JPL PUBLICATION 83-70

AD A138213

Product Assurance Technology for Procuring Reliable, Custom LSI/VLSI Electronics

Report for Period: May 1981 - October 1982

M.G. Buehler
T.W. Griswold
C.A. Pina
B.R. Blaes
C.C. Timoc
R.H. Nixon
S.F. Suszko

September 1983

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED (A)

DTIC FILE COPY

Prepared for
Defense Advanced Research Projects Agency
and
National Aeronautics and Space Administration
by
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

DTIC
ELECTE
FEB 23 1984
S A

84 02 22 067

JPL PUBLICATION 83-70

Product Assurance Technology for Procuring Reliable, Custom LSI/VLSI Electronics

Report for Period: May 1981 - October 1982

M.G. Buehler
T.W. Griswold
C.A. Pina
B.R. Blaes
C.C. Timoc
R.H. Nixon
S.F. Suszko

Contract #
NAS - 7 - 918
AO = 4665

September 1983

Prepared for

Defense Advanced Research Projects Agency

and

National Aeronautics and Space Administration

by

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

This document has been approved for public release and sale; its distribution is unlimited.

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Defense Advanced Research Projects Agency and the National Aeronautics and Space Administration.

ABSTRACT

The effort described in this report initiates the development of a product assurance methodology that is intended to be used in the process of qualifying reliable, custom, CMOS-bulk integrated circuits obtained from silicon foundries using a set of composite layout rules. A major element in this methodology is the microelectronic test chip that is fabricated along with the integrated circuits. This report describes the effort involved in further developing the test-chip methodology.

The major accomplishments include the development of a state-of-the-art parametric tester; the development of CMOS-bulk layout rules; the development of bulk test strips and test chips; the development of certain test structures, including the pinhole array capacitor for yield analysis and the split-cross-bridge resistor for process characterization; the development of a test-chip assembler software tool; the evaluation of two CMOS-bulk silicon foundry runs; and the identification of critical elements in procuring custom circuits from silicon foundries.

Letter on file

A-1

CONTENTS

1. INTRODUCTION 1-1

2. TECHNICAL ACCOMPLISHMENTS 2-1

2.1 PARAMETRIC TEST SYSTEM 2-1

2.1.1 Data-Acquisition System 2-1

2.1.2 Data Analysis 2-7

2.1.3 Data Validation 2-10

2.1.4 References 2-15

2.2 CMOS-BULK LAYOUT RULES 2-15

2.3 TEST CHIPS 2-25

2.3.1 Test-Chip Assembler 2-25

2.3.2 Test Chips and Test Strips 2-31

2.4 TEST STRUCTURES 2-42

2.4.1 Categories of Test Structures 2-42

2.4.2 Test Methods for a CMOS-bulk Test Strip 2-47

2.4.3 Split-Cross-Bridge Resistor 2-65

2.4.4 Yield Analysis Test Structures 2-87

2.5 CMOS-BULK FOUNDRY RUNS 2-99

2.5.1 Foundry Wafers and Schedule 2-99

2.5.2 Foundry Results and Conclusions 2-99

2.5.3 Data Analysis Tool 2-104

2.5.4 Reference 2-111

2.6 FAULT MODELS 2-113

2.7 PROCUREMENT TECHNOLOGY 2-121

2.8 TECHNOLOGY TRANSFER 2-124

3. DISCUSSION AND FUTURE OUTLOOK 3-1

4. APPENDIX 4-1

Figures

2.1.1-1 Data acquisition system 2-2

2.1.1-2 Relay matrix network 2-5

2.1.3-1 Example of a CRUNCH file illustrating the variation in n-channel transistor measurements 2-13

2.2-1 Cross-sectional drawing for a CMOS-bulk fabrication process 2-16

2.2-2 Six geometrical layout rule primitives and their relationship to photomask levels i and j 2-18

2.2-3 The 28 CMOS-bulk layout rules displayed using geometrical primitive figures 2-18

2.2-4 CMOS-bulk layout rules illustrated using descriptive diagrams 2-19

2.2-5 Matrix display of CMOS-bulk layout rules shown in Figures 2.2-3 and 2.2-4 2-20

2.3.1-1 Test structure location identified with a block of structures 2-28

2.3.1-2 Test structure documentation including structure location, parameters, and pad identification 2-29

2.3.1-3 Block locations within a test chip 2-30

2.3.2-1 Test strip 8205 2-32

2.3.2-2 Parametric test chip 2012 2-33

2.3.2-3 Random fault test chip 2011 2-34

2.3.2-4 Test strip 8207 2-35

2.3.2-5 Parametric test chip 2071 2-36

2.3.2-6 Random fault test chip 2072 2-37

2.3.2-7 Radiation test chip 2-38

2.4.2-1 Layout of a four-terminal n-channel MOS transistor (N - XT) 2-49

2.4.2-2 Channel leakage current, IDSO, measurement circuit 2-49

2.4.2-3 Source and drain diode leakage, IDBLEAK, measurement circuit 2-50

Figures

2.4.2-4 Source and drain diode breakdown voltage, VDBBD, measurement circuit 2-51

2.4.2-5 Gate leakage current, IGBLEAK, measurement circuit 2-51

2.4.2-6 Transistor parameter (VTO, KP, and GAMMA) measurement circuit 2-52

2.4.2-7 Maximum slope approach to evaluating VTO and KP for an NMOS transistor 2-53

2.4.2-8 Layout of a four-terminal, field-oxide, metal-gate n-channel MOS transistor (N - FMXT) 2-55

2.4.2-9 Field-oxide threshold voltage, VTFIELD, measurement circuit 2-55

2.4.2-10 Layout of a CMOS-bulk inverter (INV) 2-56

2.4.2-11 Inverter, VOH and VOL, measurement circuit 2-57

2.4.2-12 Inverter threshold voltage, VTINV, measurement circuit 2-58

2.4.2-13 Inverter gain, G, measurement circuit 2-58

2.4.2-14 Layout of an n-diffused cross-bridge resistor (ND - XBR) . . . 2-60

2.4.2-15 Sheet resistance, RS, measurement circuit 2-61

2.4.2-16 Linewidth, W, measurement circuit 2-62

2.4.2-17 Layout of a metal to n-poly contact resistor (M/NP - CR) . . . 2-63

2.4.2-18 Contact resistance, RC, measurement circuit 2-64

2.4.4-1 Contact array resistor (CAR) 2-88

2.4.4-2 Wafer maps of the resistance per contact element of the seven resistor strings of the CAR (wafer 1213) 2-90

2.4.4-3 An example of the defect type that is excluded from the CAR analysis 2-91

2.4.4-4 Photomicrographs of nonlocal global photolithographic-related flaws in contact resistors arrays 2-92

2.4.4-5 Pinhole array capacitor (PAC) 2-93

2.4.4-6 Photomicrographs of the PAC 2-95

2.4.4-7 Wafer maps of the six subarrays found in the PAC 2-97

Figures

2.4.4-8 Photomicrograph of the PAC, showing a nonrandomly occurring defect that has caused a short between the metal and poly layers (site row 5, column 7, wafer 1104) 2-98

2.4.4-9 Yield analysis of the PAC data shown in Figure 2.4.4-7 (wafer 1104) 2-98

2.5.1-1 Silicon foundry CMOS-bulk wafers 2-100

2.5.2-1 Wafer maps of CMOS-bulk inverters and contact resistors showing that some inverter outputs are stuck low due to open metal to p-diffusion contacts (wafer 1102) 2-105

2.5.2-2 Wafer maps of CMOS-bulk inverters from run 1 showing the regions where the inverters failed to toggle properly 2-105

2.5.2-3 Wafer maps of CMOS-bulk inverters from run 2 showing the regions where the inverters failed to toggle properly 2-106

2.5.3-1 Data integrity assurance scheme 2-108

2.5.3-2 Data analysis procedure 2-109

2.5.3-3 N-transistor threshold voltage wafer map and histogram in which data less than 10^{-9} volts have been excluded from the data set 2-110

2.5.3-4 N-transistor threshold voltage wafer map and histogram in which invalid data (data less than 10^{-9} volts) and outlier data have been excluded from the data set 2-110

2.7-1 Flow diagram describing the procurement of reliable, custom circuits from silicon foundries 2-122

Tables

2.3.2-1 Summary of CMOS-Bulk Test Chips and Foundry Runs 2-40

2.3.2-2 Example Plan File Documentation for Several Inverters in the Test Chip Shown in Figure 2.3.2-2 2-41

2.5.2-1 CMOS-Bulk Foundry Parametric Characterization (Run 1) 2-101

2.5.2-2 CMOS-Bulk Foundry Parametric Characterization (Run 2) 2-102

2.5.2-3 CMOS-Bulk PAC Yield Characterization 2-103

2.5.3-1 Statistical Analysis/Wafer Map Program Commands 2-106

SECTION 1

INTRODUCTION

The effort described in this report initiates the development of a product assurance methodology that can be used in the process of qualifying CMOS-bulk integrated circuits obtained from a number of silicon foundries using a set of composite layout rules.

This effort is an extension of a previous DOD-sponsored effort to establish certain standard interfaces for use in obtaining custom, integrated circuits from silicon foundries. In the previous effort the emphasis was on the development of test chips and test methods for CMOS-SOS. In this effort the knowledge gained previously was extended to the development of a product assurance technology for the CMOS-bulk processes available at silicon foundries.

The major accomplishments of this effort included the development of a state-of-the-art parametric tester; the development of CMOS-bulk layout rules that are being used by the DARPA university community; the development of CMOS-bulk test strips and test chips; the development of certain test structures, including the pinhole array capacitor for yield analysis and the split-cross-bridge resistor for measuring sheet resistance, linewidth, line spacing, and pitch of conducting layers; the development of a test-chip assembler tool for the rapid creation of test chips with different layout rules; the initiation of the first CMOS-bulk silicon foundry run for the DARPA community; the evaluation of two CMOS-bulk silicon foundry runs; and the identification of critical elements in the procurement of custom circuits from silicon foundries.

The goal of this effort is to develop the methodology that will provide a low-cost approach to qualifying custom LSI/VLSI components. Our methodology is based on a wafer-level acceptance plan rather than a lot acceptance plan. This conclusion was reached after it was observed that the semiconductor fabrication process can be highly variable. That is, results from one wafer cannot be reliably used to indicate results from another wafer. In our methodology, the test strip and test chip are used to generate wafer maps of critical parameters that characterize various aspects of the semiconductor fabrication process. From these parameter wafer maps we intend to establish those regions on a wafer where potentially good circuits can be found.

In the future we intend to extend this effort to include wafer-level reliability tests on test strips and test chips that can be used to evaluate critical failure mechanisms. These reliability tests will be combined with circuit performance tests to provide a comprehensive component acceptance plan.

ACKNOWLEDGMENT

The authors wish to acknowledge the interest, guidance, and many helpful suggestions of this effort's sponsors, namely Harry Sonneman, NASA

Office of the Chief Engineer, and Paul Losleben, of DARPA's Information Processing Technology Office. The helpful suggestions of Jim Osterritter, DOD, are also greatly appreciated. In addition, we wish to acknowledge the efforts of USC's MOS Implementation Service in obtaining some of the fabricated wafers used in this study.

SECTION 2

TECHNICAL ACCOMPLISHMENTS

2.1 PARAMETRIC TEST SYSTEM

2.1.1 Data-Acquisition System

2.1.1.1 Design Objectives. Our primary design goal was to develop a laboratory instrument for characterizing test structures on integrated-circuit wafers prior to chip separation and packaging. Using a standard probe-pad array, we were able to access all structures on the wafer with a single probe card. We wanted to be able to move to a test structure, make contact with the pads, acquire and store data, and then break contact and move to another site, all under computer control initiated by an operator or by a command file.

2.1.1.2 System Overview. A block diagram of the data-acquisition system is shown in Figure 2.1.1-1. It consists of the following major components:

1. A wafer-handling assembly, consisting of an automatic wafer prober and a probe card for making electrical contact to the wafer under test.

2. An instrument assembly, comprised of voltage and current sources, voltage and current meters, and an impedance meter.

3. A relay matrix for connecting instrument terminals to probe-card test points.

4. A test controller, consisting of a DEC LSI-11/23 computer and associated peripheral devices, including operator terminals, a disk system for program and data storage, and a printer.

5. A data link to a DEC VAX 11/780 computer.

The system is used as follows: A wafer is placed on the vacuum table of the wafer prober and brought into alignment with the table motion. A test program is loaded into the LSI-11/23 computer and started. Under program control, the computer moves the wafer-prober table to place a selected test device in position for measurement. The wafer table is raised to bring the probe pads of the test device into contact with electrical probes. Contacts in the relay matrix are closed to form measurement paths between probe pads and instrument terminals. Commands are issued to the instruments to set up measurement parameters and modes. The computer issues enabling commands to the instruments as required, reads data values returned by the meters, stores the data in memory or in a disk file, and finally issues disabling commands. The computer then proceeds to the next measurement.

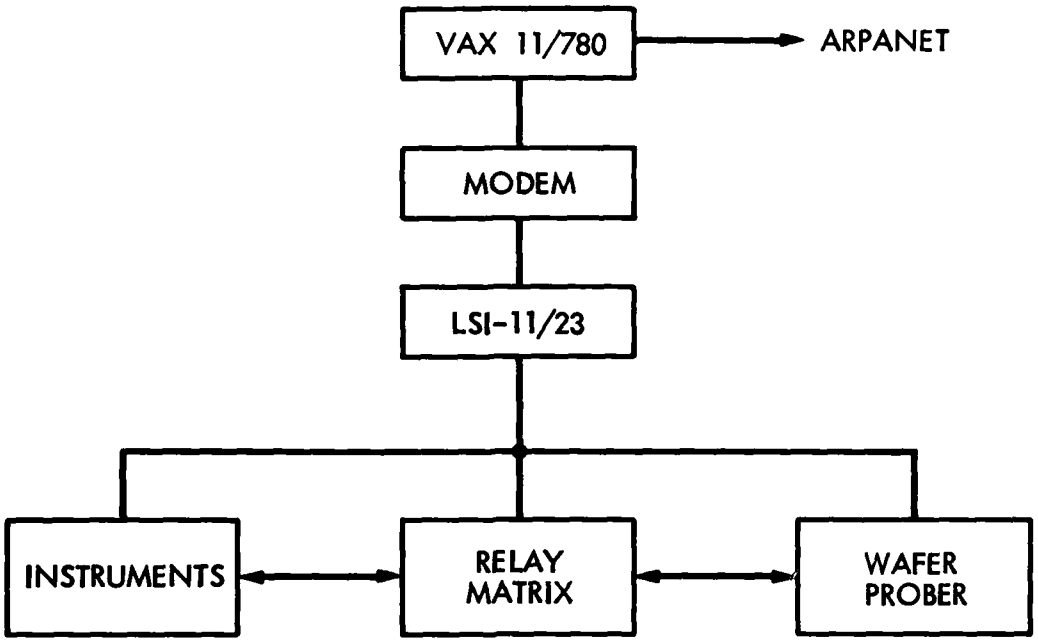


Figure 2.1.1-1. Data acquisition system.

2.1.1.3 System Elements

2.1.1.3.1 Wafer-Handling Assembly. The wafer prober consists essentially of a movable vacuum table and a probe-card holder. Its function is to position the wafer horizontally so that a particular location is directly under a set of test points on the probe card, and to raise the table so that the test points contact probe pads on the wafer. In a typical machine, horizontal motion is carried out by a pair of stepping motors driving lead screws in x- and y-directions. Step sizes are in the neighborhood of 5 or 10 micrometers (0.25 or 0.5 mil). The wafer table is equipped with a rotating arrangement for aligning the chip rows with the table motion. The vertical motion is adjusted for a slight springing action of the probe points when they contact the wafer. This springing action causes a scrubbing of the points on the pads for good electrical contact. The scrubbing action damages the pads somewhat, and limits the number of times that a particular device can be measured reliably.

The prober in the JPL system is a Teledyne TAC PR-52. It has a 10-micrometer step size, and can handle a 10-centimeter (4-inch) wafer. It has been fitted with a metal box over the wafer table to shield it from room light and to permit eventual addition of an atmosphere control system. A series of measurements is initialized by using the wafer-prober's front-panel controls to position the wafer to a preselected location (chip row and column, and position within the chip). The test program then takes over and controls subsequent

wafer motions. A particular location is reached by calculating the number and sense (plus or minus) of x- and y-steps required and issuing the appropriate pulse sequences to the stepping motors.

This application of an automatic wafer-prober is different from its most common use in industry--the screening of chips prior to packaging. The measurement procedures described here use only the primitive prober functions, namely, move one step in the +x, -x, +y, or -y direction, and raise or lower the wafer table. No use is made of inking or edge-sensing prober functions, and the front-panel controls for chip indexing and table motion are used only for convenience in alignment and initialization.

The probe card is an etched circuit board with a hole in its center. The probe points are sharpened metal wires mounted around the edge of the hole and project downward through it. They are positioned so that their ends match the bonding-pad pattern on the wafer. The card plugs into a standard card-edge connector. Small, 50-ohm coaxial cables are used between the card connector and a panel of BNC jacks mounted on the covering box. Another piece of coaxial cable connects the wafer table to the BNC panel; this is the substrate connection. Measurements can be made from DC to a few megahertz. To go to higher frequencies, the probe card must be redesigned for better high-frequency properties. In addition, a solution will be needed to the problem of impedance matching--integrated-circuit impedances are of the order of 3000 ohms (5 volts into one low-power TTL load), while standard transmission systems and many high-frequency instruments have a 50-ohm characteristic impedance. One cannot simply hang a 50-ohm resistor across a MOS driver to match the network impedance, because a 50-ohm resistor draws 200 milliamperes at 5 volts. This problem can be solved in any particular case by designing a specific matching network, but generality is lost--a single matching network will not serve all devices and all measurements.

All of the test modules described in this report use a standard probe-pad array having two rows of 10 points each on a 160-micrometer grid. A single probe-card design can thus be used for all test devices on the wafer. This standardization avoids a common problem in testing: not being able to make measurements until a special probe card has been designed and fabricated for the chip under test. Delivery of special probe cards is typically two to four months after placement of the order.

The 2 X 10 probe-pad array is a particular case of the generalized "2 X N" standard array developed at the National Bureau of Standards, where it was shown that pad sizes as small as 60 micrometers could be used [1]. The 80-micrometer pad size was selected as a reasonable compromise between test-pattern density and probing reliability, and N was fixed at 10, on the grounds that 20 pads were enough for the most complex test structure contemplated at that time. The 160-micrometer grid size is a fixed standard, as is the two-row arrangement. The length of the array (N) can be changed without causing severe problems. There are two considerations in selecting the length of the array: (1) there must be enough pads so that all of the test-device terminals can be accessed in a single set-down of the probe, and (2) unused probe-point locations should not contain test-device circuitry, since probe points can damage the surface of the wafer when they contact it.

The 80-micrometer (3.2 mils) pad size is rather small, and can cause difficulty in testing; a 10-centimeter (4-inch) wafer is 100,000 micrometers in diameter, while the placement tolerance for probe points is 40 micrometers. Extreme care must be taken in setting up the measurement so that rotational and translational errors of the wafer table, z-axis positioning, wafer taper, and manufacturing/maintenance tolerances of the probe-card-point locations do not combine to cause one or more probe points to miss their target pads somewhere on a large wafer. It has proved necessary, in fact, to include contact checks in the test programs wherever possible to detect this kind of measurement error. This precaution is of critical importance for test structures that are used to detect fabrication faults and estimate process yields.

2.1.1.3.2 Instrument Assembly. Details of the instrumentation and relay matrix are shown in Figure 2.1.1-2. The instruments provide the following functions:

1. Establish (force) desired current or voltage level.

2. Measure current or voltage.

3. Measure impedance (resistance, capacitance, inductance).

The instruments are:

Count	Instrument	Force	Measure
3	Analogic MP8116 DAC	Voltage	
1	Hewlett-Packard 6131C	Voltage	
1	Keithley 220	Current	
1	Hewlett-Packard 3456A		Voltage 2-wire Resistance
1	Keithley 619		Voltage Current
1	Hewlett-Packard 4192A	Signal	Conductance Capacitance

The DACs are 16 bits in resolution, jumpered for $\pm$ 10 volts, and are used as general-purpose voltage sources. They have good linearity over the entire 16-bit range and can supply about 2 milliamperes, which is adequate for most MOS device measurements. They are optically isolated from the computer data lines.

