true for high ratios of loop bandwidth to frequency offset.<sup>1</sup>

It is sometimes necessary to determine the time it takes the loop to lock if the signal and the VCO are offset in phase but not in frequency. Such a need arose in the evaluation of bit synchronization time in planetary relay-link communication systems,<sup>2</sup> where the synchronization time may be a significant portion of the total (only tenths of seconds for certain mission profiles) transmission time.

The second-order phase-locked loop configuration shown in Figure 1 was studied; it has a phase detector with the triangular characteristics

<sup>1</sup> Richman, D. "Color-Carrier Reference Phase Synchronization Accuracy in NTSC Color Television," *Proc. IRE*, Jan. 1954.

<sup>2</sup> *Selected Studies of UHF/VHF Communications for Planetary (Mars/Venus) Relay Links*, Final Report, Contract No. NAS2-3772.



$$F(s) = \frac{T_2 s + 1}{T_1 s + 1}$$

$$T_1 = (R_1 + R_2)C$$

$$T_2 = R_2 C$$

$$\omega_n^2 = K/T_1$$

$$\tau = T_2 \omega_n / 2$$

$$K = K_1 K_\phi K_{VCO} = \text{static gain of this loop}$$

Fig. 1—Second-order phase-locked loop diagram and related parameters.

shown in Figure 2. This investigation is considered applicable for a loop with a sine-wave detector characteristic if it is analyzed on a quasi-linear basis.

The loop is assumed to be in lock; a step function of input phase,  $\theta_i$ , is then applied to perturb the loop. The time to lock is defined in this paper as the time required by the loop to have the phase error diminish to  $\pm 10^\circ$ . Because of this criterion, the time to lock is a function of the magnitude of the step input (as shown later). As a result of the phase-detector characteristic, the loop operates in two regions: (1) the stable region, where the slope of the phase-detector characteristic is positive and the midpoint is unconditionally stable, and (2)

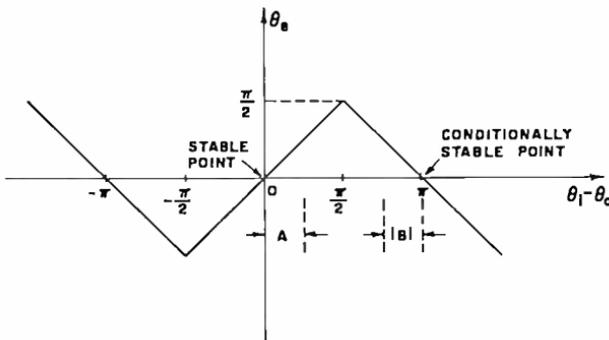


Fig. 2--Phase-detector characteristics.

the unstable region, where the slope of the phase detector characteristic is negative and the midpoint is conditionally stable.

Loop equations are presented for each of the two regions. The transient behavior of the loop in each region is obtained from these equations for a damping factor of 0.707. The time required to achieve lock, as defined previously, is found from this transient response to a step function of input phase. Thus, a functional relationship between  $\theta_i$  and the normalized time to lock ( $\omega_n t_L$ ) is obtained in the form of a graph. Based on this relationship, and based on a uniform probability density distribution of input phase,  $\theta_i$ , for  $-\pi < \theta_i \leq \pi$ , the probability density distribution,  $f_{\omega_n t_L}(\omega_n t_L)$  of the time to lock is obtained. The normalized mean time to lock ( $\langle \omega_n t_L \rangle$ ) for the  $\pm 10^\circ$  criterion is found to be approximately 2.5 radians.

### THE STABLE REGION

The transfer function of the second-order phase-locked loop shown in Figure 1 has been derived and widely discussed in the literature<sup>3-5</sup> for the linear region of a phase detector with a sine-wave characteristic. This corresponds to the region where the phase error can be approximated by  $\sin \theta_r \approx \theta_r$ . The loop transfer function for these conditions is the same as the transfer function for a loop with a triangular phase detector characteristic in the positive-slope region (see Figure 2). In Laplace notation, and assuming the system to be at rest at  $t = 0$ , the transfer function in the positive-slope region is

$$H(s) = \frac{\Theta_o(s)}{\Theta_i(s)} \approx \frac{2\zeta \omega_n s + \omega_n^2}{s^2 + 2\zeta \omega_n s + \omega_n^2} \quad (1)$$

where  $\zeta$  is a damping factor (dimensionless). The output of the phase detector is

$$\frac{\Theta_r(s)}{\Theta_i(s)} = 1 - \frac{\Theta_n(s)}{\Theta_i(s)} = \frac{s^2}{s^2 + 2\zeta \omega_n s + \omega_n^2} \quad (2)$$

<sup>3</sup> Sanneman, R. W. and Rowbotham, J.R., "Unlock Characteristics of The Optimum Type II Phase Locked Loop," *IEEE Trans. Aerospace and Navigational Electronics*, March 1964.

<sup>4</sup> Gardner, F. M., Kent, S. S. and Dasenbrock, R. D., "Theory of Phase Lock Technique," NASA N-66-10515.

<sup>5</sup> Martin, B. D., "Pioneer IV Lunar Probe: A Minimum Power FM/PM System Design," JPL TR-32-215, March 15, 1962.

Solving this equation for  $\Theta_i(s) = A/s$ , which is a step function of input phase of magnitude  $A$ , results in

$$\theta_e(t) = A \exp \{-\zeta \omega_n t\} \left[ \cosh (\omega_n \sqrt{\zeta^2 - 1}) t - \frac{\zeta}{\sqrt{\zeta^2 - 1}} \sinh (\omega_n \sqrt{\zeta^2 - 1}) t \right]. \quad (3)$$

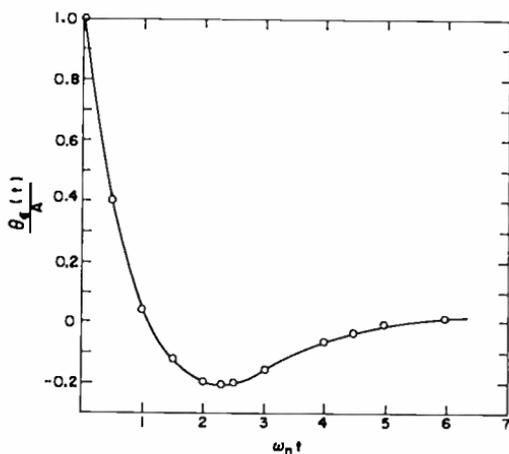


Fig. 3—Normalized response for  $\theta_e(t)$  due to step function of input phase of magnitude  $A \leq \pi/2$ .

This investigation is of a loop with a damping factor  $\zeta = 1/\sqrt{2}$ ; the substitution of this value into Equation (3) and the rearrangement of terms results in

$$\frac{\theta_e(t)}{A} = \sqrt{2} \exp \left[ -\frac{\omega_n t}{\sqrt{2}} \right] \cos \left[ \frac{\omega_n t}{\sqrt{2}} - \frac{\pi}{4} \right]. \quad (4)$$

Equation (4) describes the normalized transient response of the phase error if the loop is subjected to a step-function perturbation. Equation (4) is plotted in Figure 3.

It is reasonable at this point in the investigation to assign a criterion for phase lock subsequent to driving the loop by a step function. It is suggested that if the phase error in response to a step function  $A \leq \pi/2$  settles down to within a bound of  $\pm 10^\circ$ , the loop will be called

locked. The time to phase lock can be easily obtained from Figure 3 for any step function. For instance, if  $A$  is  $90^\circ$ , then to achieve phase lock, the value of  $\theta_c(t_L)/A$  must equal  $10/90$ , which corresponds to a normalized time-to-lock,  $\omega_n t_L$ , of 3.45 radians. The actual time to lock is obtained directly by dividing the normalized time to lock by  $\omega_n$ , the natural radian frequency of the loop. The normalized time to lock of a second-order phase-locked loop is shown in Figure 4, as a function of the step input  $A$ , where  $0 \leq A \leq \pi/2$ .

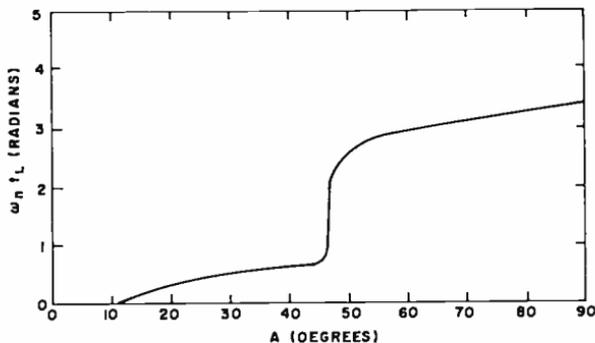


Fig. 4—Normalized time to lock as a function of initial displacement,  $A$ .

Figure 4 indicates clearly that the time to lock is a function of the step input. This is a consequence of the criterion chosen to define lock. A criterion could have been chosen that would result in a lock time independent of the step function of input phase. From Figure 3, for example, the time to lock could be chosen as the time at which the normalized error crosses the zero point ( $\omega_n t \sim 1.1$  radians). This becomes the time to lock regardless of the magnitude of the step function  $A$ , so long as  $A \leq \pi/2$ .

This criterion was not adopted here, however, because at normalized times greater than 1.1 radians, the normalized error can be as much as 0.21 ( $\omega_n t = 2.6$  radians). This corresponds to  $19^\circ$  if  $A = \pi/2$ . The justification for preferring the  $10^\circ$  criterion to the  $19^\circ$  criterion lies in the relationship between signal-to-noise ratio in the loop bandwidth and the rms phase jitter in the loop.

An rms phase jitter of  $10^\circ$  corresponds to an SNR in the loop bandwidth of about 12 dB, while a jitter of about  $19^\circ$  corresponds to 7.5 dB (see References (4) and (5)). Under noisy conditions, a phase error of  $10^\circ$  corresponds to a high signal-to-noise ratio and, therefore, to little degradation in loop performance. Hence, this phase error is chosen as a criterion of lock.

## THE UNSTABLE REGION

The value of  $\theta_c(t)$  in the unstable region can be expressed (see Figure 2) as

$$\theta_c(t) = - [\theta_i(t) - \theta_o(t)] + \pi. \quad (5)$$

where  $\theta_o$  is the phase of the output signal. To simplify the calculations, the  $(\theta_i - \theta_o)$  axis is shifted so that the conditionally stable point is at the origin. Thus, Equation (5) becomes

$$\theta_c(t) = - [\theta_i(t) - \theta_o(t)], \quad (6)$$

or, in Laplace operator notation,

$$\Theta_c(s) = - [\Theta_i(s) - \Theta_o(s)]. \quad (7)$$

The negative sign in Equation (7) can be absorbed in the open-loop gain. Thus, in the relation

$$\Theta_o(s) = \Theta_c(s) \frac{KF(s)}{s}, \quad (8)$$

the constant  $K$  can be replaced by  $-K$ , and the derivation of  $H(s)$  is then the same as in References (4) and (5). The result is

$$H(s) = \frac{-(2\zeta\omega_n s + \omega_n^2)}{s^2 - 2\zeta\omega_n s - \omega_n^2}. \quad (9)$$

Furthermore, in the unstable region,

$$\frac{\Theta_c(s)}{\Theta_i(s)} = \frac{\Theta_o(s) - \Theta_i(s)}{\Theta_i(s)} = H(s) - 1. \quad (10)$$

Thus,

$$\frac{\Theta_c(s)}{\Theta_i(s)} = \frac{-s^2}{s^2 - 2\zeta\omega_n s - \omega_n^2}. \quad (11)$$

Equation (11) is applicable only in the unstable region. Assuming that

$$\theta_i = B\mu(t) \quad (12)$$

where  $\mu(t)$  is a unit step at  $t = 0$ , then Equation (11) yields the following for a unit step of input phase of magnitude  $B$ ,

$$\Theta_c(s) = \frac{-Bs}{s^2 - 2\zeta\omega_n s - \omega_n^2}. \quad (13)$$

Solving Equation (13) for the phase error yields

$$\theta_c(t) = -B \exp\{\omega_n \zeta t\} \left[ \frac{\zeta}{\sqrt{\zeta^2 + 1}} \sinh(\sqrt{1 + \zeta^2} \omega_n t) + \cosh(\sqrt{\zeta^2 + 1} \omega_n t) \right]. \quad (14)$$

Again, letting  $\zeta = 1/\sqrt{2}$ , Equation (14) becomes

$$\frac{\theta_c(t)}{B} = -\exp\left\{\frac{\omega_n t}{\sqrt{2}}\right\} \left[ \frac{1}{\sqrt{3}} \sinh \sqrt{\frac{3}{2}} \omega_n t + \cosh \sqrt{\frac{3}{2}} \omega_n t \right]. \quad (15)$$

As in the case for the stable region, the time required to lock is defined as the time it takes for the loop error to settle down to within  $\pm 10^\circ$  after the loop has been subjected to a step-function perturbation.