The Keithley 220 current source is used in the measurement of junction (diode) breakdown voltage, and for resistance measurements that require a particular value and direction of current.

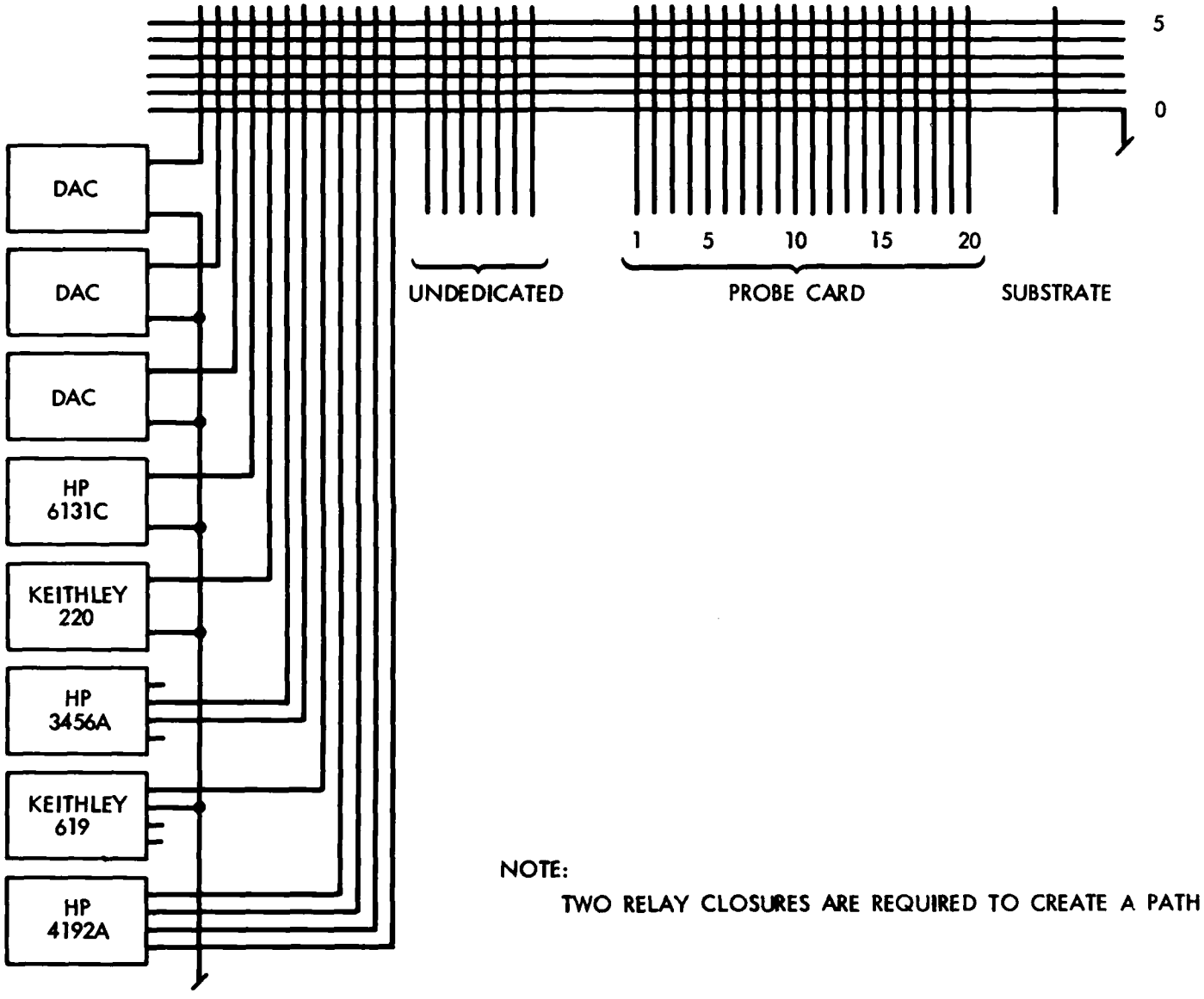


Figure 2.1.1-2. Relay matrix network.

The Hewlett-Packard (HP) 3456A is a general-purpose voltmeter and ohmmeter. Its input resistance as a voltmeter on the 0.1-, 1.0, and 10-volt scales is greater than 10,000 megohms. It can be programmed for optimum balance between reading rate and measurement accuracy. Its four-wire resistance capability is not used in the JPL system because the measurement current cannot be controlled independently; all four-terminal resistance measurements are made by explicitly connecting a voltage (current) source and measuring the current (voltage) with one of the meters. Its two-wire resistance capability is used only for checking the network for bad contacts.

The Keithley 619, an electrometer, is used for all current measurements, and for voltage measurements that require high input resistance. It is slower (in readings/second) than the HP 3456A, and therefore is not used in its voltage mode for general-purpose measurements. It has two channels, but only one is used in the JPL system.

The HP 4192A is a recent addition, replacing a Boonton 72BD capacitance meter. The Boonton is capable only of measuring capacitance at one megahertz, and does not measure the quadrature component. The HP 6131C was used formerly as the bias supply for the Boonton 72BD. It has a $\pm$ 100-volt capability, and is now available as a general-purpose high-voltage supply.

2.1.1.3.3 Relay Matrix. The relay matrix is a critical part of the system because it must make low-resistance and low-noise connections between any instrument terminal and any probe-card pin, as required by the test program. The matrix in the JPL system is manufactured by Matrix Systems Corporation of Calabasas, California. It is a coaxial (as opposed to triaxial) system, configured as in Figure 2.1.1-2. Each of the six measurement paths can be connected to any instrument terminal, to any probe pin, and to the wafer table (i.e., the back of the wafer under test), in any combination. Path 0 is permanently dedicated to analog ground; paths 1 - 5 are used as required for connection of instrument terminals to probe pads, and for connection of probe pads to other probe pads or to the substrate (the back of the wafer). The characteristic impedance is 50 ohms.

The matrix is made up of modules, each containing 20 mercury-wetted reed relays. Two modules form a path; six pairs of modules form the matrix. Each module consists of an aluminum block with 20 holes bored out to hold the 20 relays. One end of each relay goes to a BNC connector, and the other end to the common conductor. The common line is brought out through a BNC connector. Modules are placed side by side, with the relay connectors aligned. Forty crossbar straps are placed perpendicular to the modules. Each strap has connectors that mate with the relay connectors, and an end connector that is one of the matrix's 40 terminals. Relay-selection logic and driver electronics complete the assembly. Additional paths, up to a maximum of 10, can be created by adding relay modules in pairs.

The total effective coaxial-cable length between instruments and probe card is a little under two meters. Keeping the connection short was necessary for capacitance measurements with the Boonton 72BD because longer cable lengths impair the accuracy of this instrument, according to information in the manual. The wafer prober was placed close to the instrument rack that

holds the relay matrix, and the matrix was mounted at the height that made the shortest cable runs. With the change to the HP 4192A, this requirement has been eased somewhat, but it is still good practice to keep the cable lengths as short as possible.

The relay matrix has proven to be quite satisfactory in the JPL system. Two-wire channel resistances are about 4 ohms, as measured with the HP 3456A, and capacitance noise is quite low--a few femtofarads at most. There have been occasional problems with noisy channels, but these have been identified as connector problems, either at the probe-card edge connector, or in the coaxial interconnections of the matrix. Such problems have been easy to detect, locate, and repair.

2.1.1.3.4 Test Controller. The test controller is a DEC LSI-11/23 computer, running under the RT-11 operating system. (An upgrade to RSX-11 is planned.) The interface to most of the instruments and to the wafer prober is IEEE 488. Direct parallel ports are used for the relay matrix, the capacitance meter, and the DACs.

All software above the level of the operating system and the device drivers has been written by JPL in FORTRAN. A Pascal compiler (Oregon Software Systems) has been procured for installation when RSX is operational. It will be used for high-level programming, including operator interface and control, test-program generation and compilation, and data-file organization. The Oregon Pascal compiler can call FORTRAN subroutines, so the low-level software can be in either language, and existing FORTRAN modules can be used wherever they are suitable.

2.1.1.3.5 Data Link. Files are transferred to and from the VAX 11/780 by a 9600-baud data link. The VAX runs under the VMS operating system. A network connection to the VAX, through DECnet or a local-area network such as ETHERNET, will be established in the future (when RSX is operational). This link will allow the test system to access the VAX facilities directly, and thus will make the most of both computers.

2.1.2 Data Analysis

2.1.2.1, Objectives. When used in carrying out separated design and fabrication, measurements of the properties of test structures on wafers, as discussed in this section, have principally two purposes. The first objective is to support the procurement of fabrication services, establish the qualifications of vendors to provide services, and evaluate the integrity of processing for a given wafer lot to determine whether or not it meets preestablished criteria for acceptance. The second objective is to build a base of information about device and process characteristics representing many wafers, many vendors, and many technologies, to guide future decisions. For both purposes, the information will be useful and meaningful only to the extent that the methods used to obtain data and to distill and interpret it are sound, reproducible, and well-documented. The work to be described in the remainder of this section has had the following objectives:

1. Establish uniform and explicit procedures for conducting a set of measurements and acquiring data from test modules.

2. Establish methods of validating data to insure that they represent properties of the test structures, not artifacts of the measurement process.

3. Define criteria and methods to be used in "screening," that is, eliminating nonrepresentative data.

4. Define some useful standard procedures for the analysis, interpretation, and presentation of data.

2.1.2.2 Test Programs and Data Formats. Test programs for controlling data acquisition from test strips are written in FORTRAN. The results are stored in random-access binary files, with the data for each chip in a separate record. A separate file is written for each wafer. The record provides general information about the wafer, such as the number of records in the file, and includes comments (ASCII character strings).

Test-chip structures are generally measured by means of an interpreter, written in FORTRAN, named "CRUNCH" [2]. It was originally written at JPL for analysis of numerical data from the multichannel analyzer of an X-ray photoelectron spectrometer. Its basic data structure is the array, and it has a number of built-in commands for array manipulation and display. Formally, test-strip and test-chip data look like multichannel analyzer data, so CRUNCH was adapted for wafer-prober measurements. Since it is an interpreter, CRUNCH is very slow in comparison to the compiled programs used for test strips, but it has the advantage of permitting quicker and easier test-program development and debugging. CRUNCH test programs may be entered and run directly from the keyboard, or they may be loaded from command files and executed automatically or under keyboard control. Data files are written in ASCII, using standard CRUNCH format. This format is also used by JPL for data analysis and display. It is described below.

Test programs, both compiled and CRUNCH, control the wafer-prober motions, the relay-matrix connections, and the test instruments, usually in that order.

The wafer-prober position, and some of the matrix connections, are determined by (1) the location of chips on the wafer, (2) the location of the selected pad block within the chip, and (3) the probe pads to which the selected test module is connected. These items of information are obtained from the test-chip design file and from data supplied by the organization that assembled the test strip and test chips into the composite wafer for fabrication. If this information is not available, it can be determined by measurements on the fabricated wafer itself.

The remaining matrix connections, together with the test-instrument functions, depend on the kind of measurement that is to be made. Transistors, for example, have either three or four probe pads assigned to them in a pad block: GATE, DRAIN, SOURCE, and BODY (for four-terminal transistors). Each kind of transistor measurement uses a different set of paths through the relay matrix for connecting the instruments to the probe pads: threshold voltage,

leakage current, breakdown voltage, etc. Any number of measurements may be taken for each set of connections. Leakage currents and breakdown voltages typically are measured with a single reading (apply voltage or current, read current or voltage). Threshold voltages, on the other hand, involve a number of readings, because the inflection point ("steepest-slope") of the ID-VG curve must be found by a search algorithm before the actual value can be determined. A slope and an intercept, calculated at the inflection point, yield both threshold voltage (VT) and transconductance (KPRIME). Details of the measurement procedures for individual test modules are given in Section 2.4.2.

Driver modules were written for the wafer prober and the relay matrix; they are linked to the compiled test routines for test strips, providing complete control of the test system. A separate file (the "location file") is read by the test program for control of the wafer prober. The location file contains the relative positions of chips with respect to a wafer origin, and the relative positions of the pad blocks with respect to the chip origin. Any or all of the test strips on the wafer may be selected for measurement, and the order of measurement may be arranged as required. As a general rule, all test strips on the wafer are measured, but the order is often varied to provide a picture of the wafer as quickly as possible.

CRUNCH was extended to include commands for control of the wafer prober, the relay matrix, and the test instruments. Thus, a particular measurement may be carried out by a compiled program or by CRUNCH. Transistor threshold voltages are usually measured with compiled programs, because CRUNCH is very slow in making the measurements and doing the calculations involved in finding the ID-VG inflection point.

CRUNCH data files consist of FORTRAN records in ASCII format. The first record contains the file name. The next three records (which are always present) are arbitrary header lines for descriptive information. They may be empty. There follow one to 10 data files, organized as follows:

Line				
1	DATA TYPE	ARRAY NAME		(SIZE)
2	value(1)	value(2)	value(3)	...
...	value(...)	value(...)	value(...)	...
n	...	value(SIZE)		

DATA TYPE is either INTEGER or REAL; these strings serve as array delimiters. CRUNCH actually stores integer data in real format, and the data file may contain it in either form. ARRAY NAME is one to eight alphanumeric characters (beginning with a letter). ARRAY SIZE is an integer value, enclosed in parentheses, that must agree with the number of data items in the list. The first line of each array is essentially a FORTRAN declaration, and it must be in a single record (i.e., on one line). The list may be on any number of lines. Except for the restriction of the declaration to one line, the format is free, with commas or any number of blanks serving as declaration and list delimiters. (CRUNCH array lists are typically written and read in FORTRAN programs with the list-defined format, e.g., READ *,list.)

Since CRUNCH files are ASCII files, they are very large in comparison to binary files that contain the same data. They have the offsetting advantages of being human-readable and portable. They can be accessed by text editors, a useful property when one needs to modify the descriptive header lines, or to change the locations of the record boundaries of data items to improve the display format (e.g., to shorten display lines from 132 characters to 80 characters).

CRUNCH can also be used for data display by using a built-in command that plots one or more arrays against another on a video monitor (using a Matrox or a Peritek board to drive the display). Another built-in command dumps the video display to a dot-matrix printer. Axes are automatically scaled to the minimum and maximum limits of the data, and are labeled with the arrays' names. When several arrays are plotted, the vertical axis is labeled with the first array name given. Scaling of the vertical axis can be expanded by including an artificial array that has values equal to the desired axis limits.

CRUNCH is maintained and distributed by JPL. At present, it is available only for the RT-11 operating system, which runs on the PDP-11 and LSI-11 machines manufactured by Digital Equipment Corporation. (Some of its display commands are specifically for VT-100 terminals.) It will be modified for the RSX-11 operating system in the near future.

2.1.3 Data Validation

Automatic data acquisition is usually unattended data acquisition. The system is set up, initialized, and started. During a measurement, errors of various kinds can invalidate the results. One of the primary tasks of the measurement process is to identify and flag data that may be erroneous, in order to separate test-system and measurement artifacts from real physical anomalies and faults in the fabricated test structures. Some of the errors can be detected by looking for known indicators during a run, or by noting the state of the system when the run is terminated; others cannot. To avoid errors that do not clearly reveal themselves (or ones that do), it is necessary to identify their sources and take steps to reduce or eliminate them by proper test-structure design and by good measurement technique.

For the purposes of this discussion, it is convenient to place the most common errors within three groups: errors that occur in wafer-probing, errors that result from faulty measurement techniques or instruments, and errors that result from poorly designed or fabricated test structures.

2.1.3.1 Probing Error Sources. The most common source of measurement error is failure to make proper contact between the probe-card points and the probe pads. "Proper contact" means accurate positioning of the probes on the probe pads so that a "low" resistance electrical contact is obtained between the probe points and the pads.

Probe-point placement must be maintained within the 80-micrometer pad size over distances on the order of the wafer diameter, which is 10^5 micrometers for a 10-centimeter (4-inch) wafer. The tolerable error is less than half the probe-pad size, or 40 micrometers; this translates to a relative error less than 0.04 percent. Factors in probe-placement error include:

1. Rotational alignment of the wafer. If the probe-pad rows and columns are misaligned with respect to the wafer table's x-y motion, probing error can result over a portion of the wafer.

2. Initial location of the probe points within the probe pads. The points on the probe card can be offset from the probe pad centers. This error, when combined with rotational misalignment, can lead to the probes failing to contact the pads in a portion of the wafer.

3. Run-out of the probe points across the wafer. The stepping interval of the wafer prober may not match the interval between test sites on the wafer.

4. Vertical probe displacement. The probe-card points are adjusted by the manufacturer for a specified displacement after contact with the probe pads is made. This provides a scrubbing action of the probe points when they touch the pads, thus ensuring good electrical contact. The points bend, and their horizontal locations change, during probe set-down. Compensation for this bending is included in the initial adjustment of the probe-point locations. Too little displacement can result in high-resistance contacts, and too much displacement can result in the probes sliding off the pads.

5. Wafer taper (wafer surface not parallel to x-y motion). Correct vertical displacement of the probes depends on a uniform height of the wafer surface above the wafer-table surface. If the wafer is tapered, or if there is a piece of dirt under the wafer, then its surface will be tilted and the vertical displacement will vary over the wafer. The resulting errors vary from insufficient contact to excessive bending, which can drive the probes off the probe pad. (This problem can be corrected by using a "smart" vertical drive, in which a sensor detects contact with the wafer surface and adjusts the vertical displacement.)

The probe points on a probe card must be inspected regularly for cleanliness and for correct adjustment. Cards in need of adjustment or repair are returned to the manufacturer for service. A stock of cards must be kept to allow for the cards that are in the service loop.

Although bad probing is an important source of errors, it is not the only source. Other faults include bad cable connections, faulty relays in the matrix, malfunctioning instruments, and poorly designed and fabricated test structures.

2.1.3.2 Detection and Flagging of Bad Data. Some measurement anomalies can be automatically identified during probing, and the data flagged to indicate that they might be erroneous. The reason for caution in assigning "measurement error" to an anomaly is that some test-structure faults, for example, breaks in metal lines or bad ohmic contact between layers, can reveal themselves in the

same way as, for example, bad probe contacts, and additional study (e.g., visual inspection) may be necessary to resolve the ambiguity.

The information available at measurement time for detection of possible measurement errors includes the following:

1. Continuity between pads

   Some test structures are purely resistive in nature, and resistance checks can be made between pads to detect open-circuit conditions. An open circuit can occur either because the probes are not in electrical contact with the pads, or because the structure itself is faulty. Active devices, such as transistors, cannot readily be checked this way.

2. Expected range of data

   Contact resistances, sheet resistances, linewidths, and transistor threshold voltages are examples of quantities that have expected ranges of values for physically valid devices. In many cases, the arithmetic sign of the value suffices as a flag: passive resistances are never negative.

3. Correlation of data values

   Physically anomalous or bad devices usually show recognizable correlations among their parameter values. A shorted transistor, for example, has high leakage current and an unreasonable threshold-voltage value. If one parameter is anomalous and the other is normal, then there may be something wrong with the measurement (or there may not).

Test-structure data are identified with individual chips. The wafer is treated as a rectangular array of data values, indexed by chip row and column. Row 1 is at the bottom of the wafer, and column 1 is at the left (see Figure 2.1.3-1).