Thus, to find the time it takes the loop to lock (i.e., the time it takes for the phase error to be bounded by  $\pm 10^\circ$ ), it is necessary first to find the time it takes for the phase error to reach  $\pi/2$ , using Equation (15), and then to add the time it takes the loop phase error to diminish from  $90^\circ$  to  $\pm 10^\circ$ . This calculation requires knowledge of the initial conditions of the system (when  $\theta_c = \pi/2$ ) and the use of an appropriate equation for the stable region. The required equation is derived later for the proper initial conditions.

The normalized time to achieve  $\theta_c = \pi/2$  could be obtained from Equation (15), i.e., by plotting a graph of  $\omega_n t_{90}$  versus  $B$  as shown in Figure 5. The A scale shown in Figure 5 shifts the reference point to the unconditionally stable point. Thus, the case where the rest position of the loop is the conditionally stable point and a step function of  $(+B)$  is applied is considered to be the same as when the loop is at rest at the unconditionally stable point and a step function of  $180^\circ + B$  is applied.

The discussion will be confined to negative values of  $B$ ; positive values of  $B$  yield identical results.

The solution for the loop behavior after the phase error increases to  $\pi/2$  requires that the system equation be presented in the time domain. This can be obtained from Equation (2) by inspection, and is

$$\frac{d^2\theta_e(t)}{dt^2} + 2\zeta\omega_n \frac{d\theta_e(t)}{dt} + \omega_n^2\theta_e(t) = \frac{d^2\theta_i(t)}{dt^2}. \quad (16)$$

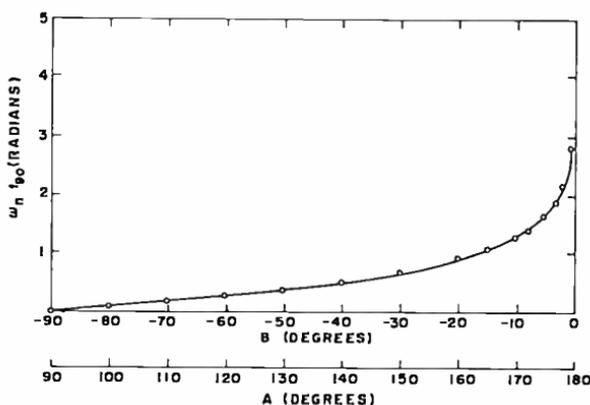


Fig. 5—Normalized time to reach phase error of  $\pi/2$  as a function of step inputs with magnitude of  $A$  such that  $\pi/2 \leq A \leq \pi$ , or of  $B$  such that  $-\pi/2 \leq B \leq 0$ .

Substituting  $\frac{d^2\theta_i(t)}{dt^2} = 0$  in Equation (16) and using Laplace trans-

formations results in

$$s^2 \Theta_e(s) - s\theta_e(0) - \dot{\theta}_e(0) + 2\zeta\omega_n [s \Theta_e(s) - \theta_e(0)] + \omega_n^2 \Theta_e(s) = 0. \quad (17)$$

Thus,

$$\Theta_e(s) = \frac{\theta_e(0) [s + 2\zeta\omega_n] + \dot{\theta}_e(0)}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \quad (18)$$

The solution to Equation (18) is

$$\theta_r(t) = \theta_r(0) \exp \{-\omega_n \zeta t\} \left[ \cosh \sqrt{\zeta^2 - 1} \omega_n t - \frac{\zeta}{\sqrt{\zeta^2 - 1}} \sinh \sqrt{\zeta^2 - 1} \omega_n t \right] \quad (19)$$

$$+ \left( \frac{2\zeta\omega_n\theta_r(0) + \dot{\theta}_r(0)}{\omega_n\sqrt{\zeta^2 - 1}} \right) \exp \{-\zeta\omega_n t\} \sinh (\sqrt{\zeta^2 - 1} \omega_n t).$$

For  $\zeta = 1/\sqrt{2}$ , Equation (19) becomes

$$\theta_r(t) = \exp \left\{ -\frac{\omega_n t}{\sqrt{2}} \right\} \theta_r(0) \left( \cos \frac{\omega_n t}{\sqrt{2}} + \sin \frac{\omega_n t}{\sqrt{2}} \right)$$

$$+ \frac{\sqrt{2}}{\omega_n} \dot{\theta}_r(0) \exp \left\{ -\frac{\omega_n t}{\sqrt{2}} \right\} \sin \frac{\omega_n t}{\sqrt{2}}, \quad (20)$$

where

$$\theta_r(0) = \frac{\pi}{2}, \quad (21)$$

and  $\dot{\theta}_r(0)$  must be obtained from the behavior of the loop in the unstable region. Since  $\dot{\theta}_r(0)$  is the negative of  $\dot{\theta}_r(\omega_n t_{90})$  in the unstable region,

$$[\theta_r(0)]_{\text{stable region}} = [-\dot{\theta}_r(\omega_n t_{90})]_{\text{unstable region}} \quad (22)$$

The change in the sign of the derivative results from the characteristics of the phase detector.

Equation (15) for the unstable region can be rewritten

$$\frac{\theta_r(t)}{B} = -\frac{1}{2} \exp \left\{ \frac{\omega_n t}{\sqrt{2}} \right\} \left[ \frac{1}{\sqrt{3}} \left( \exp \left\{ \frac{\sqrt{3}}{2} \omega_n t \right\} - \exp \left\{ -\frac{\sqrt{3}}{2} \omega_n t \right\} \right) \right. \quad (23)$$

$$\left. + \exp \left\{ \sqrt{\frac{3}{2}} \omega_n t \right\} + \exp \left\{ -\frac{\sqrt{3}}{2} \omega_n t \right\} \right]$$

The derivative of Equation (23) is

$$\frac{\dot{\theta}_r(t)}{B} = -\frac{\omega_n t}{2\sqrt{2}} \exp\left\{\frac{\omega_n t}{\sqrt{2}}\right\} \left[ \left(\frac{4 + 2\sqrt{3}}{\sqrt{3}}\right) \exp\left\{\frac{3}{2}\omega_n t\right\} - \left(\frac{4 - 2\sqrt{3}}{\sqrt{3}}\right) \exp\left\{-\sqrt{\frac{3}{2}}\omega_n t\right\} \right] \quad (24)$$

Table I—Derivative of Phase Error as a Function of Step Input ( $B$ )

$B$	$\dot{\theta}_r(0)$
$-90^\circ$	$-126.0$
$-80^\circ$	$-136.5$
$-70^\circ$	$-145.0$
$-50^\circ$	$-152.5$
$-30^\circ$	$-157.5$
$-20^\circ$	$-164.5$
$-10^\circ$	$-170.0$
$-6^\circ$	$-179.0$
$-1^\circ$	$-421.5$

The term  $\dot{\theta}_r(\omega_n t_{90})$  is of interest because it is the rate of change of  $\theta_r$  when  $\theta_r$  is  $\pi/2$  for all values of  $B$ . The value of  $\omega_n t_{90}$  can be obtained (from Figure 5) for several step functions of magnitude  $B$  and substituted in Equation (24). Values of  $\dot{\theta}_r(0)$ , the negative of the result, are presented in Table I as a function of  $B$ . Substituting  $\pi/2$  for  $\theta_c(0)$  in Equation (20) results in

$$\theta_r(t) = x(t) + y(t), \quad (25)$$

where

$$x(t) = \frac{\pi}{2} \exp\left\{-\frac{\omega_n}{\sqrt{2}} t\right\} \left(\cos \frac{\omega_n}{\sqrt{2}} t + \sin \frac{\omega_n}{\sqrt{2}} t\right) \quad (26)$$

and

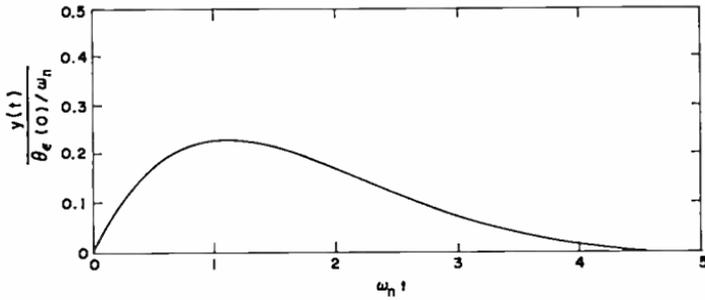


Fig. 6—The value of  $y(t)/[\dot{\theta}_c(0)/\omega_n]$  as a function of  $\omega_n t$ .

$$y(t) = \frac{\sqrt{2}}{\omega_n} \dot{\theta}_c(0) \exp\left\{-\frac{\omega_n}{\sqrt{2}} t\right\} \sin \frac{\omega_n}{\sqrt{2}} t. \quad (27)$$

The value of  $y(t)/[\dot{\theta}_c(0)/\omega_n]$  is plotted in Figure 6. For the special conditions of  $B = -90^\circ$  and  $-6.5^\circ$ , the value of  $y(t)$  is shown in Figure 7. Figure 8 shows the summation values of  $x(t)$  and  $y(t)$  for  $B = -90^\circ$  and  $-6.5^\circ$ . Figure 8 shows the transient behavior of  $\theta_c$  (starting at  $\theta_c = \pi/2$ ) due to initial conditions at that point precipitated by operation in the unstable region. A family of curves could be drawn for a set of  $B$  values, but Figure 8 indicates that for disturbances of  $-90^\circ$  and  $-6.5^\circ$ , the loop phase error moves from  $90^\circ$  to within  $\pm 10^\circ$  in normalized times differing only by 0.2 radian. This is small considering the overall time the loop takes to move from the unstable region to the  $\pm 10^\circ$  bound.

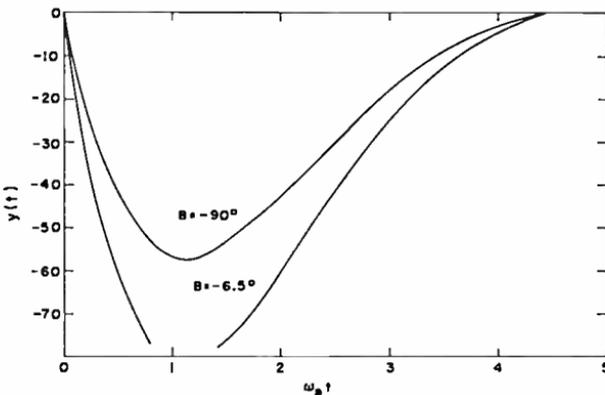


Fig. 7—The value of  $y(t)$  as a function of  $\omega_n t$  for  $B = -90^\circ$  and  $B = -6.5^\circ$ .