It is important to define "right side up" for the wafer, so that all parties agree on the chip numbering--it is the fundamental means of chip identification for data analysis and comparison. In most cases, there are lettering and numbering on the wafer, all reading the same way. The default assumption is that the wafer is right side up when the lettering is right side up. The wafer flat is <u>not</u> an acceptable standard. We have seen variations in the flat position among wafer lots. Identification of row 1 and column 1 has presented no difficulties to date. Numbering starts from the first row and first column of chips that are completely on the wafer. It is to be expected that this definition will prove inadequate because there will be borderline cases. The only error-free system is to identify row 1 and column 1 by fabricating standardized marks on the wafer, such as chip numbers or a "flag chip" that has a defined position.

```
RK2:CZSA13.107    CREATED: 09:08:42 17-SEP-82 BY CMOSDI
SILICON GATE CMOS TEST WAFER
TESTED BY JPL
4 TERMINAL NCH TRANSISTOR; LENGTH = 2 LAMBDA , WIDTH = 10 LAMBDA
INTEGER   COL     ( 138)
     4     5     6     7     8     9    10     3     4     5     6     7     8     9    10    11     2     3
     4     5     6     8     9    10    11    12     1     2     3     4     5     6     7     8     9    10
    11    12    13     1     2     3     4     5     6     7     8     9    10    11    12    13     1     2
     3     4     5     6     7     8     9    10    11    12    13     1     2     4     5     6     8     9
    10    12    13     1     2     3     4     5     6     7     8     9    10    11    12    13     1     2
     3     4     5     6     7     8     9    10    11    12    13     1     2     3     4     5     6     7
     8     9    10    11    12    13     2     3     4     5     6     8     9    10    11    12     3     4
     5     6     7     8     9    10    11     5     6     7     8     9
INTEGER   ROW     ( 138)
     1     1     1     1     1     1     1     2     2     2     2     2     2     2     2     2     3     3
     3     3     3     3     3     3     3     3     4     4     4     4     4     4     4     4     4     4
     4     4     4     5     5     5     5     5     5     5     5     5     5     5     5     5     6     6
     6     6     6     6     6     6     6     6     6     6     6     7     7     7     7     7     7     7
     7     7     7     8     8     8     8     8     8     8     8     8     8     8     8     8     9     9
     9     9     9     9     9     9     9     9     9     9     9    10    10    10    10    10    10    10
    10    10    10    10    10    10    11    11    11    11    11    11    11    11    11    11    12    12
    12    12    12    12    12    12    12    13    13    13    13    13
REAL      VT0     ( 138)
 0.8978671     0.9062845     0.9124520     0.9039939     0.9045509     0.9163378     0.9149143     0.8905141
 0.9085205     0.9121565     0.9015152     0.9086438     0.9123701     0.9244494     0.9240245     0.9164779
 0.8980778     0.9109823     0.9102269     0.9105178     0.9022723     0.9182722     0.9167645     0.9120094
 0.9083617     0.9072062     0.8940262     0.8929908     0.8998032     0.8931395     0.9019577     0.9071634
 0.9173749     0.9148381     0.9151761     0.9185899     0.9289070     0.9249299     0.9050236     0.8862377
 0.8944253     0.8891656     0.9026937     0.9007840     0.9023483     0.9047250     0.8988650     0.9080892
 0.9184324      1.625929     0.9192460     0.1142920     0.8810667     0.8889441     0.8997396     0.9046996
 0.9031140     0.9047860     0.9071295     0.9146895     0.9174542     0.9111638     0.9159162     0.9118801
  1.259466     0.8998263     0.9003201     0.8921217     0.9003214     0.9033647     0.9055185     0.9075007
 0.9052460     0.9128371     0.0000000     0.0000000     0.8959864     0.8977084     0.8906791     0.9027473
 0.8976020     0.9059175     0.8938200     0.9040734     0.9026607     0.9055387     0.4200542     0.0000000
 0.8684149     0.8853486     0.8918452     0.8952990     0.8817666     0.8934152     0.8937036     0.9001067
 0.9022825     0.9047412     0.9034353     0.9001607     0.0000000     0.8080731     0.8535551     0.8834035
 0.8859191     0.8871537     0.8866282     0.8939978     0.8932077     0.8928351     0.8884206     0.8875451
 0.8888655     0.0000000     0.8150555     0.8462253     0.8729465     0.8686230     0.8728316     0.8700008
 0.8745218     0.8738344     0.8782646     0.8790414     0.8135861     0.8229985     0.8605074     0.8391205
 0.8546422     0.8636200     0.8537188     0.8633395     0.8627559     0.8429759     0.7944040     0.8244039
 0.8441314     0.8427730
```

Figure 2.1.3-1. Example of a CRUNCH file illustrating the variation in n-channel transistor measurements. Wafer maps of this data are shown in Figures 2.5.3-3 and 2.5.3-4.

The purpose of data flagging is to segregate test-module data into three classes:

1. ND: No data, i.e., the test module on the chip was not measured.

2. ME: Measurement error detected, for any of the reasons discussed above.

3. VD: Valid data.

"Valid" means "not in error from known instrumental causes." Whether or not the data are "correct," in the sense that the test structure is "good," is a separate issue. The purpose of the ME class is to separate measurement artifacts from physical anomalies.

Prior to a measurement run, every module of every chip belongs to the class ND. As measurements are made, the data are identified by chip row and column, and classified as VD or ME.

Initially, no data flagging was done in the JPL test system. All values were stored. The need for detection of measurement errors became clear when a relatively large number of bad-probe cases were identified. At present, data values are flagged by giving them special values. The data arrays are initialized to zero; no measured value is zero. Therefore, a zero value serves as the ND (no data) flag. Measurement errors are flagged either by reversing the sign of the measurement, or by making the value some number that is unlikely to occur in practice (e.g., 999.999). This is not a satisfactory method, because in many cases the actual value, even though erroneous, contains information. An improved data structure will be used in the future, involving two values for every measurement: (1) the measured value, and (2) a code byte for the classification. This is actually a record structure, and it is handled better in a language like Pascal or `C` than in FORTRAN.

As a general rule, it is preferable to minimize the amount of data analysis that the test system has to do, so that it can perform its primary task of acquiring and storing data efficiently and economically. The principal data-analysis requirement on the JPL test system is that it provide the information needed to classify data values, as described above. Checking continuity between pads is an example of measurement for the sole purpose of error detection; no other use is made of the data. A test system could be made to respond to anomalous data by changing the measurement sequence. In this way, special data could be acquired for diagnostic purposes. We plan to explore this approach. One anticipated problem with this approach is that the situation could get out of hand, requiring complex test programs, and yielding too much information to digest.

The need for probe-contact validation has proved to be so important in practice that test-module designs have evolved to the point of dedicating pads specifically for continuity checks in structures that are not inherently testable for continuity. This results in fewer test-module functions per pad block, but also in more reliable data. A particularly important example is a random fault array that shows a short circuit for oxide pinholes. A good de-

vice is an open circuit, and it cannot be distinguished from a probe-contact failure.

Probe cards are listed in commercial catalogs with double contacts for probe-contact checking on a single pad, but they require a pad size of 100 micrometers or more. They cannot be used with the standard 2 X 10 array. However, even if such points were available for 80-micrometer pads, there would be a reluctance to use them, because the double-pad approach is more straightforward and probably more reliable. Additionally, the use of special probe points would reduce the portability of the entire test methodology.

2.1.4 References

1. Buehler, M. G., "Comprehensive Test Patterns with Modular Test Structures: The 2 by N Probe-Pad Array Approach," Solid-St. Technol., 22, 89-94 (October 1979).

2. CRUNCH is a program developed under tasks sponsored by NASA and the National Bureau of Standards. Currently the only documentation is embedded in the FORTRAN code; additional description is planned for a JPL technical memorandum. Copies of the program can be obtained through participation in the CRUNCH Users' Group. For further information, contact Dr. Frank J. Grunthaner, Mail Stop 198-231, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109.

2.2 CMOS-BULK LAYOUT RULES

The 5-micrometer CMOS-bulk layout rules we developed were used to lay out the test chips and test strips fabricated on the first four CMOS-bulk foundry runs. The rules were published in VLSI Design and a copy is included at the end of this section. Since these rules were published, we have found one change to suggest. That is, that the spacing between the p^+-source/drain doping and the thin oxide/diffusion levels should be increased from 1.5 to 2 lambda. This change reduces the likelihood that a fatal error will occur in the source/drain doping due to either a misalignment of the p^+-source/drain doping levels with respect to the thin oxide/diffusion levels or a size error in one of the layers or both.

A systematic approach was taken to the description of layout rules. The approach is based on the observation that the layout rules describe dimensional restrictions within a particular layer or between two layers. This has led to a description based on primitive figures that describe the geometrical restrictions.

The description is based on the CMOS-bulk process shown in Figure 2.2-1. The minimum set of levels for this process are noted as follows:

T = Thin Oxide/Diffusion

W = Well

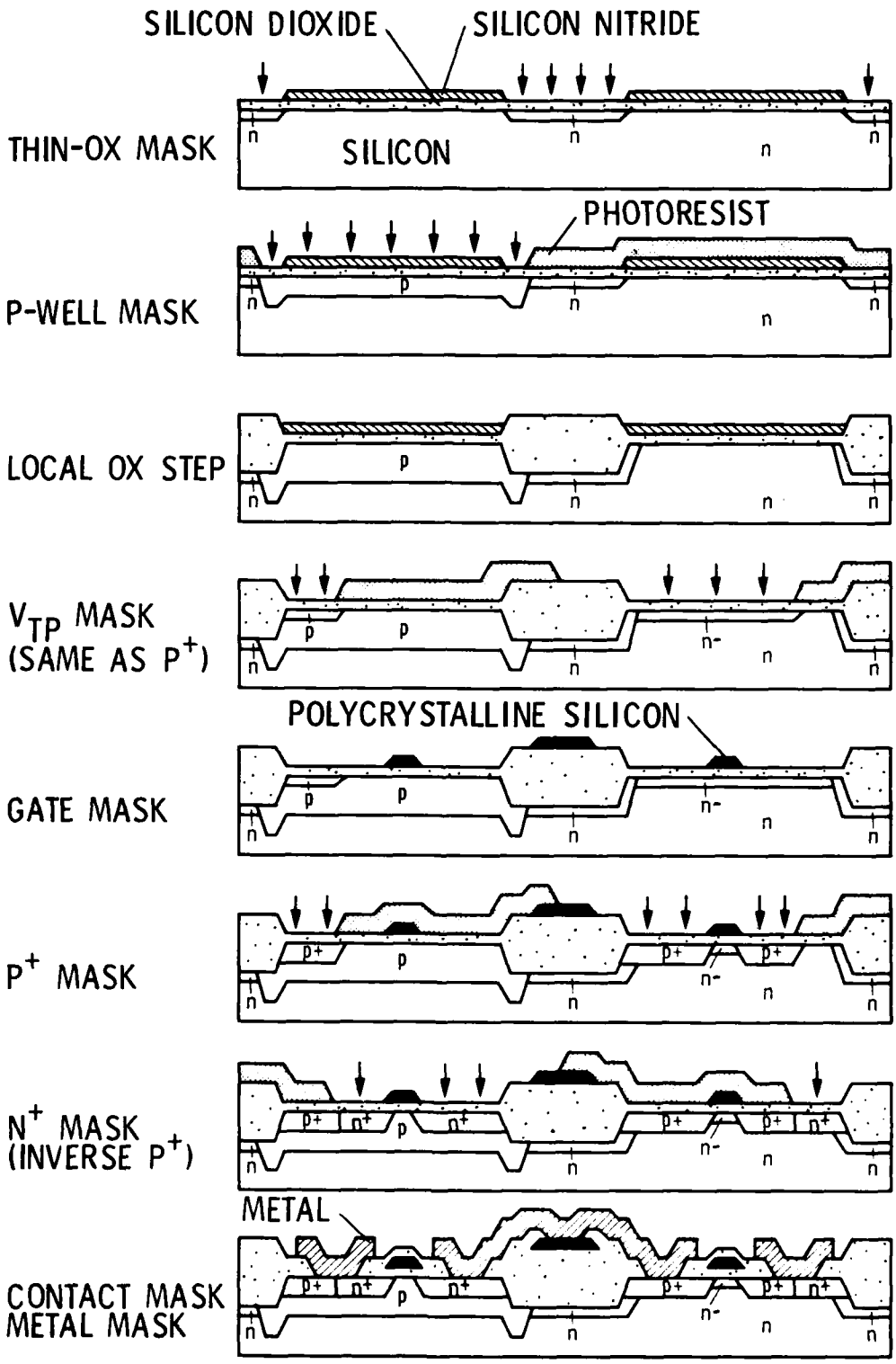


Figure 2.2-1. Cross-sectional drawing for a CMOS-bulk fabrication process.

G = Gate

P = p^+ Source/Drain Doping

C = Contact

M = Metal

The symbol G refers to all polycrystalline material used to form both gates and wires. The other levels (e.g., V_{TP}, N^+) are derived from the above set. The passivation level has been omitted from this discussion for simplicity.

The layout rules are described by six geometrical primitives as illustrated in Figure 2.2-2 and defined as follows:

W_{ii} = width of the inside of level i from inside of level i.

X_{ji} = extension of the inside of level j from outside of level i.

Y_{ji} = extension of the inside of level j from inside of level i.

S_{ii} = spacing of the outside of level i from outside of level i.

S_{ji} = spacing of the outside of level j from outside of level i.

R_{ji} = spacing of the outside of level j from inside of level i.

These primitives are illustrated in Figure 2.2-2. For the CMOS process described in Figure 2.2-1, the primitive Y is invoked only once, and the primitive R is not invoked.

The layout rules are summarized in Figure 2.2-3. This figure contains six figures but only five primitives are used. The extra figure occurs for the extension X. Note that all but four of the 28 rules involve one or two levels. The exceptions are: S_{gct}, which is the spacing of gate contact from the thin oxide; S_{tcg}, which is the spacing of the thin-oxide contact from the gate; and X_{tcp} and Y_{tcp}, which are extensions of the thin-oxide contact from the P level. These rules are also shown in Figure 2.2-4 in order to relate the primitive figures to a more conventional description.

The rules are arranged in a matrix as shown in Figure 2.2-5. There are 28 rules listed in Figure 2.2-3. Since $S_{ij} = S_{ji}$, there are an additional six entries to make a total of 34 entries in the matrix shown in Figure 2.2-5.

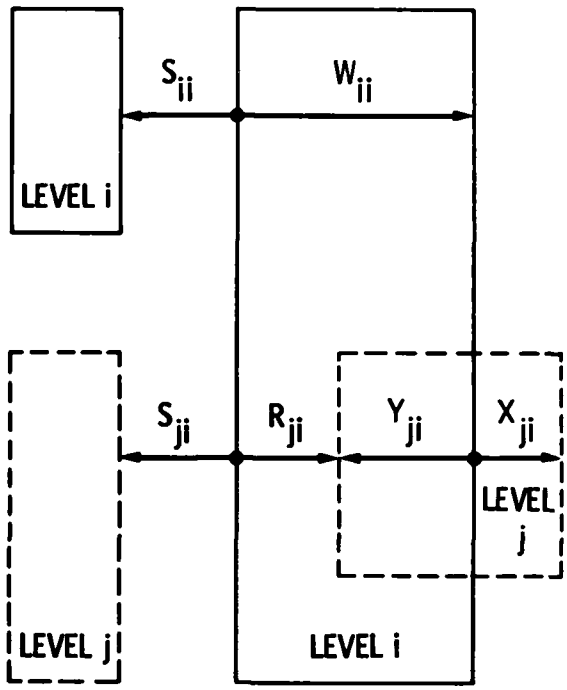
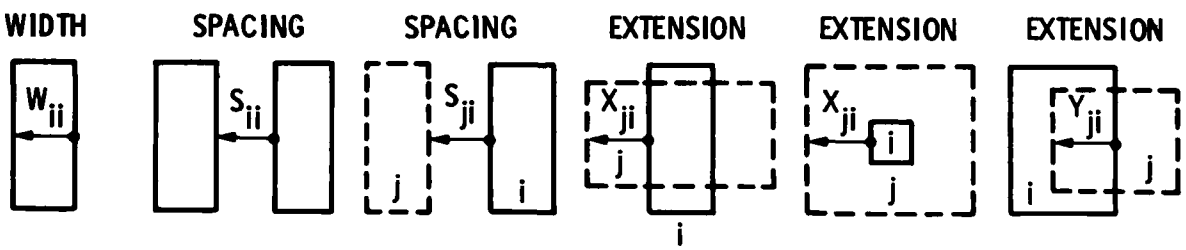


Figure 2.2-2. Six geometrical layout rule primitives and their relationship to photomask levels i and j.



$W_{ww} = 4$	$S_{ww} = 2$	$S_{tw} = 5$	$X_{gt} = 2$	$X_{tc} = 1$	$Y_{tcp} = 3$
$W_{tt} = 2$	$S_{tt} = 3$	$S_{gt} = 1$	$X_{tg} = 2$	$X_{gc} = 1$	
$W_{gg} = 2$	$S_{gg} = 2$	$S_{pt} = 2$	$X_{pg} = 2$	$X_{mc} = 1$	
$W_{pp} = 2$	$S_{pp} = 2$	$S_{gct} = 2$	$X_{tcp} = 3$	$X_{wt} = 3$	
$W_{cc} = 2$	$S_{cc} = 2$	$S_{pg} = 2$		$X_{pt} = 2$	
$W_{mm} = 3$	$S_{mm} = 3$	$S_{tcg} = 2$			

Figure 2.2-3. The 28 CMOS-bulk layout rules displayed using geometrical primitive figures. Each rule is expressed in multiples of lambda where lambda is 2.5 micrometers.

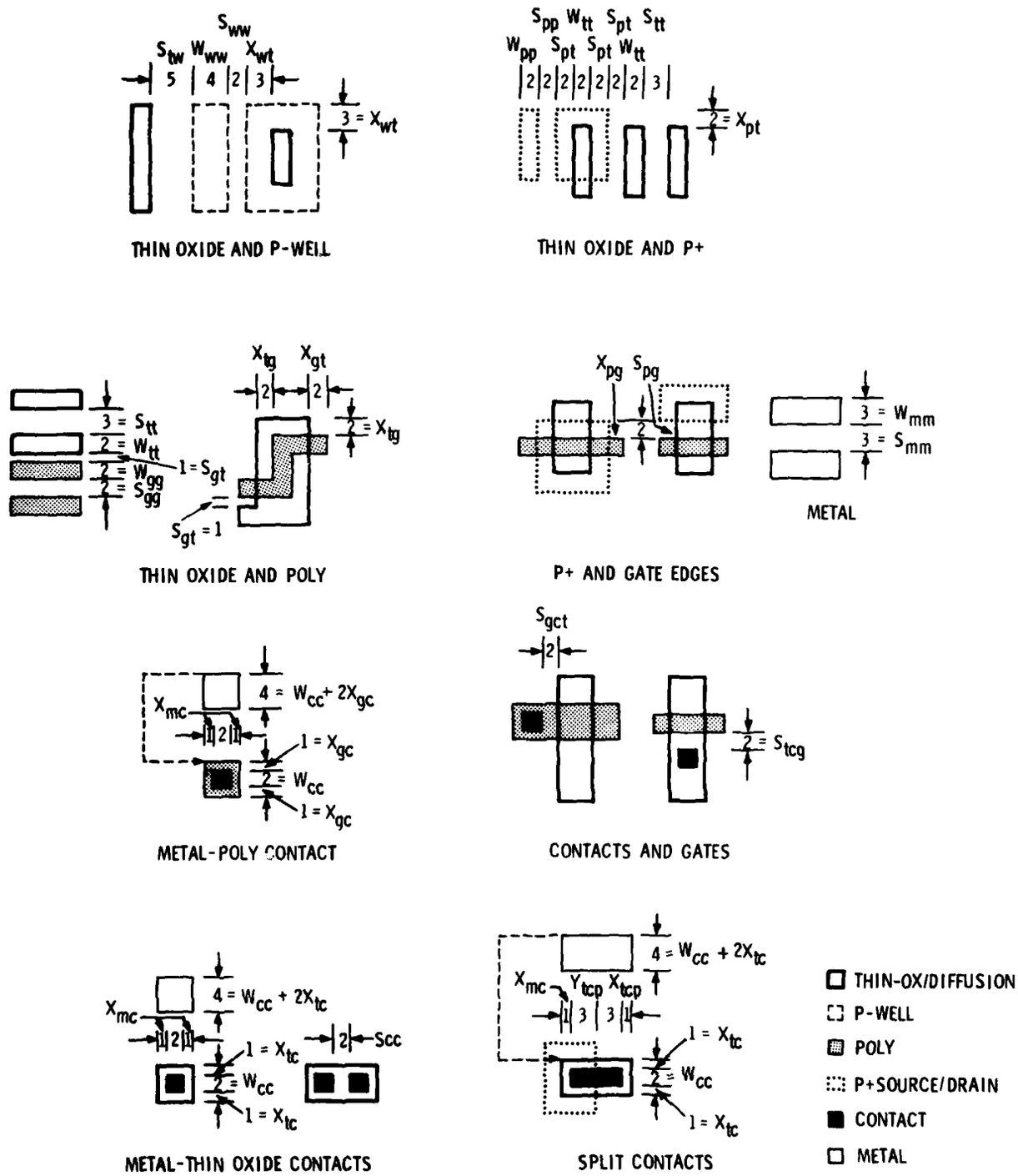


Figure 2.2-4. CMOS-bulk layout rules illustrated using descriptive diagrams. Each rule is expressed in multiples of lambda, where lambda is 2.5 micrometers.