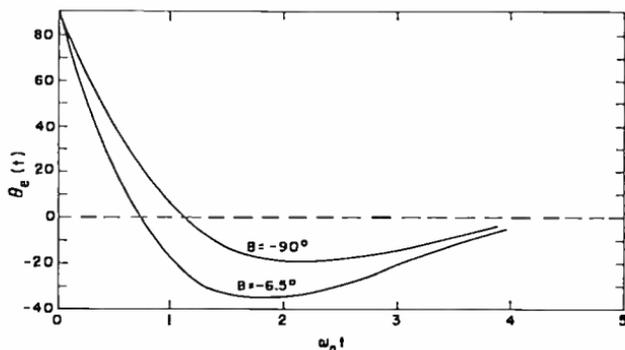


Fig. 8—The value of  $\theta_e(t)$  as a function of  $\omega_n t$  resulting from initial operation in the unstable region.

No curve for a value of  $B$  larger than  $-6.5^\circ$  is shown in Figure 8; the derived equations are not valid for  $B$  larger than  $-6.5^\circ$  because the loop would probably skip cycles. Initial conditions at  $\theta_e = -\pi/2$  can no longer be assumed to be zero, and the loop behavior will require investigation for these conditions. However, the time taken by the loop to travel from the neighborhood of the unstable point to  $\theta_e = \pi/2$  (under the influence of a small step) is so large that the time it takes the phase error to change from  $90^\circ$  to  $\pm 10^\circ$  is small in comparison. Theoretically, if the loop is at rest at the conditionally stable point, it will never move from that position, and the time to lock is infinite.

The solutions for the stable region ( $A \leq \pi/2$ ) and the unstable region ( $A \geq \pi/2$ ) are the same when  $A$  is  $\pi/2$ . The time to lock,  $\omega_n t_L$ , is equal to 3.45 radians if loop operation starts at the point of discontinuity. The time to lock for  $A \geq \pi/2$  is shown in Figure 9. This is

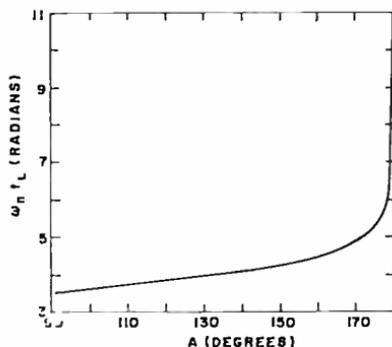


Fig. 9—Time to lock as a function of step input of magnitude  $A$  ( $A \geq \pi/2$ ).

obtained by shifting the curve of Figure 5 by 3.45 radians, as indicated by Figure 8.

The overall functional relationship between the time to lock ( $\omega_n t_L$ ) and the step input phase is shown in Figure 10 for a range of input phase from 0 to  $\pi$ . It is evenly symmetrical about the origin.

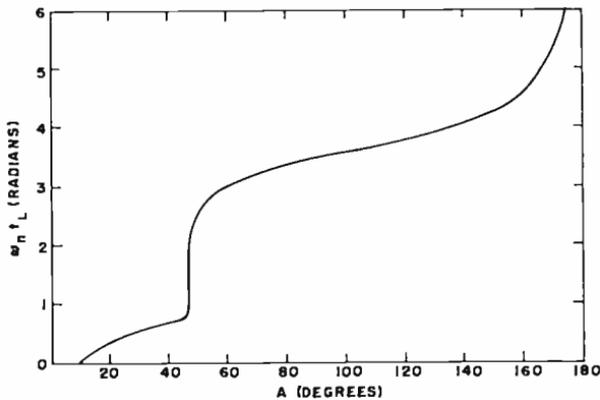


Fig. 10—Time to lock as a function of step input of magnitude  $A$  ( $0 \leq A \leq \pi$ ).

#### PROBABILITY DENSITY DISTRIBUTION OF THE TIME TO LOCK

In the preceding analysis, the loop was assumed to be initially at rest position. The step function of input phase applied to the loop is assumed to be uniformly distributed between  $-\pi$  and  $\pi$ . This is expressed as

$$f_A(A) = \frac{1}{2\pi} \quad \text{for } -\pi < A \leq \pi, \quad (28)$$

This function is the probability density distribution of  $A$ . The probability density distribution of  $\omega_n t_L$  can be obtained from  $f_A(A)$  as follows.<sup>6</sup> Let

$$\omega_n t_L = g(A). \quad (29)$$

Then the probability density distribution of the normalized time to lock is

<sup>6</sup> A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw Hill, 1965.

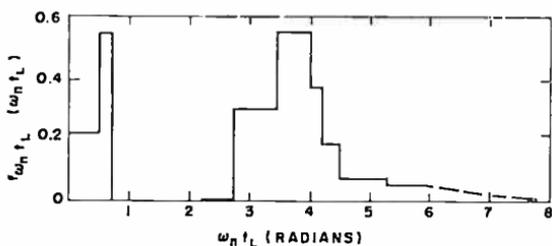


Fig. 11—Probability density distribution of time to lock.

$$f_{\omega_n t_L}(\omega_n t_L) = \frac{2f(A)}{\left| \frac{d\gamma}{dA} \right|} = \left( \frac{\Delta A}{\Delta \omega_n t_L} \right) \frac{1}{180}. \quad (30)$$

As implied by Equation (30), the probability density distribution shown in Figure 11 is obtained by dividing the time-to-lock function, shown in Figure 10, into regions where  $\omega_n t_L$  is a linear function of  $A$ . The term  $F_{\omega_n t_L}(\omega_n t_L)$  is the cumulative probability distribution of  $\omega_n t_L$ , and is shown in Figure 12; it is obtained by integration of the density function  $f_{\omega_n t_L}(\omega_n t_L)$ . As shown in Figures 11 and 12, Equation (30) is valid up to the 85% probability point; the dotted lines in Figures 11 and 12 indicate extrapolations.

The median (or 50% probability point) occurs at a value of  $\omega_n t_L = 3.6$  radians. The average value of  $\omega_n t_L$  is

$$\langle \omega_n t_L \rangle = \int_0^{\infty} \omega_n t_L f_{\omega_n t_L}(\omega_n t_L) d(\omega_n t_L). \quad (31)$$

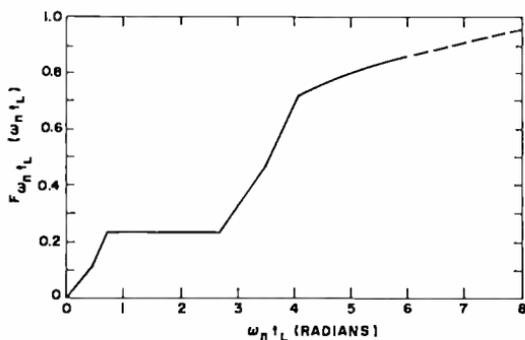


Fig. 12—Cumulative probability distribution of time to lock.

The normalized time-to-lock has been evaluated up to  $\omega_n t_L = 5.8$  radians, which corresponds to the point beyond which the equation is no longer valid. The result is

$$\langle \omega_n t_L \rangle = 2.5 \text{ radians.} \quad (32)$$

#### RELATIONSHIP BETWEEN NOISE-LOOP BANDWIDTH AND TIME TO LOCK

The relationship between the natural radian frequency of the loop and the noise-loop bandwidth can be used to further simplify Equation (32). This relationship, as derived,<sup>7</sup> is

$$2B_L \sim 1.06 \omega_n, \quad (33)$$

where  $2B_L$  is the double-sided noise-loop bandwidth in hertz. Since  $\omega_n$  is a constant, Equation (32) can be rewritten

$$\langle t_L \rangle = \frac{2.5}{\omega_n} = \frac{2.65}{2B_L}. \quad (34)$$

Thus, the time to phase lock is inversely proportional to the noise bandwidth of the loop. Equation (34) is plotted in Figure 13, as are the Richman curves.<sup>1</sup> The Richman curves show the pull-in time as a function of the single-sided noise-loop bandwidth. The pull-in time is defined by Richman as the time required for the loop to lock on a signal with a frequency ( $\Delta F$ ) away from the frequency of the free-running voltage-controlled oscillator. The time to phase lock is neglected by Richman on the assumption that it is much smaller than the time to frequency lock; Figure 13 shows, however, that the mean time required for phase lock ( $\langle t_L \rangle$ ) can be equal to or larger than the time required for frequency lock ( $t_F$ ).

The Richman equation for the pull-in time ( $t_F$ ) in terms of the single-sided noise-loop bandwidth and the initial frequency offset (between the VCO and signal frequencies) is

$$t_F = \frac{4(\Delta F)^2}{(B_L)^3}. \quad (35)$$

<sup>7</sup> Karras, T. J., "Equivalent Noise Bandwidth Analysis from Transfer Functions," NASA Technical Note TND-2842, Goddard Space Flight Center, Greenbelt, Maryland.

## CONCLUSIONS

The preceding analysis indicates that the mean time to phase lock varies inversely as the noise-loop bandwidth for a condition where no frequency offset exists between the signal frequency and the loop VCO frequency, and for a uniformly distributed input phase. The relationship is that derived in Equation (34).

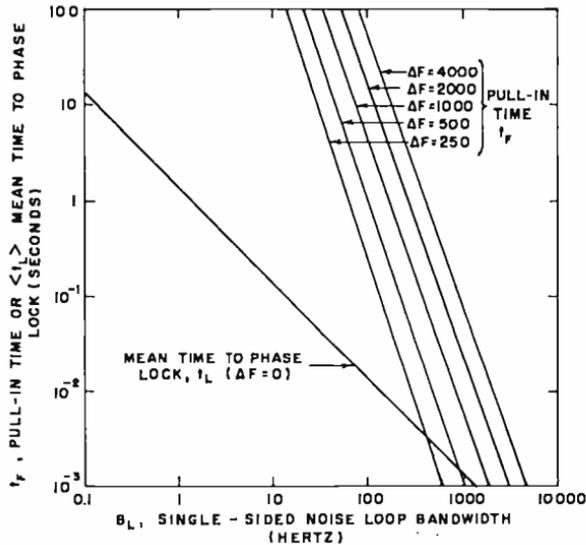


Fig. 13—Mean time to phase lock and time to frequency lock (pull-in time) as functions of single-sided noise-loop bandwidth.

It can also be concluded (using this result and the Richman results<sup>1</sup>) that the mean time to phase lock can be more significant than the pull-in time (the time to frequency lock) as determined by Richman.<sup>1</sup>

## ACKNOWLEDGMENT

The author acknowledges the support of this effort by personnel of the Ames Research Center, National Aeronautics and Space Administration, under Contract Number NAS2-3772.

# QUADRATURE DETECTION PSK SYSTEM AND INTERFERENCE

BY

V. F. VOLERTAS

RCA Defense Communications Systems Division  
Camden, N. J.

*Summary*—When perfect coherent reception of discrete phase-modulated signals is impossible, quadrature detection is used. The performance of such a binary system in white noise and similar-signal interference environments is studied. Formulas of error probabilities for various interference combinations are derived, and error rates are computed. Effects of phase and energy relations between communication and interference signals are determined. Similar-signal and white-noise interferences are compared. Results can be applied for intentional and unintentional interference as well as for an environment where reflective multipaths prevail. It is concluded that white-noise interference is surpassed by similar-signal interference in very special cases only, and, generally, the latter type of intentional interference (jamming) against this system is unlikely.

## INTRODUCTION

THE quadrature-detection signal-processing technique is described in the next section. The performance of such a spread-spectrum system, when  $n$  keying elements are integrated into one information bit, is evaluated under the following operation conditions. A transmitter sends digital, binary, orthogonal pulse-shift-keyed (PSK) signals to a receiver or group of receivers employing quadrature detection. These signals, due to operating environment, arrive at the receiver contaminated; in addition to the information-bearing wave, natural noise and interference from other radiating sources are present. The natural noise is white, gaussian, additive, of zero mean, and has a spectral density  $N_w$ . The interference from other radiating sources is of two types—similar-signal and white noise. A similar interference signal is a binary PSK wave of the same carrier frequency as the desired signal. Furthermore, it uses the same binary phases, and the duration of the keying element (the shortest time interval between the switch of the phases) equals the duration of the keying element of the information signal,  $T_w$ . The initial phase of the similar interfering signal may be random relative to the initial phase of the desired signal, or these two phases may differ by some constant angle (which may be zero). The phases of individual keying elements

of both signals may match or be opposite. Furthermore, these phase relationships may exist either for the entire duration of the keying element or for only a fraction of it. If there are several similar signals interfering with the reception, each of them may be in a different phase relationship with the desired signal. Since the number of such signals is large, the identity of individual sources may be lost and the total interference at the receiver may take the form of a white noise. Furthermore, the interfering similar signal may be of varying strength relative to the information-bearing signal.

In addition to the similar-signal and natural noise, an intentional white-noise interference is also considered. It is assumed, that this signal is statistically independent of other signals. It might be present at the receiver all times, or it might arrive interchangeably with the similar signal.

The problem, as described above, contains several possible operating conditions: (a) interference from a similar signal with little or no phase relationship with the desired signal and from two independent white-noise sources; (b) interference from a similar signal, the phase of which is strongly dependent on the phase of the desired signal (as might be caused by multipath or a jamming signal; (c) jamming from a single source with limited total jamming power divided between similar-signal and white-noise interference; (d) alternate white-noise and similar-signal interference with white noise radiated during the period when the intercepted information signal is being analyzed for its phase and other characteristics.

### QUADRATURE DETECTION

The quadrature-detection signal-processing technique assumed in this analysis is shown in Figure 1. It involves a pair of multipliers driven from quadrature phases of the carrier. The outputs of these in-phase and quadrature multipliers are filtered and then applied to matched filters (matched to the desired transmitted signal). The signals  $x(t)$  and  $\hat{x}(t)$  refer to outputs of in-phase and quadrature channels, respectively. These signals are cross-correlated with local replicas of the possible transmitted orthogonal signals. Since in the binary case, there are two possible orthogonal transmitted signals,  $S(t)$  and  $S^*(t)$ , there are four outputs. For the transmitted signal, the outputs of the matched filters are

$$y_1 = \int_0^T x(t)S(t)dt, \quad (1)$$

$$\hat{y}_1 = \int_0^T \hat{x}(t) S(t) dt . \quad (2)$$

$T = nT_n$ , and is the duration of an information bit. For the nontransmitted signal, the outputs of the matched filters are

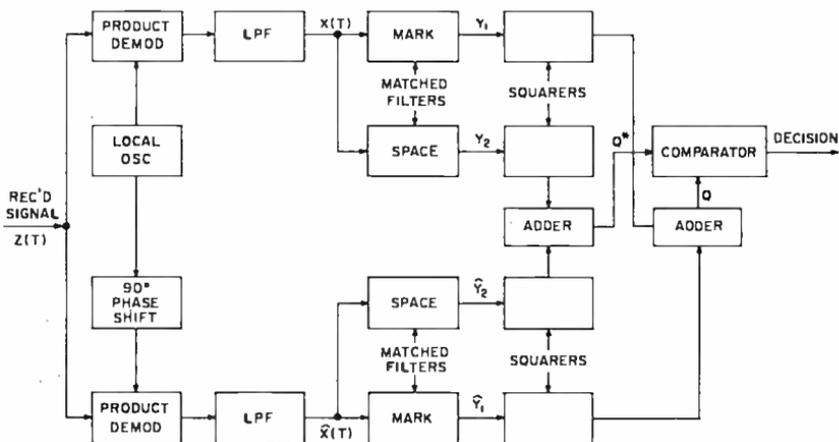


Fig. 1—Quadrature detection signal processing.

$$y_2 = \int_0^T x(t) S^*(t) dt , \quad (3)$$

$$\hat{y}_2 = \int_0^T \hat{x}(t) S^*(t) dt . \quad (4)$$

The outputs of the matched filters are squared and added. The decision procedure consists of comparing and selecting the larger of

$$Q = \sqrt{y_1^2 + \hat{y}_1^2} \quad (5)$$

and

$$Q^* = \sqrt{y_2^2 + \hat{y}_2^2} . \quad (6)$$

The probability of correct reception is

$$P_n = P(Q^* < Q) . \quad (7)$$

#### SYSTEM PERFORMANCE

The desired communications signal is a binary PSK wave of carrier frequency  $\omega_n$  and initial phase  $\theta_1$ . The modulation phase,  $\phi(t)$ , can assume only two values,  $0^\circ$  and  $180^\circ$ . The amplitude of the communication signal at the receiver is designated  $A_n$ , and the additive natural, white, gaussian noise is  $n(t)$ . The amplitude of the similar PSK interfering signal is  $A_I(t)$  and its initial phase is  $\theta_2$ . Furthermore, there exists a source transmitting white noise within the bandwidth of the receiver; the noise,  $I(t)$ , is of amplitude  $B_I(t)$  and of spectral density  $J_n$ . All signals are statistically independent and additive. Hence, the total signal at the receiver is

$$\begin{aligned} z(t) = & A_n \cos [\omega_n t + \phi(t) + \theta_1] \\ & + A_I(t) \cos [\omega_n t + \phi(t) + \theta_2] \\ & + B_I(t) I(t) + n(t) . \end{aligned} \quad (8)$$

The quadrature-detection system processes the received signal, integrates  $n$  keying elements of duration  $T_n$  into one information bit of duration  $T = nT_n$ , and makes a decision whether a mark or a space was received. The formulas for the average probability that an error occurred in decision are derived in the Appendix. Two cases are treated:

- (a) Similar signal and white noise are interfering with the desired signal continuously.
- (b) Similar signal and white noise are being radiated interchangeably, but their total power is constant.

#### CONTINUOUS SIMILAR-SIGNAL AND WHITE-NOISE INTERFERENCE

##### *Constant Relative Phase and Energy Ratio*

The use of orthogonal signals in conjunction with ideal quadrature detection in a white-noise environment leads to the same error probability as ideal noncoherent frequency shift keying.<sup>1,2</sup> When a similar PSK and a white-noise signal are interfering continuously, the average

<sup>1</sup> G. Lieberman, "Quantization in Coherent and Quadrature Reception of Orthogonal Signals," *RCA Review*, Vol. XXII, p. 461, Sept. 1961.

<sup>2</sup> S. Reiger, "Error Rates in Data Transmission," *Proc. I.R.E.*, Vol. 46, May 1958.

bit error probability of a quadrature detection system is given by Equation (31) of the Appendix:

$$P_c = \frac{1}{2} \exp \left\{ - \left[ \frac{E_s + E_I + 2 \cos \theta \sqrt{E_s E_I}}{2(N_o + I_o)} \right] \right\}. \quad (31)$$

$E_s$  is the desired signal energy per bit,  $E_I$  the PSK similar-signal interference energy per bit, and  $\theta$  is the relative phase (the difference between the initial phases of the desired and interfering similar signals). If only similar-signal interference is present,  $I_o = 0$ . If only white-noise interfering signal is present,  $E_I = 0$ . Letting  $N_o + I_o = N_o'$ , Equation (31) can be re-written

$$P_c = \frac{1}{2} \exp \left\{ - \left( \frac{E_s}{2N_o'} \right) \left[ 1 + \frac{E_I}{E_s} + 2 \cos \theta \sqrt{\frac{E_I}{E_s}} \right] \right\}. \quad (9)$$

If all variables of Equation (9) are constant,  $P_c$  can be plotted versus their magnitudes. It can be shown that the  $E_s/(2N_o')$  ratio is of least importance, and that  $E_I/E_s$  and  $\theta$  are the two quantities that, in general, determine the value of  $P_c$ . The maximum value of  $P_c$  is obtained when

$$\sqrt{\frac{E_I}{E_s}} = \begin{cases} 0 & \text{for } 0^\circ \leq \theta \leq 90^\circ \text{ and } 270^\circ \leq \theta \leq 360^\circ \\ -\cos \theta & \text{for } 90^\circ < \theta < 270^\circ \end{cases} \quad (10)$$

Thus, if  $0^\circ \leq \theta \leq 90^\circ$  or  $270^\circ \leq \theta \leq 360^\circ$ , the similar-signal interference reduces the error rate of the system. Otherwise, the error rate is increased. Figure 2 shows the variation of the bit error rate due to variations in magnitudes of  $\theta$  and  $E_s/E_I$ . Because  $E_s/(2N_o')$  is the least important variable, the bit error probability was computed for a constant value of  $E_s/(2N_o') = 6.25$ . (This ratio, when no similar-signal interference is present, results in bit error probability of  $10^{-3}$ .) In Figure 2,  $P_c$  is plotted versus  $\theta$  for  $E_s/E_I = 0.5, 1.0$ , and  $2.0$ . As can be seen, it decreases considerably below the value of no similar-signal interference in the first and fourth quadrants and in parts of the second and third quadrants, but increases very rapidly for a relative phase angle in the neighborhood of  $180^\circ$ .

### Random Relative Phase and Energy Ratio

In most practical cases, the values of  $E_I/E_s$  and  $\theta$  vary from keying element to keying element and from bit to bit, and this variation is

unpredictable. The bit error rate, as given by Equation (9), is also a random variable due to randomness of  $\theta$  and  $E_I/E_n$ . For the purposes of further analysis, it is desirable to know the probability density functions of  $\theta$  and  $E_I/E_n$ .

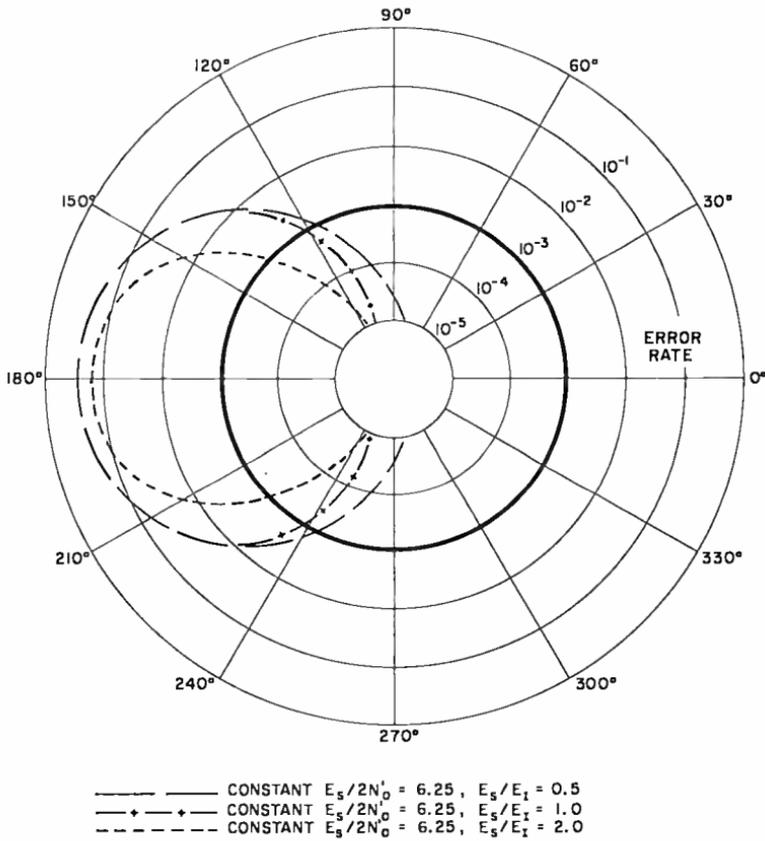


Fig. 2—Bit error probability versus  $\theta$ .

It is reasonable to assume that  $\theta$  has a uniform density distribution. The probability density function of  $\theta$  is then

$$p(\theta) = \begin{cases} 1/(2\pi), & 0^\circ \leq \theta \leq 360^\circ \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

The corresponding probability density function of  $y = \cos \theta$  is, therefore,

$$p(y) = p(\cos \theta) = \begin{cases} 1/(\pi\sqrt{1-y^2}), & -1 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

The following can be said about  $E_I/E_s$  before assuming a specific probability density distribution:

- (a) being an energy ratio, it is non-negative,
- (b) the random variation of  $E_I/E_s$  is due to the random variation in  $E_I$  and  $E_s$ .