REFERENCE LEVEL (#1)

		xW	xT	xG	xP	xC	xM
WELL	Wy	W S	X S				
THIN-OX	Ty	S	W S	X S	S	X S	
GATE	Gy		X S	W S	S	X S	
P	Py		X S	X S	W S		
CONTACT	Cy		S	S	Y X	W S	
METAL	My					X	W S

OFF DIAGONAL

Y	X
R	S

DIAGONAL

W	
	S

Figure 2.2-5. Matrix display of CMOS-bulk layout rules shown in Figures 2.2-3 and 2.2-4.

The matrix shown in Figure 2.2-5 allows one to identify which rules are not specified. One can then question whether more rules are required when designing special circuits for highly reliable applications. The matrix shown in Figure 2.2-5 also allows one to compare one rule set with another. One way this can be accomplished is to calculate a complexity factor. For instance, the total possible entries in the matrix is 6 X 6 X 4 - 6 X 2 = 132. This total must be reduced by the 15 off-diagonal spacing rules to arrive at the total possible number of rules, which is 117. The complexity factor is defined as:

$$\frac{\text{Number of layout rules}}{\text{Number of possible layout rules}} = \frac{28}{117} = 0.24.$$

Portable design rules for bulk CMOS are given in the following paper.

Portable Design Rules for Bulk CMOS

Thomas W. Griswold, Caltech Jet Propulsion Laboratory

For the past several years, one school of IC designers has used a simplified set of nMOS geometrical design rules (GDR) that is "portable," in that it can be used by many different nMOS manufacturers (Conway *et al.* 1980, Lyon 1981). The development of nMOS design technology was stimulated by the availability of fabrication services. Sources for oxide-isolated silicon-gate bulk CMOS* are now becoming available, and a similar set of GDR for bulk CMOS is of interest. CMOS is more complex and less dense than nMOS, but it has certain advantages in power dissipation, noise immunity and supply voltage range (see the accompanying box). This article describes a preliminary set of design rules for bulk CMOS that has been verified for simple test structures (transistors, inverters, ring oscillators) by fabrication at five-micron feature size.

The GDR are defined in terms of Caltech Intermediate Form (CIF) (Mead and Conway 1980): a geometry-description language that defines simple geometrical objects in layers, which are abstractions of physical mask layers. The minimum set of layers is chosen that provides at least one way to make the required structures (conductors, contacts, crossovers, transistors). In general, there are more mask layers than CIF layers. The additional mask layers are created by logical operations (AND, OR, NOT, XOR) on the CIF layers.

The CIF design-file defines the chip layout as-fabricated: that is, as the designer expects to see it. Process-dependent factors such as resist polarity and etch undercutting are taken into account in the conversion from CIF to masks. Since fabrication processes differ, portability is lost in this conversion. In general, masks are useful only in the process for which they were created.

The design rules do not presume the existence of any particular design methodology. A chip designed by any method that results in a legal CIF file can be fabricated.

p-WELL and n-WELL CMOS Processes

CMOS uses both n- and p-channel transistors. Therefore, both p- and n-type substrate material must be present on the chip. In bulk CMOS, this is achieved by deep diffusions in the starting silicon material to form wells. In a p-well process (Figure 1), the starting material is n-type, and the wells are p-type diffusions. (The types are reversed for an n-well process.) The relative physical and electrical merits of the two processes are not an issue here—they are equivalent as far as portable design rules are concerned. Early versions of the design rules were independent of the well type, but difficulties with uniform and symmetrical definition of rules for the well forced a separation for the sake of practicability. The p-well process was selected because of its wider availability. A separate but similar set of rules will be needed for n-well processes.

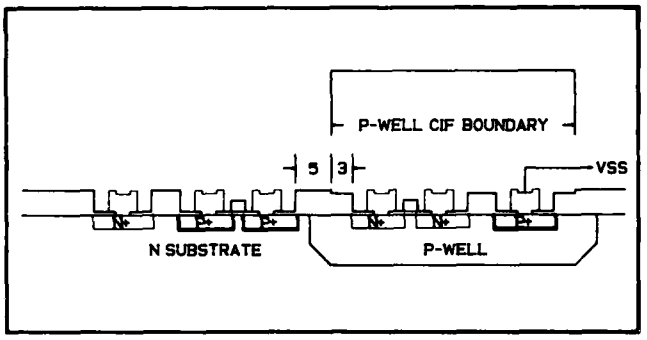


FIGURE 1. Cross-section of p-well bulk CMOS.

Layer	Code	Color	Comments
1. p-well	CW	Brown	Inside brown is p-type, outside is n-type substrate. Each p-well must be connected to V_{SS}.
2. Thin oxide	CD	Green	Thin oxide may not cross a well boundary.
3. Polysilicon	CP	Red	Single poly type (usually n+).
4. p+	CS	Orange	Inside orange is p+, outside is n+.
5. Contact	CC	Black	Contacts that are ohmic to the substrate are marked with a cross in the cut.
6. Metal	CM	Blue	
7. Glass	CG	—	Not shown. Typical spacing from glass cut to pad edge is 4 lambda.

TABLE 1. CMOS CIF layers. Colors are conventions, not standards.

Bulk CMOS and CMOS-SOS

Another CMOS technology that is technically suitable for foundry fabrication is CMOS-SOS. The design rules for CMOS-SOS are simpler than the CMOS/bulk rules, because SOS has no well. For the same reason, SOS should be denser than bulk, and easier to lay out. SOS is free of the latch-up phenomenon (Lipman 1982). SOS has a number of technical problems associated with its "exotic" nature, but its major disadvantage is the lack of commercial fabrication sources.

*The term "bulk" CMOS refers to those processes that use the starting wafer material (silicon) as the substrate. The term is used to distinguish this type of process from CMOS-SOS, in which the silicon material is a thin film on an insulating substrate.

Copyright 1982. Reprinted from VLSI DESIGN Magazine.
P.O. Box 50518, Palo Alto, CA 94303-0518. (415) 966-8340.

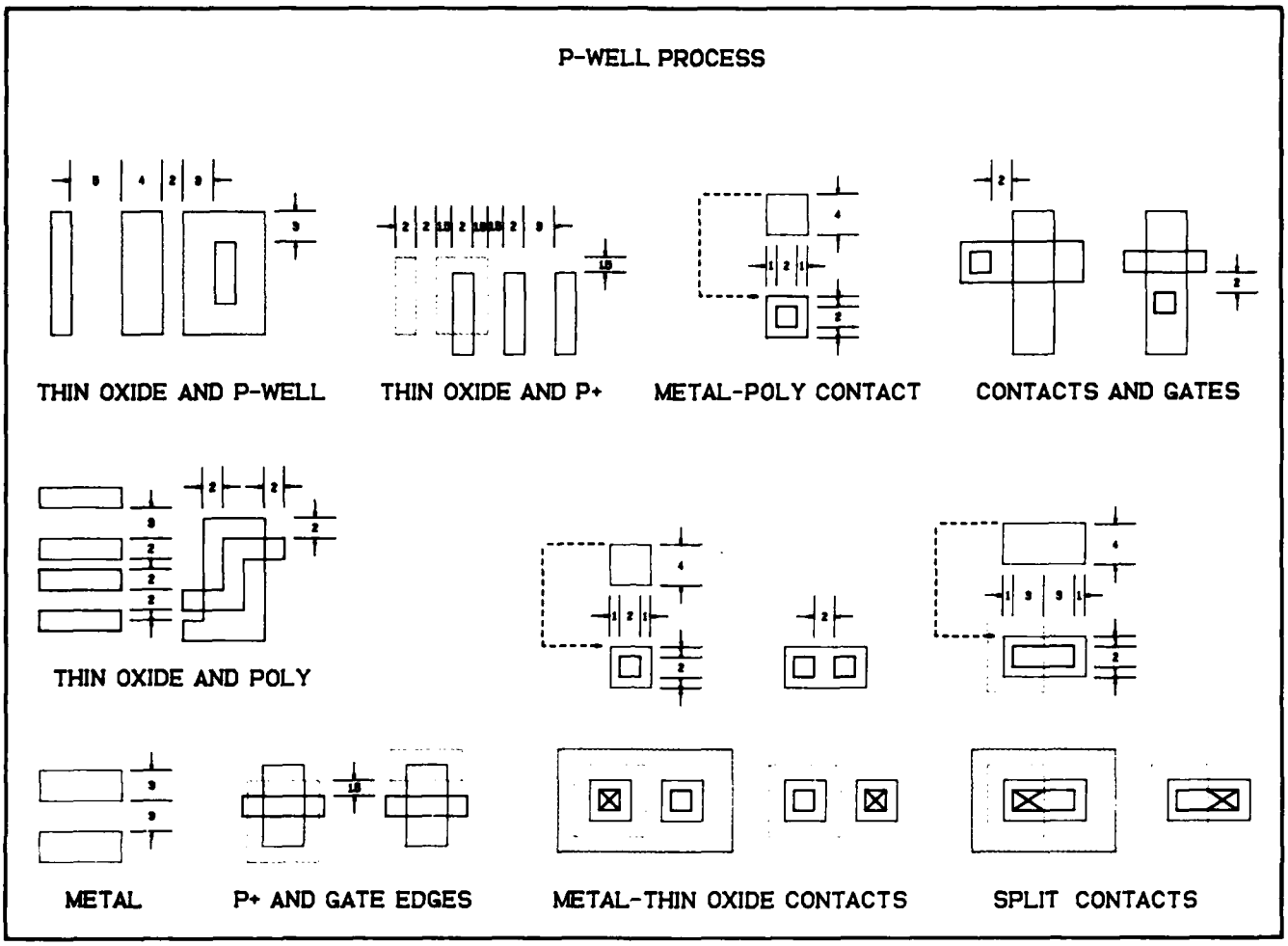


FIGURE 2. Bulk CMOS geometrical design rules (p-well process).

One reason for this lack is that starting SOS wafers are about ten times more expensive than bulk silicon wafers, and SOS parts cannot compete in high-volume low-cost markets. In the foundry environment, however, design cost, not chip cost, is the driving factor, and CMOS-SOS might well be a successful competitor if fabrication sources can be developed for it.

CMOS Geometrical Rules

Figure 2 shows the geometrical rules. Table 1 defines the CIF layers and color-code conventions. Figure 3 shows a sample NAND-invert layout. Dimensions are given in units of the scaling parameter *lambda*, which is equal to half the minimum feature size. Preliminary verification has been made at five-micron feature size (lambda = 2.5 microns). The p-well lateral diffusion, which determines the outside well clearance, is expected to be the most critical dimension for scaling.

Diffusions are n+ outside the p+ regions and p+ inside them. A transistor is p-channel if it is inside a p+ region, otherwise it is n-channel. N-channel transistors must be inside p-wells, and p-channel transistors must be outside.

The p-well is a deep diffusion that forms the substrate for n-channel transistors (Figure 1). The 5-lambda outside dimension provides clearance between the p-well edge and adjacent p+ diffusions. The 3-lambda inside clearance is a conservative estimate of the spacing needed for the transition of the field

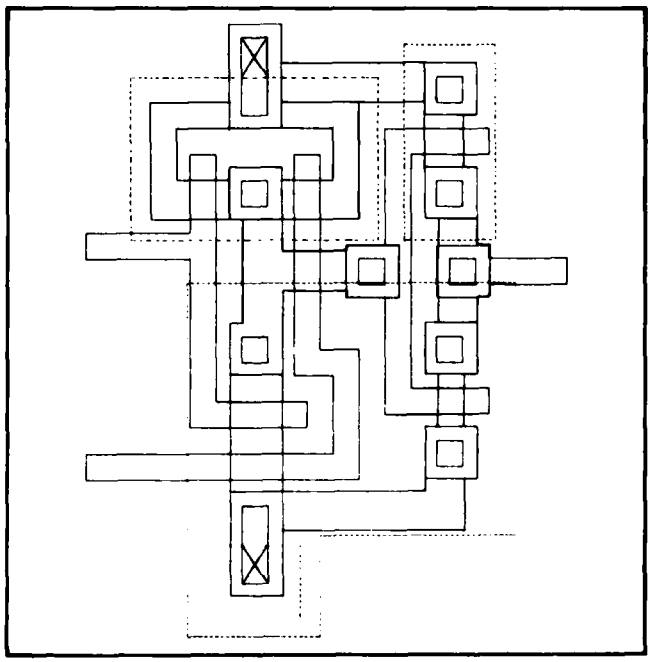


FIGURE 3. One possible layout of a CMOS NAND-plus-inverter module.

oxide across the well boundary. Some proprietary rules allow the placement of p+ diffusions at the well boundary (*i.e.*, zero inside clearance for ohmic contacts to the well). While this increases layout density, it cannot be supported until it is shown to be portable.

Thin oxide may not cross a well boundary, because a shorted condition will result. The n- and p-channel drains of inverters and gates are separated by 8 lambda (the sum of the inside and outside well clearances). They are connected by a metal strap to form the output node. The large well clearances tend to make bulk CMOS a non-dense technology. It may be possible to reduce them in future revisions.

The p-well is actually the anode of a diffused diode, and it must be grounded explicitly by contacts to the V_{SS} bus. The sheet resistance can be as high as several thousand ohms per square, and the p-well should be thoroughly grounded to minimize voltage drops from substrate currents. No quantitative rule is available; A rule of thumb is suggested: If there is space, put in a contact; if there isn't space, consider making some. Because the p-well is part of a diode, its boundary must be completely within the active area of the chip. There is no specific value for the clearance; the designer should keep well boundaries far from the chip edge.

P-wells not tied to V_{SS} are not supported. One of the reasons is prevention of latch-up (Lipman 1982). Another is that removal of this restriction would require the definition of additional rules for non-related well clearance and p+ collars, which would complicate the GDR set and reduce its portability.

The 1.5-lambda spacings for overlap and clearance of p+ and thin oxide were adopted from the nMOS rules. They have been criticized for violating the "meta rules" (Lyon 1981), and also for complicating the task of CIF-to-mask conversion with half-lambda values. They may be increased to two lambda in future revisions.

The two diffusion types and two types of substrates in CMOS result in four distinct metal-diffusion contacts. Two form rectifying junctions to the substrate, and two form ohmic contacts to the substrate. The ohmic combinations are marked with crosses in Figure 2.

The recommended maximum-size contact cut is two by six lambda. Large-area contacts should be made by multiple use of small contacts.

The split contact is used to tie transistor sources to substrates. It is equivalent to separate contacts strapped with metal, but it is smaller. Discussions with industrial sources resulted in an increase from the original 4-lambda value of the cut length to its current 6-lambda value. Although a 4-lambda cut will probably work, and a 5-lambda cut will *very* probably work, the conservative value was chosen for the preliminary set of rules. The physical reason is that maintenance of the p+/n+ doping boundary within one lambda of the center of the cut is not guaranteed for all processes, and the result of missing the cut is a rectifying contact. This is not necessarily fatal (at least for small transistors), but it certainly is not desirable.

Some of the rules have no logical or electrical significance, existing strictly for lithographic control in fabrication.

Polysilicon is typically doped n+ in current fabrication processes. Its sheet resistance in p+ areas is increased by doping compensation. This may affect the quality of metal-polysilicon contacts inside p+ regions, and force the addition of a rule that restricts them to n+ regions. Such a rule would complicate the design procedure.

Buried contacts have not been included in the preliminary GDR set. Their addition will interact with polysilicon doping. Two possibilities exist: (1) Stay with n+ polysilicon and use only n-type buried contacts; (2) Define both n+ and p+ poly, and have both n+ and p+ buried contacts. The second approach requires the addition of a contact between n+ and p+ poly. Because most (if not all) commercial processes have only n+ polysilicon, the second option would require the establishment of a new set of silicon foundry processes.

The 1.5-lambda minimum spacing of a doping change from a gate is found in both bulk and CMOS-SOS rules. It supports the creation of diodes, for whatever purpose the designer might have in mind. The transition from n+ to p+ doping is not controlled, in the sense that different processes can have different junction gradients, and thus different avalanche voltages. Another way to make a diode is to make the doping change in the middle of a gate, with at least 2-lambda clearance from either gate edge. This construction is not part of the current GDR set; it may or may not be added. It has been used successfully in CMOS-SOS test modules to create four-terminal transistors (the substrate in CMOS-SOS normally floats). Some variation of these structures might be used in standard libraries for input-protection devices.

Some of the rules have no logical or electrical significance, existing strictly for lithographic control in fabrication. These rules are the 2-lambda minimum spacings between p+ edges and between p-well edges. The general rule is that no element of photoresist (or electron resist) on a wafer should be less than 2 lambda in width. The reason is that a thin strip of resist could peel from the surface and lead to extensive damage in pattern definition. If two p+ or well edges are close together, then merge them. The minimum width of the p-well is somewhat similar—it may be difficult to fabricate such a structure.

Interest in the use of these rules for linear circuits has prompted a search for an electrode layer that will provide voltage-independent capacitors. The new layer could be either diffusion under polysilicon, or a second layer of polysilicon. Other layers that might be added include a second layer of metal and its via cut, and (for high-reliability applications) an explicit n+ doping layer to minimize the amount of dopant in the field oxide.

Acknowledgments

This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, and was supported jointly by the National Aeronautics and Space Administration and the Defense Advanced Research Projects Agency. Two sets of CMOS rules were consulted in constructing the JPL preliminary set. The first set was communicated to JPL by Hewlett-Packard (Lipman 1981), and the second by the University of California at Berkeley (Carlo Séquin).

JPL welcomes any comments regarding these preliminary design rules—either positive or negative. The author can be contacted at (213) 354-4145 or the following address:

Jet Propulsion Laboratory
Mail Stop 198-231
4800 Oak Grove Drive
Pasadena, CA 91109

References

Conway, L.A., A. Bell, M.E. Newell. Second Quarter 1980. "MPC79: A Large-scale Demonstration of a New Way to Create Systems in Silicon," *LAMBDA* (now *VLSI DESIGN*).

Lipman, J. Fourth Quarter 1981. "A CMOS Implementation of an Introductory VLSI Design Course," *VLSI DESIGN*.

Lipman, J. May/June 1982. "Latchup Prevention in Bulk p-Well CMOS Circuits," *VLSI DESIGN*.

Lyon, R.F. First Quarter 1981. "Simplified Design Rules for VLSI Layouts," *LAMBDA* (now *VLSI DESIGN*).

Mead, C.A. and L.A. Conway. 1980. *Introduction to VLSI Systems*. Reading: Addison Wesley.