On the basis of this argument, the following two functions can be used to describe the random behavior of  $E_I/E_s$ :

$$p\left(\frac{E_I}{E_s}\right) = p(x) = \begin{cases} (1/x_0)e^{-x/x_0}, & 0 \leq x \leq \infty \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

$$p\left(\frac{E_I}{E_s}\right) = p(x) = \begin{cases} (x/x_0^2)e^{-x/x_0}, & 0 \leq x \leq \infty \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

In both of these density functions, the mean value of  $x = E_I/E_s$  is assumed to be unity.

No conclusions are drawn as to which of these two density functions is better suited to represent the actual situation. Thus, both are employed in subsequent calculations.

Since  $x = E_I/E_s$  and  $y = \cos \theta$  are random, the bit error rate as given by Equation (9) is also a random quantity and must be treated as such. Any single value does not indicate the effectiveness of interference, and a statistical approach must be taken. A given bit error rate represents the probability of a single bit being in error at a particular time when interference represents a specific value of  $E_I/E_s$  and  $\theta$ . This probability (of a bit being in error) changes from bit to bit, from one time instant to the next. To obtain a more realistic picture, either the mean bit error rate or the probability that the error rate is equal to or greater than some fixed value  $z$  must be computed. In this analysis, the second approach is taken.

The bit error rate,  $P_e$ , is equal to or greater than some fixed value  $z$  only when  $x = E_I/E_s$  and  $y = \cos \theta$  are within specific ranges of variation. These ranges must be determined for each value of  $z$ . The probability that  $P_e = z$  is then equal to the probability that the two random variables  $x$  and  $y$  assume the corresponding values. This probability

can be obtained by integrating the joint probability density function of  $x$  and  $y$  over the area of the variation of these two variables for the specific value of  $z$ .

The probability that the bit error rate is equal to or greater than some fixed value  $z$  was computed for  $z = 0.1$  and  $z = 0.01$ . The probability density function of  $y = \cos \theta$  used in the calculations is given

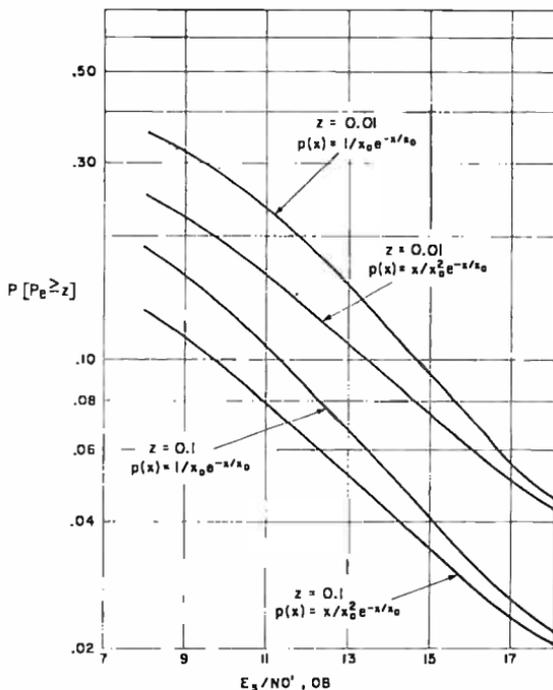


Fig. 3—Probability of bit error rate being equal to or greater than 0.1 and 0.01 versus  $E_s/N_0'$ .

by Equation (12). To describe the variation of  $x = E_1/E_s$ , both probability density functions as given by Equations (13) and (14) were employed. The results of these computations are plotted versus  $E_s/N_0'$  in Figure 3. From these curves, it can be concluded that system performance is improved when  $E_s/N_0'$  is increased. However, at  $E_s/N_0' = 18$  dB the curves start to level off, and a further increase in energy contrast is of little (if any) value. Also, it is seen that the density function of  $E_1/E_s$  as given by Equation (14) produces a smaller probability of error. As mentioned above, however, we do not know which of the two assumed density functions is more realistic.

DISCONTINUOUS SIMILAR-SIGNAL AND WHITE-NOISE INTERFERENCE  
FROM A SINGLE SOURCE WITH LIMITED TOTAL POWER

Consider a single interference source with fixed total power radiating two types of signals in the following fashion. For a fraction of the keying element duration,  $KT_n$  (here  $0 \leq K \leq 1.0$ ), the entire power is radiated in the form of the white noise; for the remaining portion of the keying element duration,  $(1 - K)T_n$ , the total available power is divided between the pseudo-random PSK signal and white noise. Such situations may arise in some multipath environments and in the case of an intentional interference source employing look-through capabilities.

For the case where the total available power is equally divided between similar signal and white noise, the bit error probability of a quadrature detection system is given by Equation (38) of the Appendix:

$$P_c = \frac{1}{2} \exp \left\{ - \left[ \frac{E_s + E_I + 2 \cos \theta \sqrt{E_s E_I}}{2|N_n + (K + 1)I_n|} \right] \right\} \quad (38)$$

where

$$E_I = \frac{A_I^2 T}{2} (K - 1)^2.$$

In this case,  $A_I(t)$  and  $B_I(t)$ , the quantities appearing in Equation (8), are

$$A_I(t) = \begin{cases} 0 & \text{for } 0 \leq t < KT_n \\ A_I & \text{for } KT_n \leq t \leq T_n \end{cases} \quad (32)$$

$$B_I(t) = \begin{cases} \sqrt{2} B_I & \text{for } 0 \leq t < KT_n \\ B_I & \text{for } KT_n \leq t \leq T_n \end{cases} \quad (33)$$

COMPARISON OF EFFECTIVENESS

*Constant Relative Phase, Variable Interference Energy*

It is of interest to see how the bit error probability is affected when the proportion of white-noise to similar-signal interference is varied. For this purpose, the bit error probability, as given by Equation (38) of the Appendix, is computed for several situations.

Figure 4 shows the bit error probability versus  $E_I/(K + 1)I_n$  where

$E_s/N_o = 12.5$  and  $(K + 1)I_o$  is a constant equal to  $N_o$ . Three values of  $\theta$  are used. For the ratio  $E_I/(K + 1)I_o$  to vary,  $E_I$  must vary (since  $(K + 1)I_o$  is constant). Therefore, the total interference energy varies. Curves are given showing the error rate for  $\theta = 120^\circ$ ,  $\theta = 150^\circ$ , and  $\theta = 180^\circ$ . A curve is also given that shows the error rate for white, gaussian noise interference only. This curve represents the case when

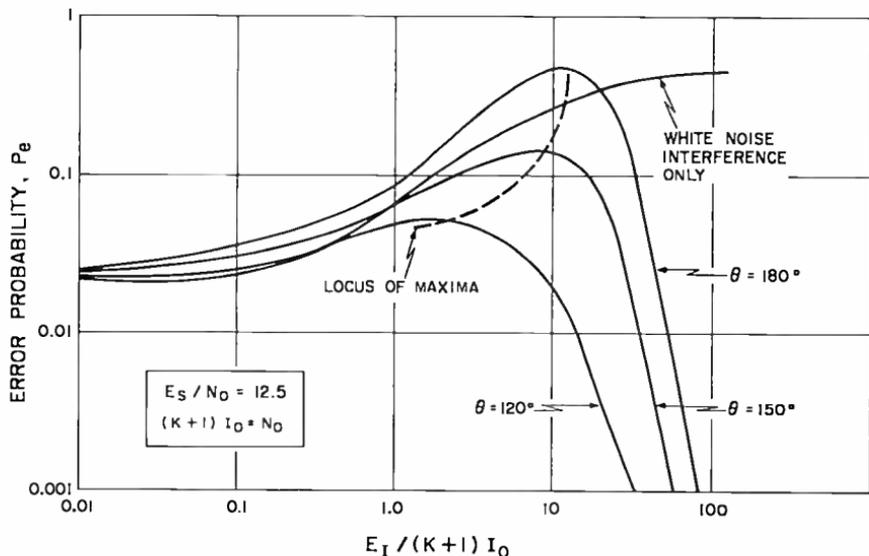


Fig. 4—Error probability versus  $E_I/[(K + 1)I_o]$  for constant  $(K + 1)I_o = N_o$  and three values of  $\theta$  compared with equal energy white noise interference.

no similar signal is interfering, i.e., the total available interference energy is converted to white noise. The spectral density of this white noise is

$$\left[ 1 + \left( \frac{E_I}{(K + 1)I_o} \right) \right] N_o$$

It is seen from the figure that for each  $\theta$  there is a value of  $E_I/(K + 1)I_o$  for which  $P_e$  is maximum. This optimum (from a jammer's point of view) value is higher for larger phase differences. The dashed curve gives the locus of the optimum value of  $E_I/(K + 1)I_o$ . When this quantity is increased beyond its optimum value, the error rate goes down very rapidly.

Figure 5 shows the error probabilities versus  $E_I/(K + 1)I_o$  for

$E_s/N_o = 25.0$  (again,  $(K+1)I_o = \text{constant} = N_o$ , and variation in  $E_I/(K+1)I_o$  means the variation in total interference energy). The general picture is identical to that of Figure 4, except that the optimum values of  $E_I/(K+1)I_o$  are higher because a higher  $E_s/N_o$  was used.

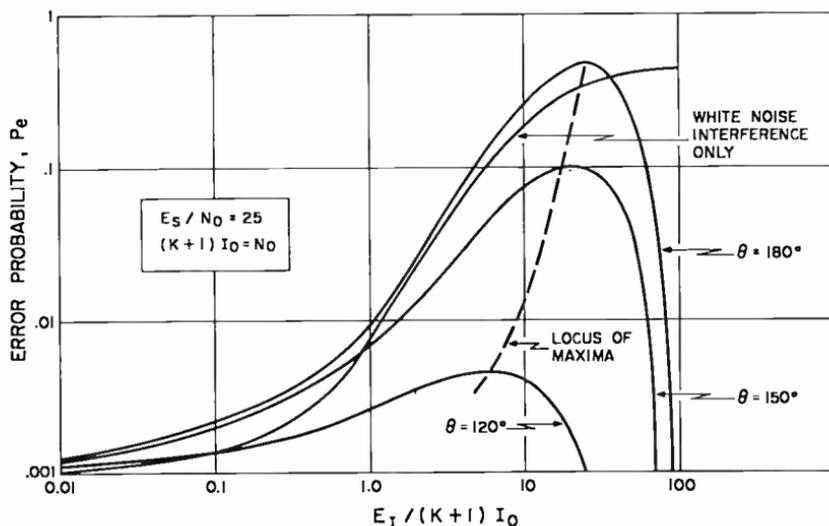


Fig. 5—Error probability versus  $E_I/[(K+1)I_o]$  for constant  $(K+1)I_o = N_o$  and three values of  $\theta$  compared with equal energy white noise interference.

Comparing this combined interference with the white-noise interference of equal energy, one sees that white noise produces a smaller error probability for small values of  $E_I/(K+1)I_o$ , regardless of the relative phase angle. When  $E_I/(K+1)I_o$  is high, however,  $P_e$  for white noise is always larger than for the combined interference. The cross-over point is a function of relative phase angle.