About the Author

Tom Griswold received a Ph.D. degree in solid-state physics from the University of California at Berkeley in 1953. He worked in the semiconductor industry from 1953 to 1975, mainly in the development of technology for discrete silicon devices, at Hughes Aircraft Company, Continental Device Corporation, Teledyne Semiconductor, and Semtech Corporation. Since 1975, he has been with Caltech Jet Propulsion Laboratory, working on the development of test methodology and standards for high-reliability integrated circuits.

2.3 TEST CHIPS

2.3.1 Test-Chip Assembler

2.3.1.1 Introduction

In the manufacture of custom semiconductor LSI and VLSI devices, a method is needed to provide assurance that the manufacturer has met his contractual obligations and similarly that design errors on the part of the designer do not get blamed on the manufacturer's process [1]. Also, if the designer is interested in high-reliability applications, as would be the case for NASA spacecraft, the assessment of the device reliability becomes one of extreme importance [2]. One possible method of meeting both of these requirements is through the more extensive use of test structures and test chips designed to provide the required information. The design of these structures is naturally very close to the geometry level of description, and hence highly sophisticated methods of computer-aided design (CAD) may not be suitable. The process, however, can be very labor-intensive unless special techniques are used. This section describes a test structure and test-chip assembler (TCA) used to generate geometrical descriptions of these structures in the Caltech Intermediate Form (CIF) [3], which requires only access to a minicomputer such as a DEC VAX with a Pascal compiler.

2.3.1.2 Historical Background

The TCA is based upon an early JPL design tool named ART. In a similar fashion to that used by the TCA, ART was a collection of Pascal procedures used to generate a CIF file. ART, however, suffered from the following shortcomings:

1. It was too closely tied to the geometry of the structures, all of which were generated by specifying "boxes" and "wires." This led to long design times because of the detail required of the designer.

2. It was difficult to generate designs with varying values of the scale factor lambda. A change in the value of lambda required a complete regeneration of the structure since connecting wire lengths would also scale.

3. Probe-pad dimensions and spacing did not remain constant with different values of the scale factor lambda.

4. It was difficult to convert designs from one technology to another without extensive regeneration of the structures.

Despite these deficiencies, ART was nevertheless a valuable design tool and laid the groundwork for the present design system. Using this design tool, several test chips were generated for NMOS (JPL MPC880, DRPIN11) and for CMOS-SOS (SOSONE, SOSTWO, and SOSTRP). The difference in labor required to

generate comparative designs, however, is significant. For example, DRPIN11 took approximately two months to lay out and code. The latest NMOS test chip, NM3034, took approximately two days, including the generation of check plots.

2.3.1.3 Purpose

The purpose of the test-chip assembler is fivefold:

1. To generate design files of test structures and test vehicles in the CIF.

2. To generate human-readable descriptions of these structures and chips at run time.

3. To generate a file containing the geometrical coordinates of the individual structures in the chip with respect to some predetermined origin. This file is used to provide the structure location information necessary for the wafer probe test.

4. To provide nonscaling features and wiring for individual structures. In the interests of accurate testing, it is very important that some dimensions such as those of the pad spacing do not scale with a change in the scale factor lambda.

5. To minimize the routine computations required of the designer.

2.3.1.4 Description

The TCA consists of a collection of Pascal [4] procedures, which provide a convenient method of generating a description of a design file in the CIF. Since CIF-type primitives such as `box,` `layer,` etc. are embedded in Pascal, the use of parameterized procedures and control structures is made available to the designer. The TCA is specifically designed to generate test structures and chips subject to certain restrictions which will be outlined below. In its present form it is not a silicon compiler [5] nor a general VLSI circuit design tool. Structure dimensions are expressed in terms of a scaling parameter lambda as described in [3].

Since the primary function of the TCA is to generate test structures, it cannot be too far removed from the geometry level. This sets some bounds on the degree of abstractness permitted. The TCA`s most important feature is the methodology employed and not necessarily the code used to generate the structures themselves. These features can be summarized as follows:

1. The generation of the test structures with a commonly available language, Pascal, using a structured, procedural approach.

2. The generation of the structure descriptions and structure coordinates in the chip at run time, at the same time as the CIF files themselves are being generated.

3. The generation of nonscaling features such as probe pads and pad wiring at the same time as the scalable structures of the test chip.

4. Design speed. A 6 X 6 millimeter test chip can be generated in one day.

5. Since the procedures that generate the test structures are parameterized with respect to the geometrical design rules as well as with the structure geometrical configuration, the structures are independent of the particular set of geometrical design rules.

As discussed earlier, the TCA consists of a collection of parameterized Pascal procedures with embedded CIF primitives. The program consists of a Pascal program shell with three to five main modules plus a number of auxiliary modules which depend on the specific design being generated. These modules will be briefly described below:

1. Mos.xxxgdr : Contains the geometrical design rules for the technology being used. `xxx` refers to the set of rules being used; for example, cm gdr contains the JPL geometrical design rules for CMOS-bulk [8] (as published in VLSI Design), isicm gdr contains the ISI CM(OS-bulk) 3-micrometer design rules, and nm gdr contains the NM(OS) design rules from [3].

2. MosGen.pas : Contains gen(eral) design and housekeeping procedures such as "boxes," "pads," etc.

3. MosPar.pas : Contains procedures to generate par(ametric) structures such as transistors, resistors, cross-bridges, etc.

4. MosRan.pas : Contains procedures used to generate ran(dom) fault structures.

5. ModelP.pas : Contains procedures used to generate structures used to obtain data for model(ing) or simulation purposes.

6. In addition to the first five described above, other procedures may be generated for the evaluation of special structures.

Figure 2.3.1-1, which depicts the first part of a design program for a test chip, shows the method of implementing the program. The separation of the program into separate modules permits the modification of individual modules without affecting the main program. Additional structures can also be generated and evaluated without having to modify existing modules that have already been evaluated. The scale factor lambda is specified in the main program shell before any of the design procedures are invoked. This feature makes it possible to generate test chips with different values of lambda but otherwise identical structures by merely changing the value of one number. The placement of a structure in a block is accomplished by the pdraw command.

Since the main purpose of the TCA is the generation of test structures and chips, certain restrictions have been placed on the assembler:

```
program cm3021(input,output,cif,plan,list);
 %include '[pina.mos]isicmgdr.pas'
 %include '[pina.mos]mosgen.pas'
 %include '[pina.mos]mospar.pas'
 %include '[pina.mos]modelp.pas'
 %include '[pina.expdoc]expresp.pas'

 begin
  lambda := 1.5;
      aa := 150;
      bb :=  10;
  startdesign;

  Block(1,1,50);
   inverterheader;
   inverter(30,2,2,3,2,iwire,0.8,0.8,0.8,0.8,0.8,6.0,0,2);
   inverter(35,2,2,3,2,iwire,1.0,1.0,1.0,1.0,1.0,6.0,0,6);
   inverter(40,2,2,3,2,iwire,1.2,1.2,1.2,1.2,1.2,6.0,0,10);
   inverter(45,2,2,3,2,iwire,1.4,1.4,1.4,1.4,1.4,6.0,0,14);
  define(50);
   pdraw(30,2);
   pdraw(35,6);
   pdraw(40,10);
   pdraw(45,14);
   subcont(592*s,48*s,6*s,4*s,19);
   twobynpad90(0,0,10);
  enddef;

Block(1,2,75);
 inverterheader;
 inverter(55,2,2,3,2,iwire,2.0,2.0,2.0,2.0,2.0,4.0,0,2);
 inverter(60,2,2,3,2,iwire,2.0,2.0,2.0,2.0,2.0,4.5,0,6);
 inverter(65,2,2,3,2,iwire,2.0,2.0,2.0,2.0,2.0,5.0,0,10);
 inverter(70,2,2,3,2,iwire,2.0,2.0,2.0,2.0,2.0,5.5,0,14);
define(75);
 pdraw(55,2);
 pdraw(60,6);
 pdraw(65,10);
 pdraw(70,14);
 subcont(592*s,48*s,6*s,4*s,19);
 twobynpad90(0,0,10);
enddef;
```

Figure 2.3.1-1. Test structure location identified with a block of structures.

1. To promote the standardization of test structure methodology, the 2 X N probe pad array [9] was chosen as the standard pad configuration for all JPL test chip designs. Therefore, all of the structures designed must be accessible for probing by means of a 2 X 10 or 2 X N pad array. This also includes the random fault arrays.

2. Pad dimensions and spacing must not scale with lambda.

3. The assembler must generate a structure description and structure location file at the time the CIF files are generated. Figure 2.3.1-2 shows a portion of such a file.

These conditions have resulted in test chips laid out in very regular arrangements, as a collection of blocks and columns. Each block is defined as consisting of a 2 X N pad block array and its associated structures. In order to minimize computations, the designer also has the option of specifying the geometrical location of the structures in the test chips as points in a grid with nodes 80 micrometers apart (one-half of the probe pad spacing). This last feature reduces the computation of the structure locations to that of counting with integers from 1 to 20. A similar provision has been made for locating the separate blocks within the test chip, by simply specifying the block location as the indexes of a 2 X 2 matrix. The placement of a block in a test chip is accomplished by the cdraw command. Examples are shown in Figure 2.3.1-3.

BLOCK (1,1) ; Coordinates llx,lly: (0.0 , 0.0)

CIF Symbol	N-CH W(u)	N-CH L(u)	P-CH W(u)	P-CH L(u)	Pad Numbers for: VDD	VSS	VIN	VOUT	Contact Sizes(u): cs1	cs2	cs3	cs4	cs5	cs6
			INVERTERS											
30	5.0	5.0	7.5	5.0	2	3	4	1	2.0	2.0	2.0	2.0	2.0	15.0
35	5.0	5.0	7.5	5.0	6	7	8	5	2.5	2.5	2.5	2.5	2.5	15.0
40	5.0	5.0	7.5	5.0	10	11	12	9	3.0	3.0	3.0	3.0	3.0	15.0
45	5.0	5.0	7.5	5.0	14	15	16	13	3.5	3.5	3.5	3.5	3.5	15.0

BLOCK (1,2) ; Coordinates llx,lly: (0.0 , 270.0)

CIF Symbol	N-CH W(u)	N-CH L(u)	P-CH W(u)	P-CH L(u)	Pad Numbers for: VDD	VSS	VIN	VOUT	Contact Sizes(u): cs1	cs2	cs3	cs4	cs5	cs6
			INVERTERS											
55	5.0	5.0	7.5	5.0	2	3	4	1	5.0	5.0	5.0	5.0	5.0	10.0
60	5.0	5.0	7.5	5.0	6	7	8	5	5.0	5.0	5.0	5.0	5.0	11.3
65	5.0	5.0	7.5	5.0	10	11	12	9	5.0	5.0	5.0	5.0	5.0	12.5
70	5.0	5.0	7.5	5.0	14	15	16	13	5.0	5.0	5.0	5.0	5.0	13.8

Figure 2.3.1-2. Test structure documentation including structure location, parameters, and pad identification.

```
define(1100);
 cdraw(50,1,1);
 cdraw(75,1,2);
 cdraw(100,1,3);
 cdraw(125,1,4);
 cdraw(150,1,5);
 cdraw(175,1,6);
 cdraw(200,1,7);
 cdraw(225,1,8);
 cdraw(250,1,9);
 cdraw(275,1,10);
 cdraw(300,1,11);
 cdraw(325,1,12);
 cdraw(865,1,13);
 cdraw(805,1,14);
 cdraw(810,1,15);
 cdraw(815,1,16);
 cdraw(820,1,17);
 cdraw(825,1,18);
enddef;
```

Figure 2.3.1-3. Block locations within a test chip.

Since our design system resides in a DEC VAX 11/780, the separate modules are combined together using the `%include` option of VAX VMS Pascal. Other computer implementations have similar facilities since Pascal`s I/O features are rather primitive.

2.3.1.5 Language

Pascal was chosen as the language for the TCA for a number of reasons:

1. Pascal`s richness of control and data structures [7] are far superior to the other language, FORTRAN, that was available in the JPL VAX at the time the initial work was started. It offers a wide range of iterative and control statements, as well as a wide variety of data types, including user-defined types.

2. `C` was so dense (to quote Kernighan and Ritchie [10], "`C` features economy of expression") that programs could be so compact that they would be hard to maintain. In addition, the lack of a good `C` compiler in the JPL VAX at the time eliminated it from further consideration.

3. `SIMULA,` a language that was being used at Caltech for some of its design tools, required a DEC-20 system to run.

4. `Mainsail` was not available at the time.

5. ICL, a language developed by R. Ayres of Caltech and used in some of the Caltech design tools of the time, required a DEC-20.

2.3.1.6 References

1. Buehler, M. G., and L. W. Linholm, "Toward a Standard Test Chip Methodology for Reliable, Custom Integrated Circuits," Proceedings of the 1981 Custom Integrated Circuits Conference, 142-146 (1981).

2. Scott, W. R., L. M. Hess, and F. R. Ault, "Concerns in Space Flight Qualification of LSI/VLSI Devices," Proceedings of the 1980 IEEE Test Conference, 494-495 (1980).

3. Mead, C. A., and L. A. Conway, Introduction to VLSI Systems, Reading, MA: Addison-Wesley (1980).

4. Jensen, K., and N. Wirth, Pascal-User Manual and Report, New York: Springer-Verlag (1974).

5. Johannsen, D., "Bristle Blocks: A Silicon Compiler," Proceedings of the Caltech Conference on VLSI, 303-310 (January 1979).

6. Grogono, P., Programming in Pascal, Reading, MA: Addison-Wesley (1978).

7. Schneider, G. M., and S. C. Bruell, Advanced Programming and Problem Solving with Pascal, New York: J. Wiley & Sons (1981).

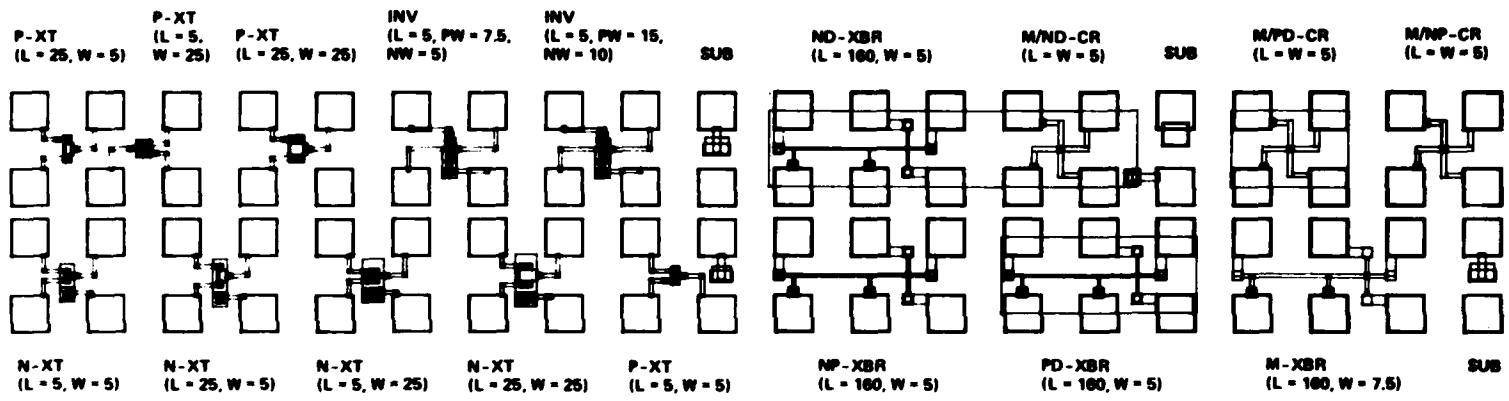
8. Griswold, T. W., "Portable Design Rules for Bulk CMOS," VLSI Design, 3, 62-67 (September/October 1982).

9. Buehler, M. G., "Comprehensive Test Patterns with Modular Test Structures: The 2 by N Probe-Pad Array Approach," Solid-St. Technol., 22, 89-94 (October 1979).

10. Kernighan, B. W., and D. M. Ritchie, The C Programming Language, Englewood Cliffs, NJ: Prentice-Hall (1978).

2.3.2 Test Chips and Test Strips

During the contractual period, five test chips and two test strips were developed for a 5-micrometer, p-well, CMOS-bulk process. The strip and chips for the first two foundry runs are shown in Figures 2.3.2-1 to 2.3.2-3, where the feature size is 2.5 lambda. Before these strips and chips were evaluated, a second set was prepared. These are shown in Figures 2.3.2-4 to 2.3.2-7. The second set includes test structures for evaluating capacitors



D	=	DIFFUSION	XT	=	TRANSISTOR
P	=	POLY, TYPE	CR	=	CONTACT RESISTOR
M	=	METAL	XBR	=	CROSS-BRIDGE RESISTOR
F	=	FIELD	INV	=	INVERTER
N	=	TYPE	SUB	=	SUBSTRATE
W	=	WIDTH			
L	=	LENGTH			

ALL DIMENSIONS IN MICROMETERS

Figure 2.3.2-1. Test strip 8205.

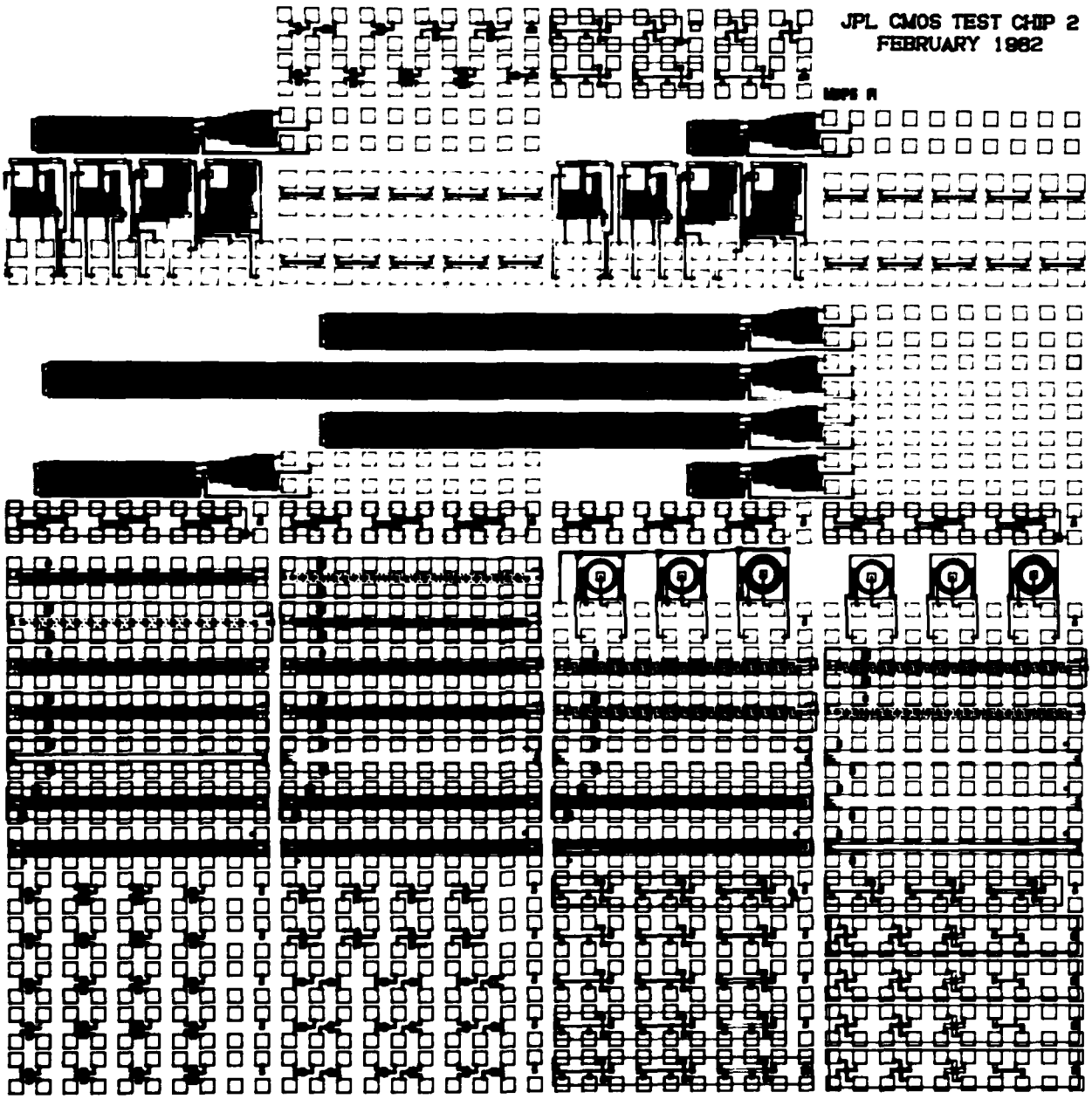


Figure 2.3.2-2. Parametric test chip 2012.