### Constant Relative Phase, Constant Interference Energy

We will now assume that the total interference energy is kept constant, and only the ratio  $E_I/(K+1)I_o$  is varied. The effect on the error probability is shown in Figures 6 and 7. A constant  $E_s/N_o = 25.0$  is assumed. In Figure 6, total interference energy is selected to be equal to the signal energy. The error rate,  $P_e$ , is computed for  $\theta = 120^\circ$ ,  $150^\circ$ , and  $180^\circ$ . For  $\theta = 180^\circ$ , maximum interference occurs when  $E_I/(K+1)I_o = \infty$ , i.e., when all of the energy is in the form of similar-

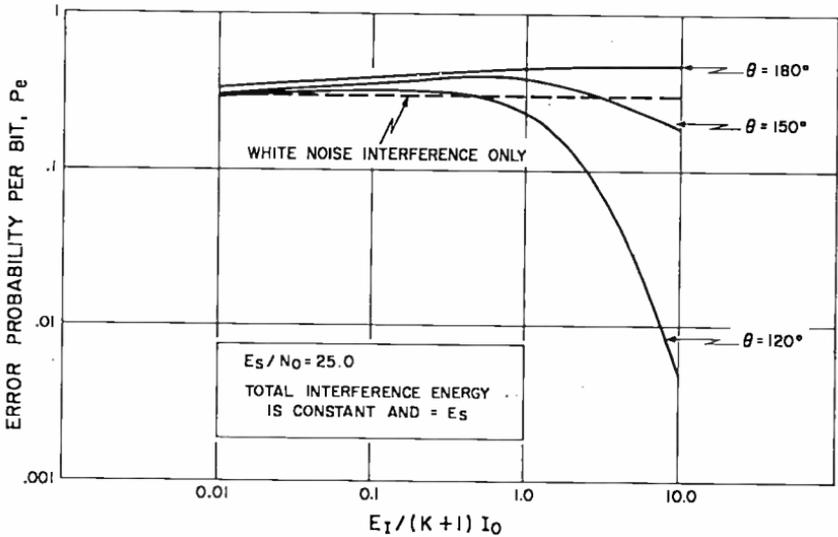


Fig. 6—Error probability versus  $E_I/[(K + 1)I_n]$  for constant total interference energy and three values of  $\theta$ .

signal interference. This is in agreement with the results of Figure 2. However, when the phase angle is decreased, optimum  $E_I/(K + 1)I_n$  is finite.

In Figure 7, the total interference energy is assumed to be  $E_n/2$ .

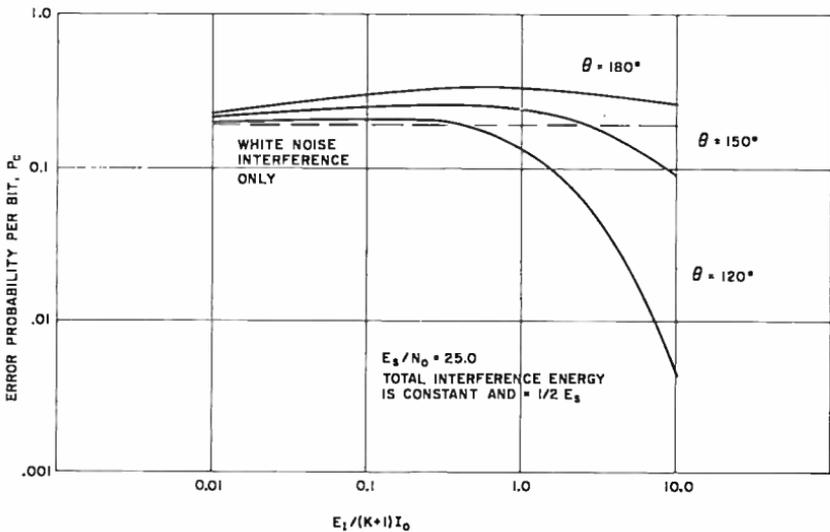


Fig. 7—Error probability versus  $E_I/[(K + 1)I_n]$  for constant total interference energy and three values of  $\theta$ .

For this case, highest interference occurs when  $E_I/(K+1)I_o$  is finite for all values of the phase angle.

In both figures, dashed lines indicate the error rates of white noise interference only, i.e., when all energy is transmitted in the form of white noise.

### Random Relative Phase

The remainder of the paper deals with the combined interference (white noise and similar signal) when both the relative phase and  $E_I/E_s$  vary randomly. Here a single value of  $P_e$  is meaningless. Instead of computing  $P_e$ , we look for the probability that  $P_e$  is equal to or greater than some fixed value  $z$ . The variation of  $P[P_e \geq z]$  will be determined.

It was shown previously that a uniform density distribution properly describes random variation of  $\theta$  (see Equation (11)). Two density distributions were considered for  $E_I/E_s$ , as given by Equations (13) and (14). Here, only the Rayleigh density distribution given by Equation (14) will be considered.

It was mentioned above that the transmitted signal energy and the total energy available to the jammer are assumed to be constant. With these assumptions,  $E_I/(K+1)I_o$  is varied. Therefore, the energy in similar signal and the white-noise density changes each time  $E_I/(K+1)I_o$  changes. Thus the average value of the random variable  $E_I/E_s$  varies with  $E_I/(K+1)I_o$ .

The probability of the error rate being equal to or greater than 0.1,  $P[P_e \geq 0.1]$ , was computed for total interference energy equal to  $E_s/2$  and for  $E_s/(2N_o) = 12.5, 25.0, \text{ and } 50.0$ . Results are shown in Figure 8. Each of the three curves plotted in this figure has a maximum. For each value of  $E_s/(2N_o)$  there is an optimum value of  $E_I/(K+1)I_o$  that produces the highest probability that  $P_e$  will be equal to or greater than 0.1.

The results of Figure 8 are replotted in Figure 9. Here, the percentage of time when the error rate is smaller than or equal to 0.1 is shown, rather than the probability that error rate is equal to or greater than 0.1.

### CONCLUSIONS

The following conclusions are drawn from the results obtained.

If at the receiver the relative phase,  $\theta$ , and the similar-signal interference to communication signal ratio,  $E_I/E_s$ , are fixed, the bit error

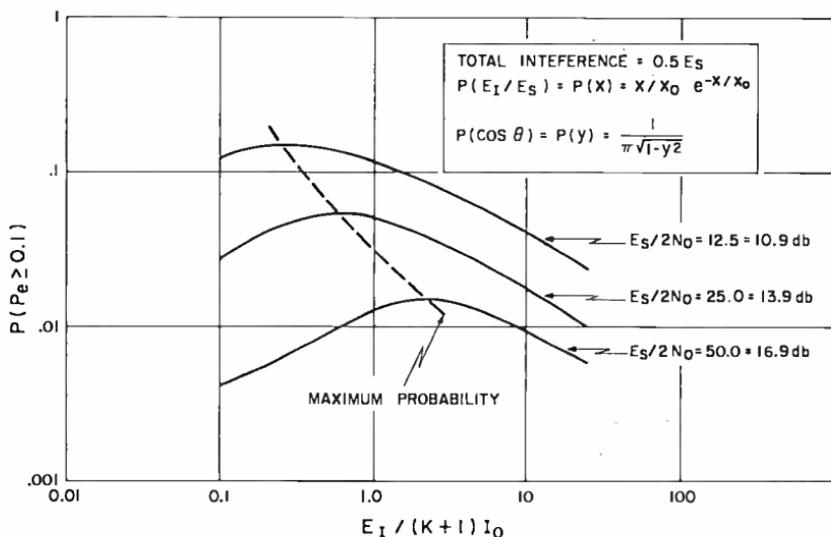


Fig. 8—Probability of error rate being equal to or greater than 0.1 versus  $E_I/[(K+1)I_0]$  (similar signal and white noise interference only).

probability depends on the magnitudes of these quantities. For the approximate interval  $110^\circ \leq \theta \leq 250^\circ$ , the similar-signal interference is more effective than white noise. (The end points of this interval are only slightly dependent on  $E/N_n$ .) Other values of  $\theta$  favor the communicator.

If the relative phase and interference to signal energy ratio are

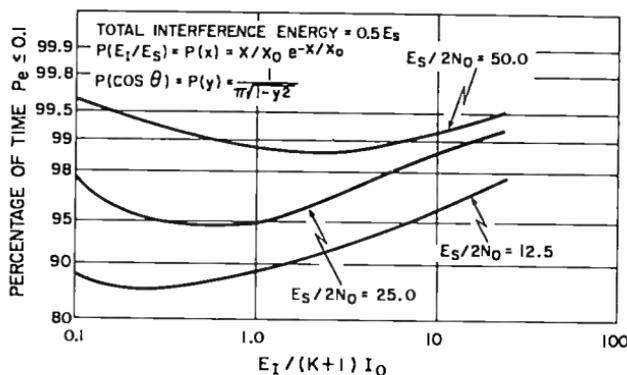


Fig. 9—Percentage of time  $P_e \leq 0.1$  versus  $E_I/[(K+1)I_0]$  (similar signal and white noise interference).

random at the receiver, similar-signal interference does not cause a higher error probability than white-noise interference of equal power.

Because of the small range of effective values of  $\theta$ , jamming with similar-signal interference can be expected only against fixed-terminal systems.

#### ACKNOWLEDGMENT

The author wishes to thank G. Lieberman for suggestions relating to the derivation of equations in the Appendix.

#### APPENDIX—ERROR RATES

A pseudo-random binary PSK signal and strong white gaussian noise are interfering with the reception of a semicoherent, spread-spectrum binary PSK communication system operating in natural noise environment. Both interference sources may be radiating simultaneously and continuously; the two interfering signals may arrive at the receiver interchangeably; one interference source may be in continuous operation, and the other one may be switching on and off periodically; finally, both interfering signals may be generated by the same transmitter simultaneously or interchangeably. The general expression for the combined signals at the receiver is

$$\begin{aligned} z(t) = & A_s \cos [\omega_0 t + \phi(t) + \theta_1] \\ & + A_I(t) \cos [\omega_0 t + \phi(t) + \theta_2] \\ & + B_I(t) I(t) + n(t), \end{aligned} \quad (15)$$

where

$A_s$  = desired signal amplitude,

$\omega_0$  = carrier frequency of the desired and interfering PSK signals,

$\phi(t)$  = modulation phase of the desired and interfering PSK signals,

$\theta_1$  = initial desired PSK signal phase,

$\theta_2$  = initial interfering PSK signal phase,

$A_I(t)$  = amplitude of the interfering PSK signal,

$B_I(t)$  = amplitude of the white-noise interference,

$I(t)$  = white noise of double-sided spectral density  $I_0$ ,

$n(t)$  = additive natural white gaussian noise.

**Case 1. Continuous Similar-Signal and White-Noise Transmission**

*Correlator Outputs*

After low-pass filtering (see Figure 1), the in-phase component of the received signal is

$$x(t) = A_s \cos [\phi(t) + \theta_1] + A_I(t) \cos [\phi(t) + \theta_2] + B_I(t) I_c(t) + N_c(t) \tag{16}$$

and the quadrature component is

$$-\hat{x}(t) = A_s \sin [\phi(t) + \theta_1] + A_I(t) \sin [\phi(t) + \theta_2] + B_I(t) I_s(t) + N_s(t) \tag{17}$$

Here,  $I_c(t)$  and  $N_c(t)$  are the in-phase components and  $I_s(t)$  and  $N_s(t)$  are the quadrature components of the white interference and natural noise, respectively.

Correlating with the stored reference  $S(t)$  and integrating for period  $T = nT_o$  ( $T_o =$  duration of a keying element, and  $n =$  number of keying elements in one information bit), one obtains

$$\begin{aligned} y_1 &= \int_0^T x(t) S(t) dt \\ &= \int_0^T A_s \cos [\phi(t) + \theta_1] \cos \phi(t) dt + \\ &\quad \int_0^T A_I(t) \cos [\phi(t) + \theta_2] \cos \phi(t) dt \\ &\quad + \int_0^T B_I(t) I_c(t) \cos \phi(t) dt + \int_0^T N_c(t) \cos \phi(t) dt \\ &= A_s T \cos \theta_1 + \sum_{i=1}^n \int_{(i-1)T_o}^{iT_o} A_I(t_i) \cos (\phi_i + \theta_2) \cos \phi_i dt \\ &\quad + \frac{1}{2W} \sum_{i=1}^{2Wn} B_I(t_i) I_c(t_i) \cos \phi(t_i) + \delta_1 \end{aligned} \tag{18}$$

where  $W$  is the bandwidth of the signal.