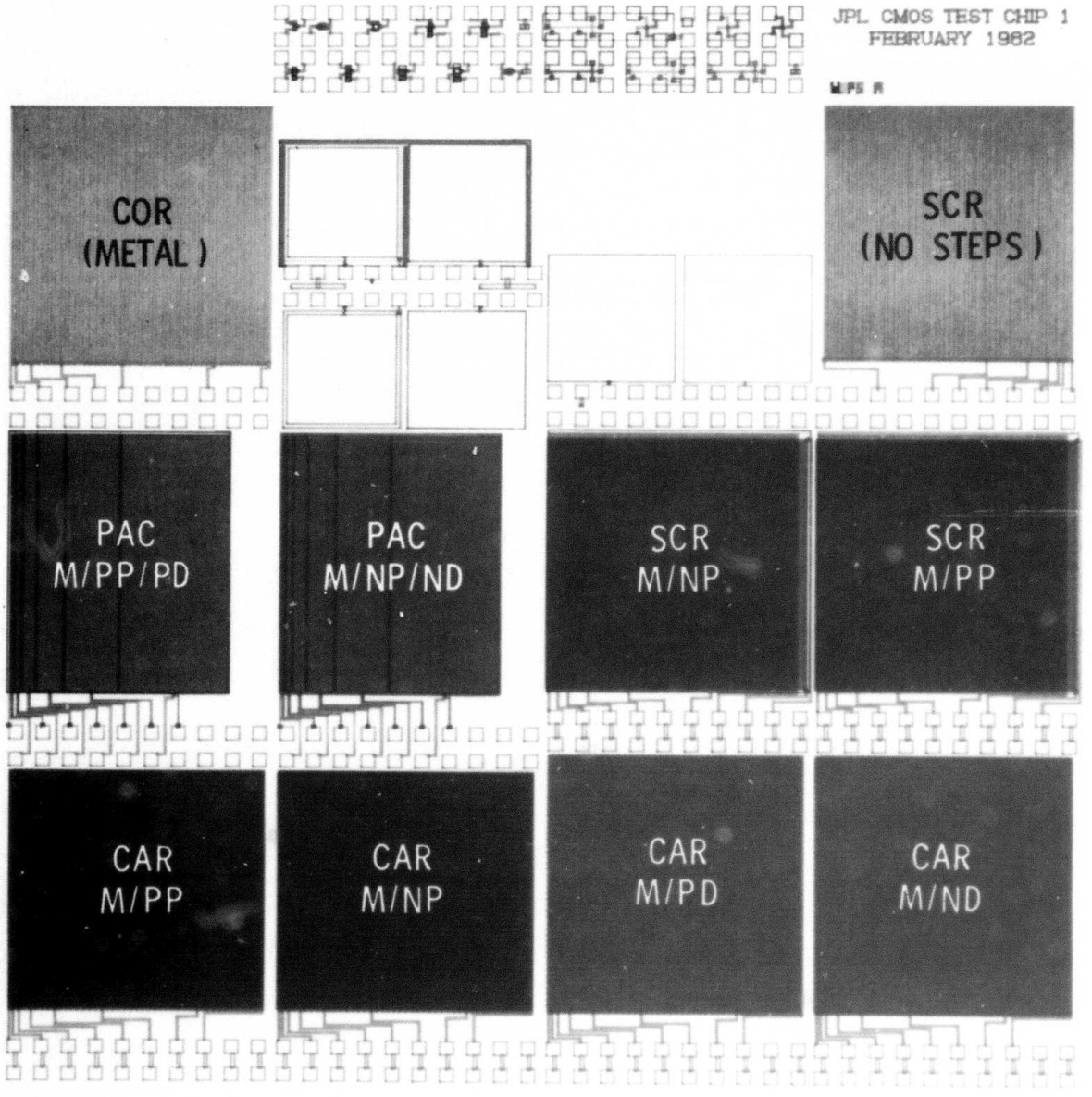
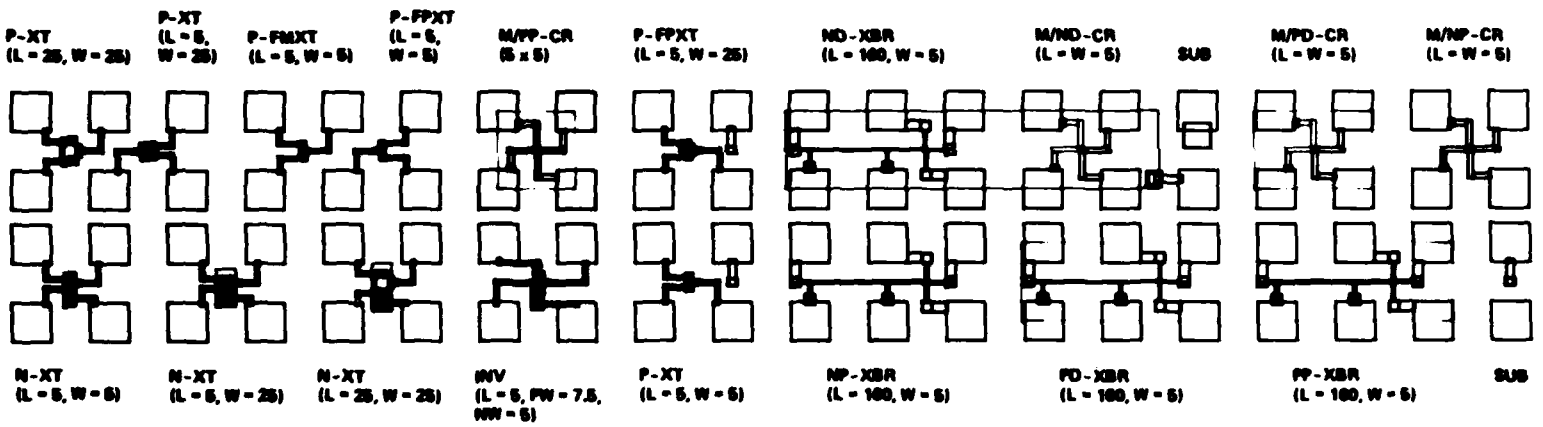


Figure 2.3.2-3. Random fault test chip 2011.



D	=	DIFFUSION	XT	=	TRANSISTOR
P	=	POLY. TYPE	CR	=	CONTACT RESISTOR
M	=	METAL	XBR	=	CROSS-BRIDGE RESISTOR
F	=	FIELD	INV	=	INVERTER
N	=	TYPE	SUB	=	SUBSTRATE
W	=	WIDTH			
L	=	LENGTH			

ALL DIMENSIONS IN MICROMETERS

Figure 2.3.2-4. Test strip 8207.

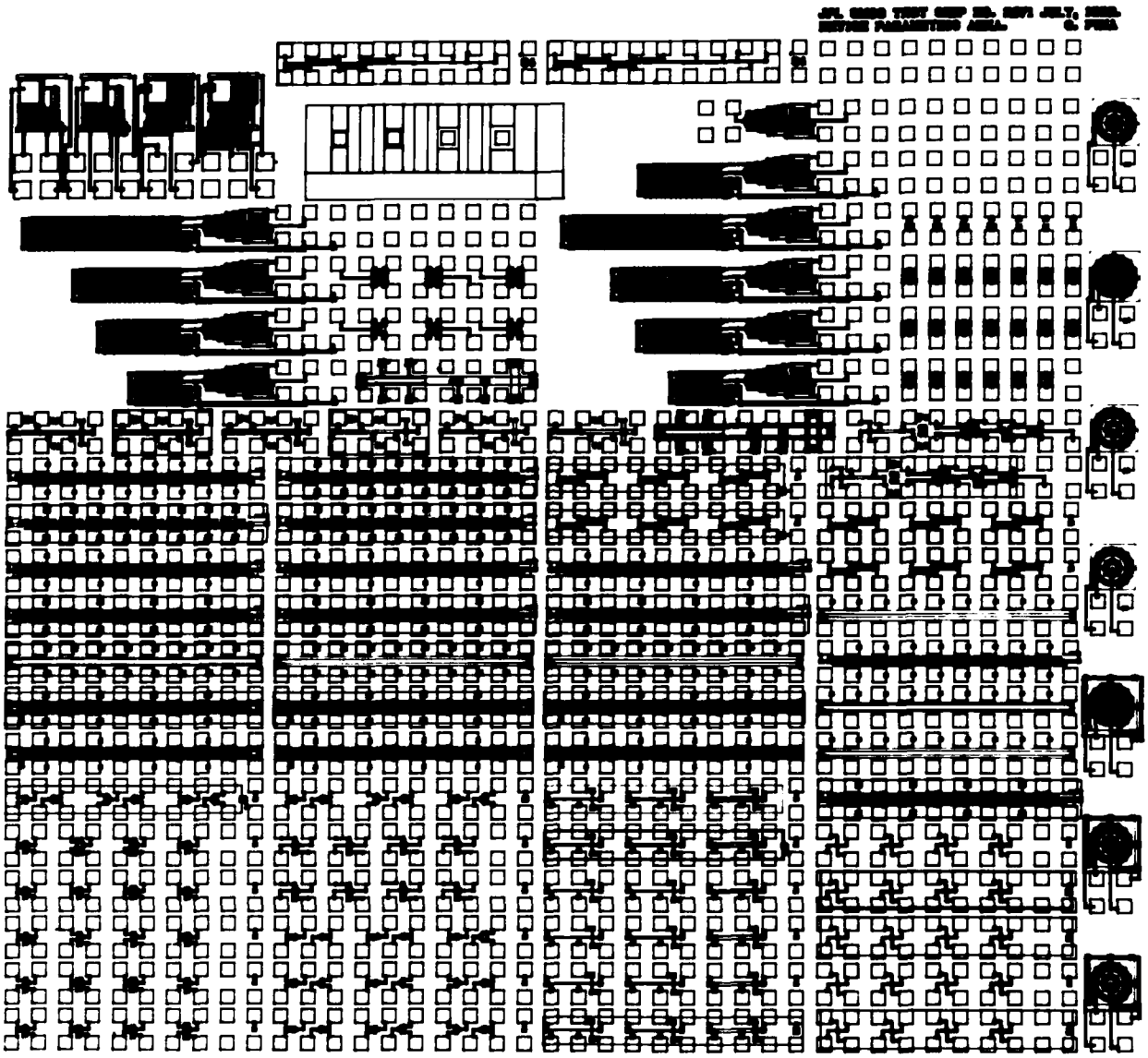


Figure 2.3.2-5. Parametric test chip 2071.

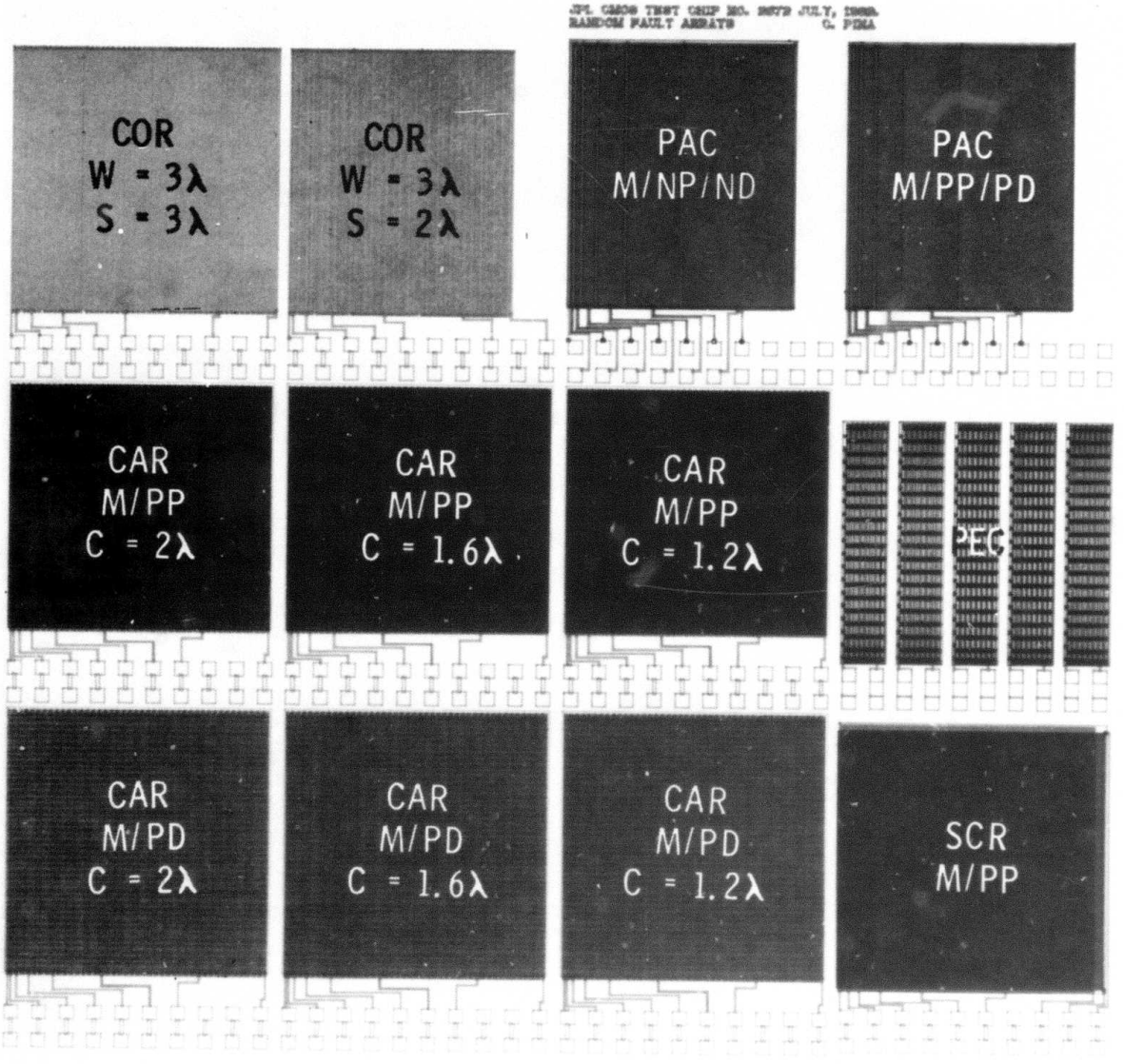


Figure 2.3.2-6. Random fault test chip 2072. Lambda is 2.5 micrometers.

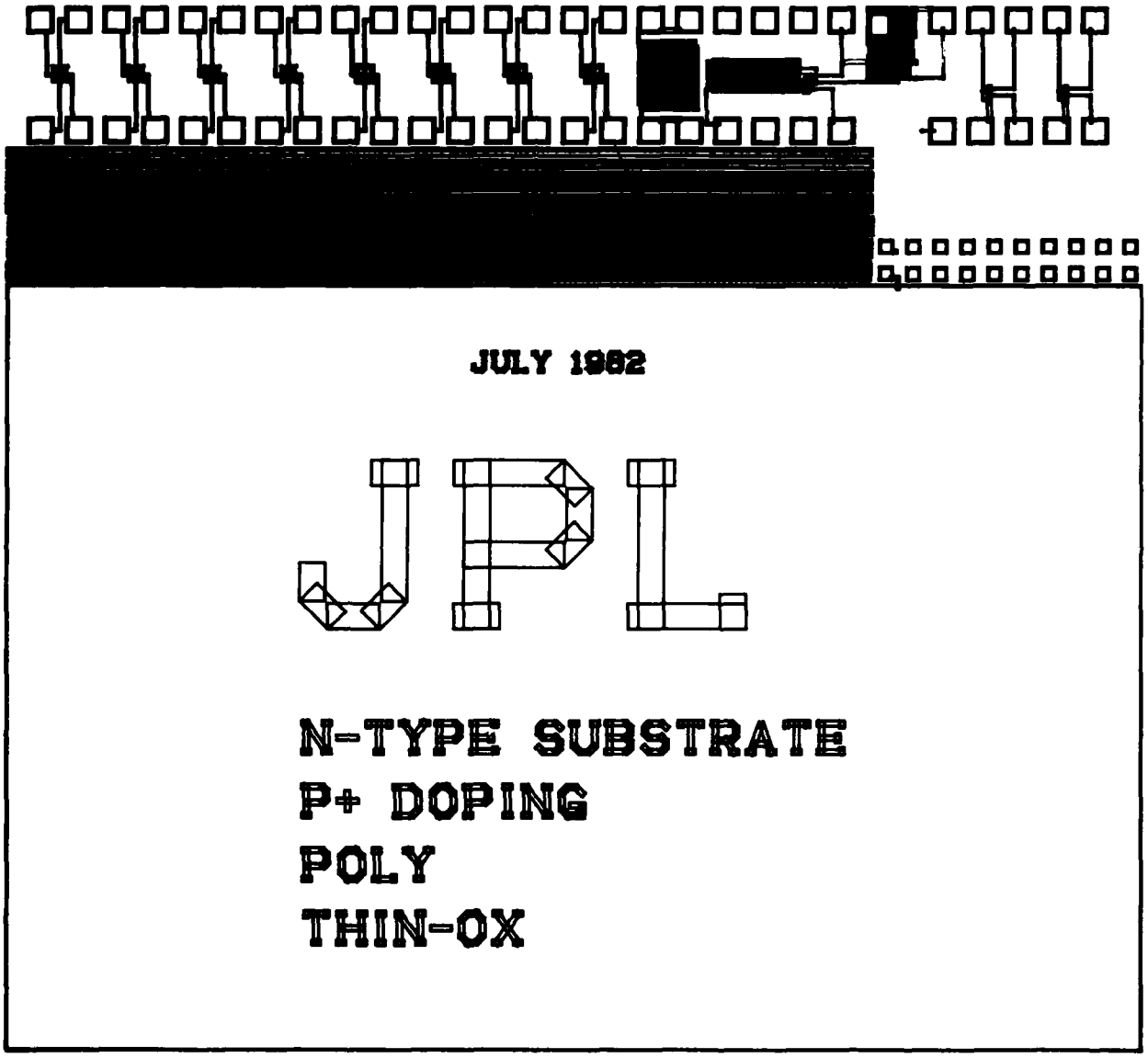


Figure 2.3.2-7. Radiation test chip.

formed by two levels of polycrystalline silicon. A summary of CMOS test chips and foundry runs is listed in Table 2.3.2-1.

The notation used to describe the random fault test chips shown in Figures 2.3.2-3 and 2.3.2-6 is:

CAR = Contact array resistor

SCR = Step-coverage resistor

COR = Comb resistor

PAC = Pinhole array capacitor

PEC = Polyelectrode capacitor

Additional notation consists of:

D = Diffusion

P = Poly or p-type

M = Metal

N = n-type

W = Linewidth

S = Line spacing

C = Contact window width

The documentation of the test chips' geometrical details is contained in a plan file. A sample of the plan file is given in Table 2.3.2-2 for the test chip shown in Figure 2.3.2-2. Results from selected test structures are presented in Section 2.5.2.

Future work will involve developing generalized approaches that reduce the labor-intensive nature in documenting test structures. The difficulty of the problem is illustrated by the documentation required for the NMOS test chip described in Section 2.4.2.

Table 2.3.2-1. Summary of CMOS-Bulk Test Chips and Foundry Runs

FOUNDRY RUN NO.	TECHNOLOGY	FILES CLOSED	WAFERS RECEIVED	TEST STRIP	TEST CHIP NUMBER		
					PARAMETRIC	YIELD	RADIATION
1	5 μm	2/28/82	5/12/82	8205	2012	2011	
2	5 μm	2/28/82	9/20/82	8205	2012	2011	
3	5 μm	7/22/82	1/14/83	ISI	2071	2072	1, 2, 3, 4
4	5 μm	10/15/82	1/14/83	ISI	2071	2072	1, 3, 3, 4
5	5 μm	11/16/82			2111	2112	
6	3 μm	12/20/82		ISI	2113, 2115	2114	1
7	5 μm	1/27/83		ISI	3011	3012	

Table 2.3.2-2. Example Plan File Documentation for Several Inverters in the Test Chip Shown in Figure 2.3.2-2

BLOCK (1,1) ; Coordinates llx,lly: (0.0 , 0.0)

INVERTERS

CIF Symbol	N-CH W(u)	N-CH L(u)	P-CH W(u)	P-CH L(u)	Pad Numbers for: VDD	VSS	VIN	VOUT	Contact Sizes(u): cs1	cs2	cs3	cs4	cs5	cs6
30	5.0	5.0	7.5	5.0	2	3	4	1	2.0	2.0	2.0	2.0	2.0	15.0
35	5.0	5.0	7.5	5.0	6	7	8	5	2.5	2.5	2.5	2.5	2.5	15.0
40	5.0	5.0	7.5	5.0	10	11	12	9	3.0	3.0	3.0	3.0	3.0	15.0
45	5.0	5.0	7.5	5.0	14	15	16	13	3.5	3.5	3.5	3.5	3.5	15.0

BLOCK (1,2) ; Coordinates llx,lly: (0.0 , 270.0)

INVERTERS

CIF Symbol	N-CH W(u)	N-CH L(u)	P-CH W(u)	P-CH L(u)	Pad Numbers for: VDD	VSS	VIN	VOUT	Contact Sizes(u): cs1	cs2	cs3	cs4	cs5	cs6
55	5.0	5.0	7.5	5.0	2	3	4	1	5.0	5.0	5.0	5.0	5.0	10.0
60	5.0	5.0	7.5	5.0	6	7	8	5	5.0	5.0	5.0	5.0	5.0	11.3
65	5.0	5.0	7.5	5.0	10	11	12	9	5.0	5.0	5.0	5.0	5.0	12.5
70	5.0	5.0	7.5	5.0	14	15	16	13	5.0	5.0	5.0	5.0	5.0	13.8

BLOCK (1,3) ; Coordinates llx,lly: (0.0 , 540.0)

INVERTERS

CIF Symbol	N-CH W(u)	N-CH L(u)	P-CH W(u)	P-CH L(u)	Pad Numbers for: VDD	VSS	VIN	VOUT	Contact Sizes(u): cs1	cs2	cs3	cs4	cs5	cs6
80	5.0	5.0	7.5	5.0	2	3	4	1	5.0	5.0	5.0	5.0	5.0	15.0
85	5.0	5.0	7.5	5.0	6	7	8	5	5.0	5.0	5.0	5.0	2.0	15.0
90	5.0	5.0	7.5	5.0	10	11	12	9	5.0	5.0	5.0	5.0	2.5	15.0
95	5.0	5.0	7.5	5.0	14	15	16	13	5.0	5.0	5.0	5.0	3.0	15.0

BLOCK (1,4) ; Coordinates llx,lly: (0.0 , 810.0)

INVERTERS

CIF Symbol	N-CH W(u)	N-CH L(u)	P-CH W(u)	P-CH L(u)	Pad Numbers for: VDD	VSS	VIN	VOUT	Contact Sizes(u): cs1	cs2	cs3	cs4	cs5	cs6
105	5.0	5.0	7.5	5.0	2	3	4	1	5.0	5.0	5.0	5.0	3.5	15.0
110	5.0	5.0	7.5	5.0	6	7	8	5	5.0	5.0	5.0	5.0	4.0	15.0
115	5.0	5.0	7.5	5.0	10	11	12	9	5.0	5.0	5.0	2.0	5.0	15.0
120	5.0	5.0	7.5	5.0	14	15	16	13	5.0	5.0	5.0	2.5	5.0	15.0

2.4 TEST STRUCTURES

The categories of test structures is discussed in Section 2.4.1 and the test module concept for describing the elements used to design and evaluate a test structure is described in Section 2.4.2. The use of the test module concept is illustrated by the documentation for an NMOS test strip. Two additional test modules are described: the split-cross-bridge resistor, in Section 2.4.3, and the yield analysis test structures, in Section 2.4.4.

2.4.1 Categories of Test Structures

In order to organize the development of the test structures, the test structures have been divided among the following six categories:

1. Process parameter extraction
2. Device parameter extraction
3. Circuit parameter extraction
4. Layout rule checking
5. Yield analysis
6. Reliability analysis

These categories are described in the following paper.

by M.G. Buehler, T.W. Griswold, C.A. Pina, and C. Timoc, California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA.

Test Chips for Custom ICs

six kinds of test structures

In working with a silicon foundry, one needs information to qualify the fab line to produce the desired custom ICs. Most of this data can be gleaned from test chips located on the wafer. Test chips then become information gathering systems encoded by the wafer fabrication process. But it remains for the test engineer to devise the methods for retrieving test chip data.

Microelectronic test chips contain a number of test structures that are used for a variety of purposes in the fabrication of integrated circuits. For convenience, their use is divided into six categories: layout-rule evaluating, process-parameter extraction, device-parameter extraction, circuit-parameter extraction, initial-fabrication failure analysis, and reliability failure analysis. A particular test structure can be used to gather information in a number of categories mentioned above. For example, a serpentine contact-resistor array can be used to check layout rules, identify the occurrence of open contacts (an initial fabrication failure), and evaluate the reliability of the contacting technology. This paper gives examples of the kinds of parameters that can be obtained in each of these categories.

Evaluating Layout Rules

A set of layout rules has been developed for a bulk CMOS process that uses local oxidation, p-well, self-aligned polycrystalline silicon gate, and aluminum metallization (Fig 1). In this process, a nitride layer is deposited early, but after the local oxidation step, which forms the field oxide, the nitride layer is removed and does not appear on the finished product. Butting pn junction contacts short the source of the n-channel transistor to the p-well and the source of the p-channel transistor to the substrate. Doping is achieved by ion implantation, and the polycrystalline silicon is doped with phosphorus. Nine photomasks, including the passivation layer, are required.

Describing the rules is a compact notation based on six geometrical primitives that define a feature and its relationship to another feature on either the same or a different photomask level.

For this CMOS process, one primitive is absent. The notation for the minimum set of levels consists of:

T = Thin-Oxide
W = Well
G = Gate
P = p+ Source/Drain Doping
C = Contact
M = Metal

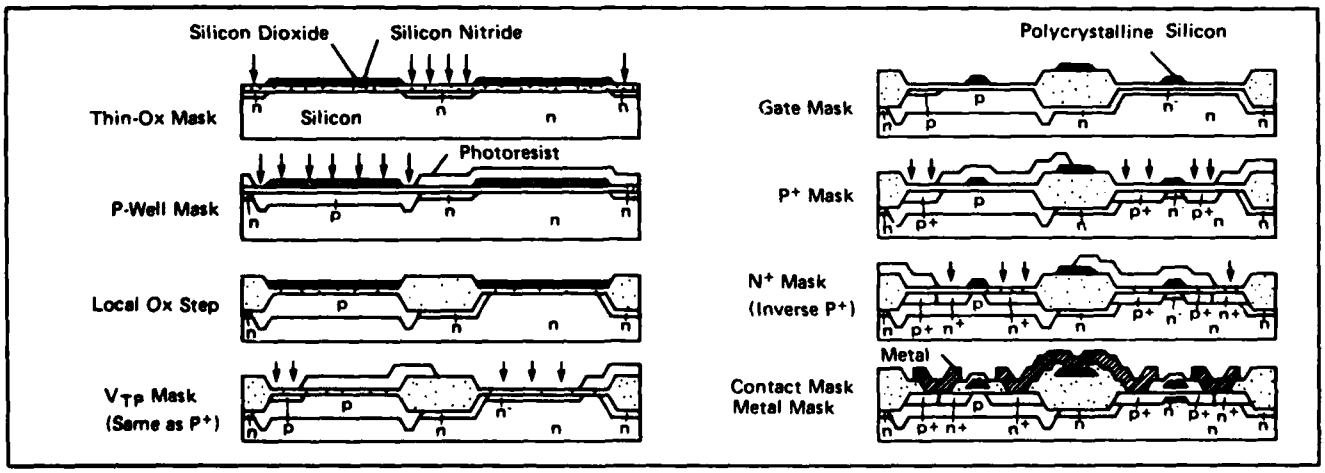
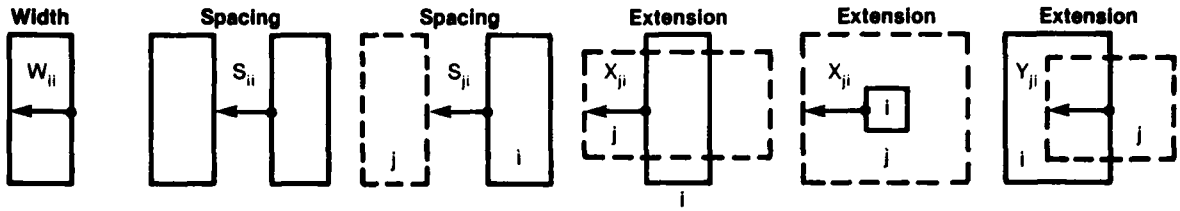


Fig 1 shows cross sections of a CMOS bulk inverter being fabricated by using local oxidation, p-well, self-aligned gate of polycrystalline silicon, and aluminum metallization. Layout-rule checking was developed for this process. Note that the diagrams are unscaled.

Reprinted from CIRCUITS MANUFACTURING June 1982 © Morgan-Grampian Publishing Company

Table 1
Test structures for checking various CMOS bulk layout rules.



Width	Spacing	Spacing	Extension	Extension	Extension
$W_{ww}(1)$	$S_{ww}(1)$	S_{tw}	$X_{gt}(1,2)$	$X_{tc}(11)$	$Y_{tcp}(12)$
$W_{tt}(1)$	$S_{tt}(1)$	S_{gt}	$X_{tg}(1)$	$X_{gc}(11)$	
$W_{gg}(1)$	$S_{gg}(1,15)$	$S_{pt}(16)$	$X_{pg}(1)$	$X_{mc}(12)$	
$W_{pp}(1)$	$S_{pp}(1)$	S_{gct}	$X_{tcp}(12)$	X_{wt}	
$W_{cc}(13)$	S_{cc}	$S_{pg}(1)$		$X_{pt}(16)$	
$W_{mm}(1,14)$	$S_{mm}(1,15)$	S_{tcg}			

Direct Measurement

1. Cross-Bridge Resistor
2. Capacitor

Indirect Measurement

11. Alignment Resistor
12. Contact Resistor
13. Contact Resistor Array
14. Serpentine Resistor
15. Comb Resistor
16. Collision Resistor

The symbol G refers to all polycrystalline material used to form both gates and wires. The other levels (e.g., V_{TP}, N^+) are derived from the above set. For simplicity, the passivation level has been omitted from this discussion.

Table 1 lists the 28 layout rules required to define this CMOS process. Also shown in this table are the electrical test structures used to evaluate certain rules. Note that by using electrical test structures 14 rules can be evaluated directly, and eight more indirectly. The remaining six rules can be evaluated using optical inspection techniques. More difficult measurements involve the contacts and the P level. The P level leaves no mark in the oxide that can be observed visually.

All but four of the 28 rules involve two layer interactions. The exceptions are: S_{gct}, the spacing of gate contact from the thin-oxide; S_{tcg}, the spacing of the thin-oxide contact from the gate; and X_{tcp} and Y_{tcp}, extensions of the thin-oxide contact from the P level.

Test Structures for Failure Analysis

Initial-fabrication failures result from processing defects that appear before parts are stressed, whereas reliability failures are caused by processing defects that shorten the life of a part due to stress the part receives during operation. For example, a metal step-coverage problem is classified as an initial-fabrication failure if a metal line is broken prior to stress. If the metal line breaks after a current stress, then the problem is classified as a reliability failure.

Table 2 summarizes various device failures common to bulk CMOS and indicates failure mechanisms appearing after wafer fabrication and after stress. The types of stress conditions consist of voltage, electric field, current density, temperature, humidity, and radiation. Test structures used to detect the failures are also listed in the table. The testing required for these structures is often statistical in nature, for one is usually trying to characterize t. frequency of a fault.

Extracting Process and Device Parameters

The test structures in Table 3 are used to obtain process and device parameters. Those parameters listed in the table are needed by the level-two SPICE MOSFET model. As seen from the table, a variety of test structures is necessary to extract the data, and both DC and AC measurements are required. Test procedures used to obtain each parameter vary among users. Tradeoffs must be made between data acquired rapidly in a production environment and accurate data acquired in a research setting.

A production-oriented measurement scheme for characterizing MOSFETs has been documented by Ham. In this procedure only 12 quantities (either currents or voltages) are measured; the small number of measurements allows rapid data acquisition. Before measuring specific device parameters, the device is given a "suitability test," which determines the following:

- Threshold voltage must be stable.
- Gate leakage current must be low.
- Channel must conduct sufficient current.
- Contact resistance must be low.

If the device does not pass these initial tests, then further testing is halted, and the reason for the failure recorded. If it does pass, the threshold voltage, conduction factor, and transconductance of the MOSFET are determined in both the linear and saturated regions of operation.

Testing for Circuit Parameters

Test structures are used to extract circuit parameters that characterize

Table 2

Test Structures for Failure Analysis of CMOS Failure Mechanisms.

	Failure Mechanism				
Region	**Device Failure**	**Initial Fab Failure**	**Reliability Failure**	**Test Structure**	**Quantity Measured**
Silicon	Excessive leakage, premature junction breakdown	Stacking faults, deep traps, metallic precipitates	Radiation induced interface states (R), inversion layers induced by surface ion migration (E,H)	Diode, ring oscillator	Leakage current, stage delay
Metal, poly	Metal open at oxide step, poly open at oxide step	Improper metal or poly deposition, oxide tapering	Electromigration (T,J)	Serpentine resistor	Resistance, frequency of opens
Metal, poly	Metal to metal short, poly to poly short	Poor lithography, incomplete etch	Corrosion due to residual contamination (H,V,T)	Comb resistor	Resistance, frequency of shorts
Contact	Metal to silicon open, metal to poly open	Residual oxide at contact due to poor contact etch	Charge injection into residual oxide (E,T)	Serpentine contact resistor	Resistance, frequency of opens
Oxide	Metal to poly short, metal to silicon short, poly to silicon short	Pinholes in oxide due to poor lithography or poor oxide growth	Clustering of Na^+ ions at oxide/silicon interface leading to premature dielectric breakdown (T,V)	Overlapping serpentine resistor, capacitor	Resistance, frequency of shorts, time dependent oxide breakdown
Oxide	Excessive shift in threshold voltage, channeling	Contaminated oxides, improper metal or poly deposition	Oxide trap charging due to electron injection from silicon (E,T) or radiation (R), ion migration within oxide (E,T) or on surface of oxide (E,T,H)	MOSFET	Threshold voltage
Package	Device burn-out, drift in parameters	Excessive power dissipation in chip due to poor die attachment	Drift in parameters due to temperature activated failure modes	Diode	Chip temperature

E = Electric Field, V = Voltage, J = Current Density, T = Temperature, H = Humidity, R = Radiation

AC and DC circuit performance and to verify that the wafer fabrication process can produce functional circuits. For example, inverters measure inverter threshold, gain, and noise immunity, while ring oscillators measure the oscillation frequency and stage delay.

Full characterization of circuit performance involves evaluating such parameters as power dissipation, maximum clock frequency, fan-out capability, and propagation delay.

CMOS Test Chips

In general, test structures naturally divide between two test chips. Structures for process, device and circuit extraction appear on one chip composed of discrete test structures such as resistors, diodes, and transistors. On the other hand, structures for layout-rule checking, initial-fabrication failure analysis, and reliability failure analysis are situated on another chip containing elements connected in arrays or large area structures. Separating the structures into these two types of test chips allows the chip with arrays to be reproduced across the wafer more frequently than the one with discrete structures so that increased statistical information can be obtained from the wafer.

The test chips' design required that all structures be accessible through a single probe card. For this layout a 2 by 10 probe pad array is used where the probe pads are 80-μm squares separated by 80 μm.

Sources for Additional Reading

1. M.G. Buehler and L.W. Linholm, "Toward a Standard Test Chip Methodology for Reliable, Custom Integrated Circuits," *Proc. 1981 Custom IC Conf.*, p. 142 (May 1981).
2. M.G. Buehler, "Microelectronic Test Chips for VLSI Electronics," in *VLSI Electronics: Microstructure Science*, supplementary volume on Materials and Process Characterization in VLSI Electronics, ed. N. Einspruch and G. Larabee, to be published by Academic Press.
3. L.W. Nagel, SPICE2: A Computer Program to Simulate Semiconductor Circuits, Memorandum No. ERL-MS10, Electronics Research Laboratory, University of California, Berkeley, CA (May 9, 1975).
4. W.E. Ham, "Comprehensive Test Pattern and Approach for Characterizing SOS Technology," NBS Spec. Publ. 400-56 (January 1980).
5. M.G. Buehler, "Comprehensive Test Patterns with Modular Test Structures: The 2 by N Probe Pad Array Approach," *Solid State Technology*, 22, 89 (October 1979). ■

Martin G. Buehler received a BSEE and MS from Duke University and in 1966 received a PhD in electrical engineering from Stanford University. Prior to joining the National Bureau of Standards in 1972, he was a staff member of the Semiconduc-

Table 3
Test Structures for Extracting CMOS Bulk Process and Device Parameters

Parameter	Name	Test Structure
VTO	Threshold voltage (linear)	Transistor
KP	Conduction factor (linear)	Transistor
Gamma	Body effect factor (linear)	Transistor
PHI	Surface built-in potential	Transistor
Lambda	Channel narrowing (saturation)	Long transistor
CGSO	Gate-source capacitance	Meander transistor
CGDO	Gate-drain capacitance	Meander transistor
CGBO	Gate-body capacitance	Field-oxide capacitor
RS DIFF	Diffused layer sheet resistance	Cross resistor
CJ	Junction capacitance	Large diode
MJ	CJ exponential factor	Large diode
CJSW	Junction side-wall capacitance	Meander diode
MJSW	CJSW exponential factor	Meander diode
PB	Junction built-in potential	Meander diode
JS	Junction saturation current	Diode/transistor
TOX	Gate oxide thickness	Large gate capacitor
NSUB	Body dopant density	Transistor capacitor

tor Research and Development Laboratories, Texas Instruments, Inc., where he applied electrical measurement techniques to the detection of defects and the profiling of dopants in semiconductor materials. At NBS he led a group in process metrology, designed nearly a dozen semiconductor test structures, was awarded the Department of Commerce's Silver Medal for creative contributions to semiconductor metrology, and initiated a program on self-test techniques for VLSI. Currently he is the principal investigator of the Product Assurance Technology Program for LSI/VLSI at the Jet Propulsion Laboratory. Buehler is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, the Electromechanical Society, and IEEE.