But  $\phi(t_i)$  is either 0 or  $\pi$  due to PSK modulation. For constant  $A_I(t_i) = A_I$ ,

$$y_1 = A_s T \cos \theta_1 + A_I T \cos \theta_2 + \frac{1}{2W} \sum_{i=1}^{2WT} B_I(t_i) I_c(t_i) \cos \phi(t_i) + \delta_1$$

$$= A_s T \cos \theta_1 + A_I T \cos \theta_2 + R(t) + \delta_1 \quad (19)$$

Similarly,

$$-\hat{y}_1 = \int_0^T \hat{x}(t) S(t) dt = A_s T \sin \theta_1 + A_I T \sin \theta_2$$

$$+ \frac{1}{2W} \sum_{i=1}^{2WT} B_I(t_i) I_c(t_i) \sin \phi(t_i) + \hat{\delta}_1 \quad (20)$$

However,  $\sin \phi(t_i) = \theta$ , because  $\phi(t_i)$  is either 0 or  $\pi$ . Thus,

$$-\hat{y}_1 = A_s T \sin \theta_1 + A_I T \sin \theta_2 + \hat{\delta}_1 \quad (21)$$

*Variance of  $y$ .*

The variance of  $R(t)$  in Equation (19) is

$$\sigma^2(R) = \left( \frac{1}{2W} \right)^2 \sum_{i=1}^{2WT} B_I^2(t_i) \overline{I_c^2(t_i)}. \quad (22)$$

Let  $I_o$  be the double-sided spectral density of  $I(t)$ . Then

$$\overline{I_c^2(t_i)} = \frac{I_o}{B_I^2 T_o}. \quad (23)$$

Then

$$\sigma^2(R) = \frac{I_o}{2WB_I^2 T_o} \int_0^T B_I^2(t) dt = \frac{I_o}{2WB_I^2 T_o} \overline{B_I^2(t)} T. \quad (24)$$

For constant  $B_I(t) = B_I$ ,

$$\sigma^2(R) = I_o T, \quad (25)$$

because  $2WT_o = 1$ .

### Error Rates

Let 
$$m = A_s T \cos \theta_1 + A_I T \cos \theta_2 \quad (26)$$

$$\dot{m} = A_s T \sin \theta_1 + A_I T \sin \theta_2. \quad (27)$$

It has been shown<sup>3,4</sup> that the bit error rate for a binary PSK quadrature detection system can be expressed by

$$P_e = \frac{1}{2} \exp \left[ - \frac{1}{4\sigma^2} (m^2 + \dot{m}^2) \right]. \quad (28)$$

But

$$\begin{aligned} \sigma^2 &= \sigma^2 [n(t)] + \sigma^2 [R(t)] = \sigma^2 (\delta_1) + \sigma^2 [R(t)] \\ &= N_o T + I_o T = T(N_n + I_n) \end{aligned} \quad (29)$$

Substituting Equations (26), (27), and (29) into Equation (28) and letting

$$\frac{A_s^2 T}{2} = E_s, \quad \text{signal energy per bit} \quad (30)$$

$$\frac{A_I^2 T}{2} = E_I, \quad \text{pseudo-random PSK interference energy per bit}$$

$$\theta_1 - \theta_2 = \theta, \quad \text{relative phase}$$

we obtain

$$P_e = \frac{1}{2} \exp \left\{ - \left[ \frac{E_s + E_I + 2 \cos \theta \sqrt{E_s E_I}}{2(N_n + I_n)} \right] \right\}. \quad (31)$$

<sup>3</sup> G. Lieberman, "Effects of Fading on Quadrature Reception of Orthogonal Signals," *RCA Review*, Vol. XXIII, p. 353, Sept. 1962.

<sup>4</sup> F. G. Splitt, "Comparative Performance of Digital Data Transmission Systems in the Presence of CW Interference," *I.R.E. Trans. on Communications Systems*, June, 1962, No. 2.

This equation gives the average bit error probability when both interference sources are radiating simultaneously and continuously. In case of a pseudo-random PSK signal interference only,  $I_o$  in Equation (31) is zero; when only white noise interference is present,  $E_I$  in Equation (31) is zero.

### Case 2. Discontinuous Similar Signal and White Noise of Limited Total Power

Consider one interference source with fixed total power radiating two types of signals in the following fashion: for a fraction of the keying element duration,  $KT_o$  (where  $0 \leq K \leq 1$ ), the entire power is radiated in the form of the white noise; for the remaining time of the keying element,  $(1 - K)T_o$ , the total available power is equally divided between the pseudo-random PSK signal and white noise. Then,

$$A_I(t) = \begin{cases} 0 & \text{for } 0 \leq t < KT_o \\ A_I & \text{for } KT_o \leq t \leq T_o \end{cases} \quad (32)$$

and

$$B_I(t) = \begin{cases} \sqrt{2}B_I & \text{for } 0 \leq t < KT_o \\ B_I & \text{for } KT_o \leq t \leq T_o \end{cases} \quad (33)$$

Employing Equations (32) and (33), one obtains

$$y_1 = A_s T \cos \theta_1 + (1 - K)A_I T \cos \theta_2 + R(t) + \delta_1, \quad (34)$$

$$-\dot{y}_1 = A_s T \sin \theta_1 + (1 - K)A_I T \sin \theta_2 + \dot{\delta}_1 \quad (35)$$

$$\overline{B_I^2(t)} = \frac{(\sqrt{2}B_I)^2 KT_o + B_I^2(1 - K)T_o}{T_o} = (K + 1)B_I^2 \quad (36)$$

and

$$\sigma^2 = T[N_o + (K + 1)I_o]. \quad (37)$$

With these expressions, the average error probability is

$$P_e = \frac{1}{2} \exp \left\{ - \left[ \frac{E_s + E_I + 2 \cos \theta \sqrt{E_s E_I}}{2[N_o + (K + 1)I_o]} \right] \right\} \quad (38)$$

where

$$E_I = \frac{A_I^2 T}{2} (K - 1)^2. \quad (39)$$

## RCA Technical Papers

Fourth Quarter, 1967

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Addendum to 'Failure Correction Decoding'" ( <i>Trans. IEEE PTGCOM</i> , February, 1967), H. E. White, <i>Trans. IEEE PTGCOM</i> (Correspondence) (October) .....	1967
"Conduction Bands in GaSb-InSb Alloys," I. Kudman and T. E. Seidel, <i>Jour. Appl. Phys.</i> (October) .....	1967
"Low-Temperature Vacuum Deposition of Homoepitaxial Silicon," L. R. Weisberg, <i>Jour. Appl. Phys.</i> (Communications) (October) .....	1967
"New UHF-TV Pylon Antennas," A. J. Galinus, <i>Broadcast News</i> (October) .....	1967
"Radiative Recombination in Melt-Grown n-Type, Ge-Doped GaAs," H. Kressel, <i>Jour. Appl. Phys.</i> (October) .....	1967
"Redundancy Reduction Applied to Coarse-Fine Encoded Video," G. P. Richards and W. T. Bisignani, <i>Proc. IEEE</i> (October) ....	1967
"Reliability Aspects of the Variable Instruction Computer," E. H. Miller, <i>Trans. IEEE PTGEC</i> (October) .....	1967
"System Utilization of Large-Scale Integration," S. Y. Levy, R. J. Linhardt, H. S. Miiller, and R. D. Sidnam, <i>Trans. IEEE PTGEC</i> (October) .....	1967
"Television Interlace Pairing: Its Effect on Detail Response and Its Measurement," W. L. Hurford, <i>Trans. IEEE PTGBC</i> (October) .....	1967
"Television's Role in Tomorrow's World," G. H. Brown, <i>IEEE Spectrum</i> (October) .....	1967
"Microstrip Plus Equations Adds Up to Fast Designs," A. Schwarzmans, <i>Electronics</i> (October 2) .....	1967
"Effect of Fermi-Level Motion on Normal-State Properties of $\beta$ -Tungsten Superconductors," R. W. Cohen, G. D. Cody, and J. J. Halloran, <i>Phys. Rev. Letters</i> (9 October) .....	1967
"Application of a Dual-Gate MOS Field-Effect Transistor to Linear Color Demodulation," W. S. Cranmer, <i>Trans. IEEE PTGBTR</i> (November) .....	1967
"A Comparison of Solar-Cell and Battery-Type Power Systems for Spacecraft," L. Pessin and D. Rusta, <i>Trans. IEEE PTGAES</i> (November) .....	1967
"Electrical Properties of Niobium-Doped Ferroelectric Pb(Zr,Sn, Ti) O <sub>3</sub> Ceramics," G. W. Taylor, <i>Jour. Appl. Phys.</i> (November) .....	1967
"A Feedback-Stabilized Vertical-Deflection Circuit," J. A. Dean, <i>Trans. IEEE PTGBTR</i> (November) .....	1967
"High-Performance Transistor Video Circuits for Color TV," W. M. Austin, <i>Trans. IEEE PTGBTR</i> (November) .....	1967

- "High-Temperature Thermal and Electrical Properties of GaSb-InSb Alloys," I. Kudman, L. Ekstrom, and T. Seidel, *Jour. Appl. Phys.* (November) ..... 1967
- "An Integrated Circuit for Remote Control Receivers," L. A. Harwood, *Trans. IEEE PTGBTR* (November) ..... 1967
- "Intermodulation Distortion Improvement in Parametric Upconverter," E. Markard, P. Levine, and B. Bossard, *Proc. IEEE* (Letters) (November) ..... 1967
- "Photoemission in Thin-Film Metal-SiO<sub>2</sub>-CdSe and Metal-SiO<sub>2</sub>-Metal Contacts," A. Waxman, *Jour. Appl. Phys.* (November) ..... 1967
- "Recombination Processes following Impact Ionization by High-Field Domains in Gallium Arsenide," P. D. Southgate, *Jour. Appl. Phys.* (November) ..... 1967
- "RF and IF Amplifier Design Considerations," R. A. Santilli, *Trans. IEEE PTGBTR* (November) ..... 1967
- "Semiconducting Properties of Ferromagnetic CdCr<sub>2</sub>Se<sub>4</sub>," H. W. Lehmann, *Phys. Rev.* (10 November) ..... 1967
- "Electromagnetic Excitation of Transverse Microwave Phonons in Metals," B. Abeles, *Phys. Rev. Letters* (13 November) ..... 1967
- "Special Report: The Many Facets of Gallium Arsenide. Devices III, Fast-Moving FET's Can Outpace Rivals," H. Becke, *Electronics* (November 13) ..... 1967
- "Special Report: The Many Facets of Gallium Arsenide. Materials I, Some Special Features Make the Difference," D. Richman, *Electronics* (November 13) ..... 1967
- "Special Report: The Many Facets of Gallium Arsenide. Materials III, An All-in-One Process for Building Junctions," J. J. Tietjen and L. R. Weisberg, *Electronics* (November 13) ..... 1967
- "Spatially Inhomogeneous Phonon Amplification in Solids," L. Friedman, *Phys. Rev.* (15 November) ..... 1967
- "The Acoustoelectric Effects and the Energy Losses by Hot Electrons—Part III," A. Rose, *RCA Review* (December) ..... 1967
- "Adaptive Detection with Regulated Error Probabilities," H. M. Finn, *RCA Review* (December) ..... 1967
- "An Adaptive Thin-Film Transistor," S. S. Perlman and K. H. Ludwig, *Trans. IEEE PTGED* (December) ..... 1967
- "Capacitance Changes in Thin CdS Crystals under dc Bias," B. Binggeli and H. Kiess, *Jour. Appl. Phys.* (December) ..... 1967
- "Catastrophic Degradation in GaAs Injection Lasers," H. Kressel and H. Mierop, *Jour. Appl. Phys.* (Communications) (December) ..... 1967
- "Coplanar-Contact Gunn-Effect Devices," R. Dean, J. F. Dienst, R. Enstrom, and A. Kokkas, *RCA Review* (December) ..... 1967
- "A Critique of Reading Improvement Training," W. J. Underwood, *Trans. IEEE PTGEWS* (December) ..... 1967
- "Crossed-Film Cryotron Switching Speed," A. R. Sass, W. C. Stewart, L. N. Dworsky, V. Hoffstein, and L. B. Schein, *Trans. IEEE PTGMAG* (December) ..... 1967
- "Crystal Growth and Electro-Optic Effect of Copper-Tantalum-Selenide, Cu<sub>2</sub>TaSe<sub>4</sub>," R. Nitsche and P. Wild, *Jour. Appl. Phys.* (Communications) (December) ..... 1967
- "Device Applications of Charge Rectifiers," T. F. Dwyer and R. Shahbender, *Trans. IEEE PTGEC* (December) ..... 1967
- "Effect of Noisy Synchronization Signals on Split Phase Coded System Performance," A. B. Glenn and G. Lieberman, *RCA Review* (December) ..... 1967
- "GaAs<sub>1-x</sub>P<sub>x</sub> Injection Lasers," J. I. Pankove, H. Nelson, J. J. Tietjen, I. J. Hegyi, and H. P. Maruska, *RCA Review* (December) .. 1967