***Thomas W. Griswold** earned from the University of California, Berkeley, an AB in physics and in 1953 a PhD in solid state physics. From 1953 to 1958 he worked on developing germanium and silicon diodes and transistors at the Semiconductor Division, Hughes Aircraft Co. He then moved to Continental Device Corp. as technical director for silicon transistor and IC development until 1972. Griswold spent the next three years in high-voltage rectifier technology at Semtech Corp. In 1975 he joined the Jet Propulsion Laboratory where he has supervised since 1978 the LSI Technology Group in the Information Systems Research Section.*

***Cesar Pina** received a BS in mathematics from CSLUB in 1968. From 1959 to 1963 he was employed in development and process engineering by Hoffman Electronics Corp., and from 1963 to 1965 he worked there as director of reliability and quality assurance. In 1966 Pina joined Teledyne Semiconductor and later became engineering manager. In 1971 he left Teledyne to start Regulus Semiconductor Corp., where he remained until 1975, when he joined Microsemiconductor Corp. as manufacturing manager. Pina stayed there until joining JPL in 1980.*

***C. Timoc** received the EE degree from the Polytechnic Institute, Bucharest, Romania, in 1966 and an MS in electrical engineering and computer science from Columbia University in 1972. From 1972 to 1976 he worked for IBM, East Fishkill, NY, in IC fault detection and location. In 1976 he founded a consulting firm presently called Spaceborne, Inc., engaged in test generation and manufacturing of hardware fault simulators for VLSI. His other research interests are design for testability and fault tolerant systems.*

2.4.2 Test Methods for a CMOS-bulk Test Strip

The purpose of this section is to give a brief description of the test procedures used to measure parameters from the CMOS-bulk test strip shown in Figure 2.3.2-4. Included are procedures for:

1. A gate-oxide transistor
2. A field-oxide transistor
3. An inverter
4. A cross-bridge resistor
5. A contact resistor

The description of the test structure is contained within a test module that includes the following elements:

1. Test module name
2. Purpose or function
3. Geometrical description
4. Test procedures
   a. Circuit diagram
   b. Test conditions
   c. Data reduction algorithm
5. References

In future test procedures we intend to include a section on suggested screening procedures that can be used to identify a faulty test. It has been observed that wafer-level measurements of test structures can lead to invalid data caused by instrumentation problems (due to probes failing to contact probe pads), test structure problems (due to missing parts of a structure because of poor lithography), or electrically unstable structures (due to contamination). Such screening procedures can save overall measurement time by allowing one to abort a faulty measurement early in the test sequence.

In the following description of transistor measurements, the transistor elements are noted by: S for source, D for drain, G for gate, and B for body.

2.4.2.1 Gate-Oxide Transistor

2.4.2.1.1 Test Module: Four terminal n-channel enhancement mode MOS transistor (gate oxide).

2.4.2.1.2 Purpose: To evaluate the transistor

a. Channel leakage current, IDSO

b. Source and drain diode leakage current, IDBLEAK

c. Source and drain diode breakdown voltage, VDBBD

d. Gate leakage current, IGBLEAK

e. Threshold voltage, VTO

f. Conduction factor, KP

g. Body threshold parameter, GAMMA

2.4.2.1.3 Geometrical Description: See Figure 2.4.2-1.

2.4.2.1.4 Test Procedures:

a. Transistor channel leakage current (OFF current), IDSO:

1. Circuit Diagram: See Figure 2.4.2-2.

2. Test Conditions: IDSO is the current that flows in the reverse direction through the drain diode with the gate biased so that the channel is turned off. IDSO is measured with VGS = VBS = 0 and VDS = 5 volts.

3. Data Reduction Algorithm: IDSO is measured directly.

b. Source and drain diode leakage current, IDBLEAK:

1. Circuit Diagram: See Figure 2.4.2-3.

2. Test Conditions: IDBLEAK is the sum of the leakage currents that flow in the reverse-biased source and drain diodes. IDBLEAK will also include leakage currents that flow through the gate oxide that overlaps the source and drain junctions. IDBLEAK is measured with VDS = 0, VDB = -5 volts, and VGS = -5 volts.

3. Data Reduction Algorithm: IDBLEAK is determined directly from the measured value.

c. Source and drain breakdown voltage, VDBBD:

1. Circuit Diagram: See Figure 2.4.2-4.

2. Test Conditions: VDBBD is the breakdown voltage measured across both the source and drain diodes with 1 microampere forced in the reverse-bias direction through the diodes. VDBBD is measured with VDS = 0, VGB = 0, and IDB = 1 microampere.

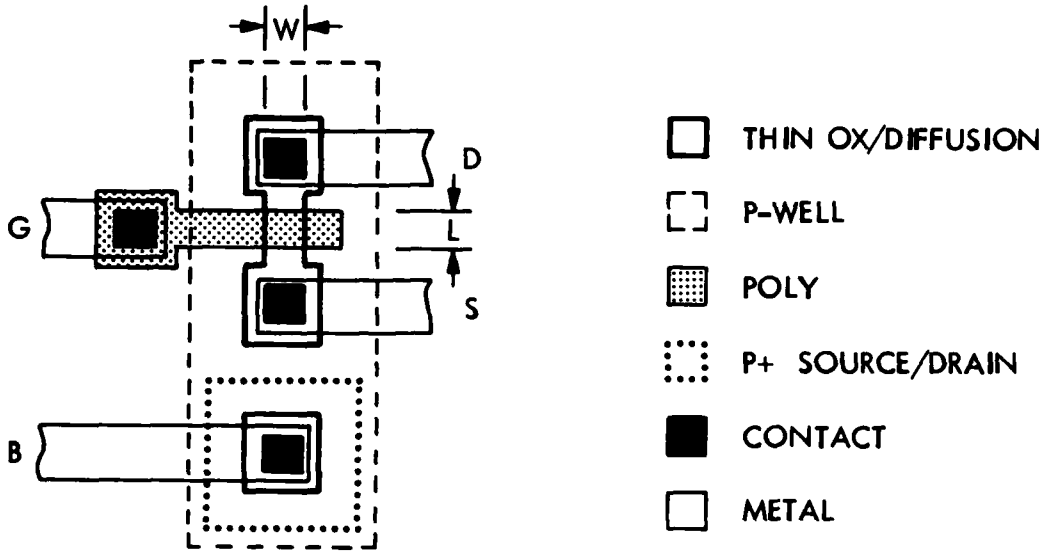


Figure 2.4.2-1. Layout of a four-terminal n-channel MOS transistor (N - XT), where L = W = 5 micrometers.

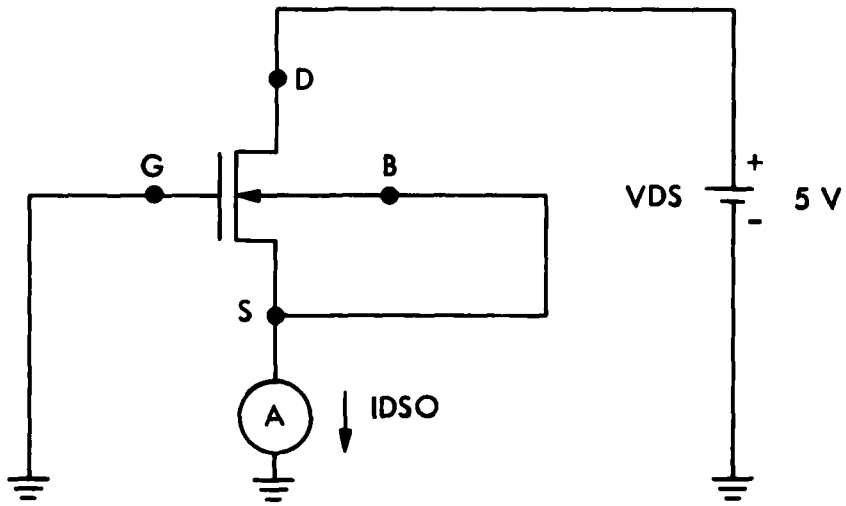


Figure 2.4.2-2. Channel leakage current, IDSO, measurement circuit.

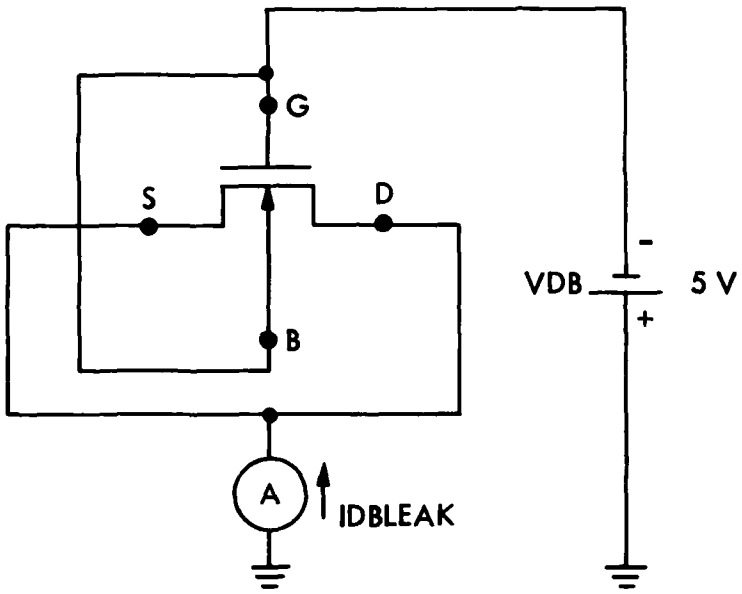


Figure 2.4.2-3. Source and drain diode leakage, IDBLEAK, measurement circuit.

3. Data Reduction Algorithm: VDBBD is determined directly from the measured value.

d. Gate leakage current, IGBLEAK:

1. Circuit Diagram: See Figure 2.4.2-5.

2. Test Conditions: IGBLEAK is the sum of the currents that flow from the gate to body, source, and drain with the gate biased so that the channel is turned off. IGBLEAK is measured with VDB = 0, VBS = 0, and VGB = 10 volts.

3. Data Reduction Algorithm: IGBLEAK is determined directly from the measured value.

e. Threshold voltage, VTO, condition factor, KP, and body threshold parameter, GAMMA:

1. Circuit Diagram: See Figure 2.4.2-6.

2. Test Conditions: The parameters VTO, KP, and GAMMA are measured in the linear region of transistor operation where VGS - VT > VDS. In this region the transistor equation is:

$$IDS = KP(W/L)(VGS - VT - VDS/2)VDS$$

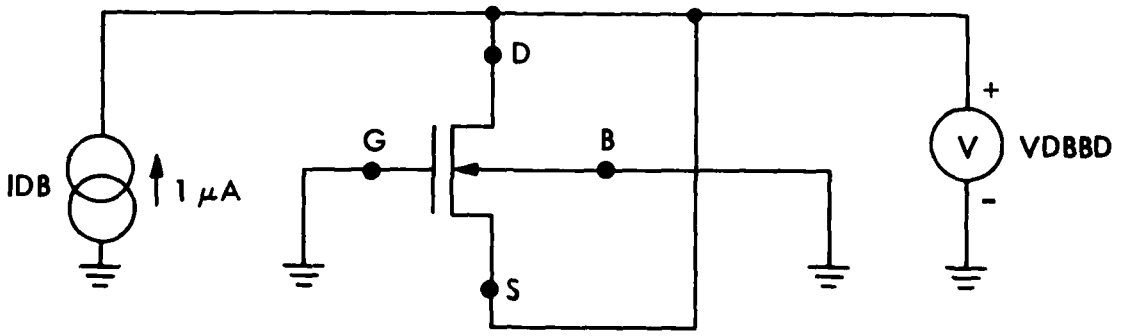


Figure 2.4.2-4. Source and drain diode breakdown voltage, VDBBD, measurement circuit.

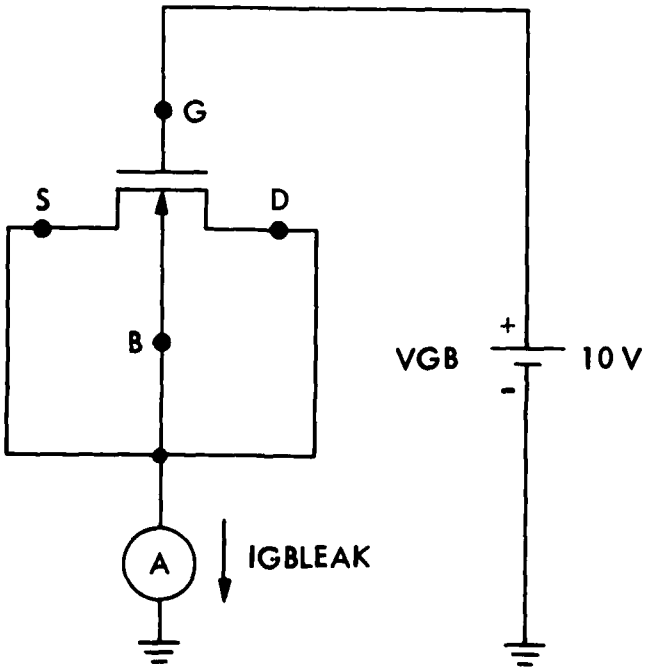


Figure 2.4.2-5. Gate leakage current, IGBLEAK, measurement circuit.

where the threshold voltage is:

$$VT = VTO - GAMMA\ \phi^{1/2} + GAMMA\ (\phi - VBS)^{1/2}.$$

The parameter W is the as-drawn channel width, L is the as-drawn channel length, and ϕ is two times the Fermi potential. The three parameters are determined from the steepest slope of the IDS versus VGS curve for VDS = 50 millivolts and VBS = constant.

The method is illustrated in Figure 2.4.2-7, where IDS and its first and second derivatives with respect to VGS are plotted against VGS. The voltage where the maximum slope occurs is determined from the zero crossing of the $d^2IDS/dVGS^2$ versus VGS curve. The maximum slope is determined from a total of 13 IDS, VGS data points taken on either side of the maximum slope point. A special fitting routine is used to determine the coefficient of a cubic equation that best fits the 13 IDS, VGS data points. The straight-line equation for the maximum slope is determined at the cubic equation's midpoint. (The fitting procedure was derived from a paper in Analytical Chemistry, 36 (8), 1627-1639 (July 1965), and Analytical Chemistry, 44 (11), 1906-1909 (September 1972).)

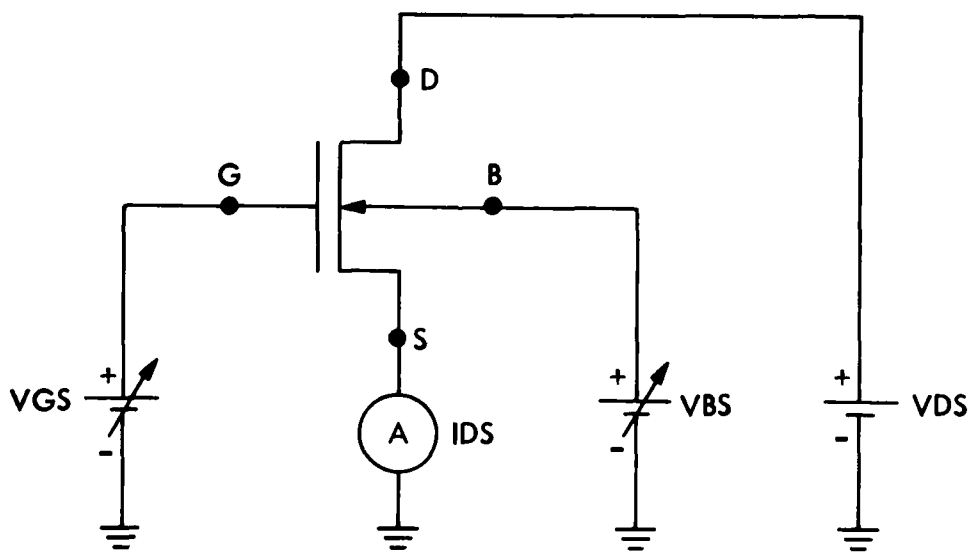


Figure 2.4.2-6. Transistor parameter (VTO, KP, and GAMMA) measurement circuit.

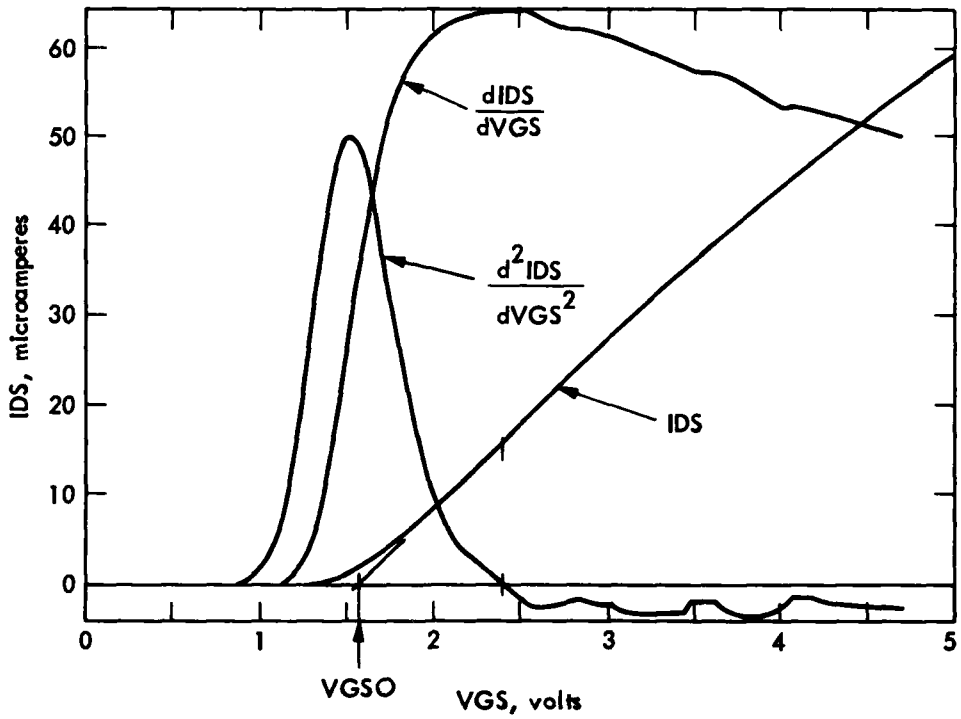


Figure 2.4.2-7. Maximum slope approach to evaluating VTO and KP for an NMOS transistor, where W = 75 micrometers, L = 5 micrometers, and VBS = 0.

The procedure used to determine VTO consists of successive measurements where ΔVGS is reduced in magnitude on successive passes. On the first pass, the procedure begins with default values of VGS = 0 and ΔVGS = 0.2 volts. A set of 13 IDS, VDS data points are stored in a buffer memory and $d^2IGS/dVGS^2$ is calculated at the middle of the data set. The data set is "moved" along the IDS versus VGS curve. When a zero crossing is detected in the second derivative, the threshold voltage, VT, is calculated from the IDS = 0 intercept of the straight line derived from the midpoint of the cubic equation fitted to the data. A new threshold voltage is sought but on the second pass the data collection is started at 0.9 * VT and $\Delta VGS = (VGS_{max} - VT)/(F * 13)$, where VGS_{max} is the maximum VGS value achieved in the previous pass, VT is the previously measured threshold voltage, and F is a factor taken as 1.5. The factor F is used to reduce the value of ΔVGS so that data is not collected too close to IDS = 0. The second threshold voltage is compared to the first value, and if the difference between them is greater than 5 millivolts, then the procedure is repeated until successive values of VT are within 5 millivolts of each other.

The measurement procedure for additional transistors on a wafer uses the VT and VGS_{max} values from the previous transistor to establish ΔVGS as described above and the starting point of 0.9 * VT. A typical value for ΔVGS is about 0.1 volt.

3. Data Reduction Algorithm: The VTO is determined for VBS = 0 and VDS = 50 millivolts. The voltage, VGSO, is determined from the IDS = 0 intercept of the maximum slope line fitted to the IDS versus VGS curve. The VTO follows from:

$$VTO = VGSO - (VDS/2)$$

The KP is determined from the maximum slope line fitted to the IDS versus VGS curve for VBS = 0 and VDS = 50 millivolts. The KP follows from:

$$KP = (L/W)(\Delta IDS/\Delta VGS)/VDS$$

where L and W values are usually derived from the as-drawn dimensions. The GAMMA is determined from a second value of VGSO derived from the maximum slope line fitted to a second IDS versus VGS curve for VBS = -5 volts and VDS = 50 millivolts.

$$GAMMA = \Delta VGSO/\Delta(\phi - VBS)^{1/2}$$

2.4.2.2 Field-Oxide Transistor

2.4.2.2.1 Test Module: Four-terminal MOS transistor (field oxide).

2.4.2.2.2 