- "Gelatin-Free Ion-Sensitive Plates for Mass Spectrography," R. E. Honig, J. R. Woolston, and D. A. Kramer, *Rev. Sci. Inst.* (December) ..... 1967
- "An Improved Cassegrain Monopulse Feed System," C. E. Profera, Jr. and L. H. Yorinks, *RCA Review* (December) ..... 1967
- "Inexpensive Audio Power Amplifiers Can Be Reliable," J. C. Sondermeyer, *The Electronic Engineer* (December) ..... 1967
- "Low-Temperature Photoconductivity in CdS," R. Crandall, *Jour. Appl. Phys.* (Communications) (December) ..... 1967
- "Lumped Elements in Microwave Integrated Circuits," D. A. Daly, S. P. Knight, M. Caulton, and R. Ekholdt, *Trans. IEEE PTGMTT* (December) ..... 1967
- "Methods of Deposition and Physical Properties of Polycrystalline II-VI Films," F. V. Shallcross, *RCA Review* (December) .... 1967
- "Microstrip Junction Circulator for Microwave Integrated Circuits," B. Hershenov, *Trans. IEEE PTGMTT* (Correspondence) (December) ..... 1967
- "The Origin of Photoelectrons in the Silver-Oxygen-Cesium Photocathode," A. H. Sommer, *RCA Review* (December) ..... 1967
- "Practical Aspects of Injection-Laser Communication Systems," W. J. Hannan, J. Bordogna, and D. Karlsons, *RCA Review* (December) ..... 1967
- "Preparation and Properties of Eutectic Bi-MnBi Single Crystals," W. M. Yim and E. J. Stofko, *Jour. Appl. Phys.* (December) ..... 1967
- "Process Techniques and Radiation Effects in Metal-Insulator-Semiconductor Structures," W. J. Dennehy, A. G. Holmes-Siedle, and K. H. Zaininger, *Trans. IEEE PTGNS* (December) ..... 1967
- "Response of Ion-Sensitive Plates as a Function of Ion Energy," J. R. Woolston, R. E. Honig, and E. M. Botnick, *Rev. Sci. Instr.* (December) ..... 1967
- "Role of Lithium in Damage and Recovery of Irradiated Silicon Solar Cells," J. J. Wysocki, *Trans. IEEE PTGNS* (December) ..... 1967
- "The Simulation of Time-Dispersed Fading Channels," L. W. Martinson and J. E. Courtney, *RCA Review* (December) ..... 1967
- "Solution of Ginzburg-Landau Equations for Arbitrary Tangential Magnetic Fields and Experimental Verification," G. A. Alphonse and Co-author, *Jour. Appl. Phys.* (December) ..... 1967
- "A Systems Look at Satellite-Borne High-Resolution Radar," J. S. Greenberg, *RCA Review* (December) ..... 1967
- "Visual Detection of Chart-Recorded Signals in the Presence of Noise," H. Blatter, *Trans. IEEE PTGGE* (December) ..... 1967
- "Visual Sensor Systems in Space," J. J. Dishler, *Trans. IEEE PTGCOM* (December) ..... 1967
- "Electron Hall Effect in Silicon Dioxide," A. M. Goodman, *Phys. Rev.* (15 December) ..... 1967
- "Setting It Straight," L. R. Weisberg and J. J. Tietjen, *Electronics* (Readers Comment) (December 25) ..... 1967

## AUTHORS



S. H. DURRANI received the B.A. and B.Sc.Engg. (Hons) degrees from the Punjab University, Lahore, Pakistan in 1946 and 1949 respectively, and a Master's degree from the College of Technology, Manchester, England in 1953. He taught at the Engineering College, Lahore, for most of 1949-51 and 1953-59, after which he came to the U.S. and obtained a doctorate from the University of New Mexico, Albuquerque in 1962. Dr. Durrani joined the Communications Products Department of General Electric Co. in Lynchburg, Va., where he worked on propagation and modulation problems in the Advance Engineering Group.

He served as Professor and Chairman of Electrical Engineering Dept. at the Engineering University, Lahore, for 1964-65. For a year after that he was at the Kansas State University, Manhattan, Kansas, as an Associate Professor and Associate Supervisor of a NASA research project on the use of radar as a remote sensor for geophysical studies. In June 1966 he joined the Astro Electronics Division of RCA at Princeton, N. J., as a Senior Engineer in the Systems Engineering and Advanced Projects group. His current work involves conceptual studies, definition, and analysis of various space communications systems.

Dr. Durrani is a Senior Member of IEEE, and a Member of IEE (London), AIAA, and Sigma Xi.

FAYSEL S. KEBLAWI received the BSEE degree in February 1957 from the American University of Beirut, Lebanon. He received the MSEE degree from North Carolina State University in 1963, and the Ph.D. degree from the same university in 1965. Dr. Keblawi joined RCA Astro-Electronics Division in 1966, and has been working on deep-space communication techniques and associated problems, with primary emphasis on the Voyager Spacecraft and probe Lander missions. Before joining RCA, Dr. Keblawi was engaged in consulting work for Kellogg, the transmission products division of International Telephone and Telegraph Co., where he worked on problems associated with frequency translators. Prior to obtaining the Ph.D. degree, Dr. Keblawi was an instructor at North Carolina State University. He is a member of the Institute of Electrical and Electronics Engineers, Phi Kappa Phi, Tau Beta Pi, and Eta Kappa Nu.





AKOS G. REVESZ received the Dipl. Ing. degree from the Technical University of Budapest in 1950. He was associated with the Physical Chemistry Department of this University as a Research Assistant and with the Iron Research Institute as a staff member. In 1951 he joined the Tungfram Company, Budapest, where he became head of the Semiconductor Development group. In 1957 he joined the Philips Company, Eindhoven, Holland, engaging in research and development of solid-state capacitors. He has been with the RCA Laboratories, Princeton, N. J., since 1959, working on thin-film problems, especially oxidation of semiconductors, epitaxial growth, and ellipsometry.

Mr. Revesz is a member of the American Association for the Advancement of Science, the American Physical Society, and the American Vacuum Society.

OTTO H. SCHADE was born and educated in Germany; he came to the United States in 1926. He joined RCA Electronic Components and Devices, Harrison, New Jersey, in 1931. Since 1938, he has specialized in television circuits, camera tubes, and picture tubes. From 1944 to 1957 he worked on a unified general method of image analysis and specification, including practical methods for measuring the "aperture" effect (square-wave and sine-wave response function) and fluctuation levels (noise) of optical, photographic, and electronic image-system components and the eye. He has had the responsibility for the thermal and electrical design of nivistor tubes. More recently, he has developed an accurate method for calculating the resolving power of television and photographic imaging systems to assist in the evaluation of high-definition television systems, and a new electron optic providing minimum aberrations and uniform focus in television camera tubes with larger ( $50 \times 50$  mm) image surfaces. Dr. Schade has received numerous honors, including the Modern Pioneers Award of the National Association of Manufacturers (1940), the Morris Liebmann Memorial Prize of the Institute of Radio Engineers (1950) and a Fellowship (1951) from the Institute of Radio Engineers. In 1951 he was made a Fellow, and also was the first recipient of the David Sarnoff Gold Medal Award, of the Society of Motion Picture and Television Engineers. In June 1953, he was invested with the honorary degree of Doctor of Engineering by Rensselaer Polytechnic Institute. In 1960, he received the Progress Medal Award of the Society of Motion Picture and Television Engineers for his outstanding technical contribution in the engineering phases of the motion picture and television industries.



HAROLD STARAS received the B.S. degree from CCNY in 1944, the M.S. from NYU in 1948, and the Ph.D. in Physics from the U. of Md. in 1955. He joined the National Bureau of Standards in 1948 and served as group leader in radio-propagation research and in the performance evaluation of radio systems. In 1954, he joined RCA where his initial assignment was in tropo-scatter propagation studies and in the systems design of tropo-scatter circuits. He has participated in basic studies of electromagnetic theory including "meteor-burst" communication, an island as a natural VLF slot antenna,

and dipole characteristics in magneto-ionic media. He performed analyses of the effect of high-altitude nuclear explosions on communication systems, and more recently, he succeeded in obtaining closed-form solutions for estimating the effect of a large extended surface on communications performance. During 1961-62 Dr. Staras was a Guggenheim Fellow and visiting professor at the Technion-Israel Institute of Technology.

V. F. VOLERTAS was graduated from the University of Kaunas, Lithuania in 1943 with a BS degree and from the University of Pennsylvania in 1954 with an MA; he also did graduate work at the University of Maryland. He joined RCA's Communications Systems Division in 1961 and began work on passive satellite communications studies. Later, his assignments included satellite systems analysis, electrical and optical correlation techniques, antijam techniques, analog and digital modulation, and filtering methods. Prior to his RCA employment, he was a Project Engineer on communications systems and performed airborne weapons system analysis and airborne radar performance analysis at Westinghouse. From 1943 to 1954, he taught mathematics and physics at European secondary schools. Mr. Volertas is a member of the Institute of Electrical and Electronics Engineers and the American Mathematical Society.



KARL H. ZAININGER received the B.E.E. degree (magna cum laude) from City College of New York in 1959; the M.S.E. in 1961, the M.A. in 1962, and the Ph.D. in Engineering Physics from Princeton University. In 1959 Dr Zaininger joined the staff of RCA Laboratories where he has been engaged in research on various semiconductor devices. He has been involved in research on silicon-based MOS devices since their original inception. His research activities with the MOS devices have included basic studies in the area of semiconductor surface physics.

He is presently concerned with MIS device physics and technology, with measurement techniques, and with the physics of radiation damage in MIS systems. Dr. Zaininger is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, the Institute of Electrical and Electronics Engineers, and the Shevchenko Scientific Society.





## STATEMENT OF OWNERSHIP

Statement of Ownership, Management and Circulation (Act of October 23, 1962, Section 4369, Title 39, United States Code).

1. Date of filing: September 28, 1967. 2. Title of publication: RCA REVIEW. 3. Frequency of issue: Quarterly. 4. Location of known office of publication: RCA Laboratories, Princeton, New Jersey, 08540. 5. Location of the headquarters or general business offices of the publishers: RCA Laboratories, Princeton, New Jersey, 08540. 6. Names and addresses of Publisher, Editor, and Managing Editor: Publisher, RCA Laboratories, Princeton, New Jersey; Editor, C. C. Foster, RCA Laboratories, Princeton, New Jersey; Managing Editor, R. F. Ciafone, RCA Laboratories, Princeton, New Jersey. 7. Owner: Radio Corporation of America, 30 Rockefeller Plaza, New York, New York, 10020. 8. Known Bondholders, Mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities: Metropolitan Life Ins. Co., 1 Madison Ave., New York, New York; New York Life Ins. Co., 51 Madison Ave., New York, New York.

I certify that the statements made by me above are correct and complete.

RALPH F. CIAFONE, *Managing Editor*

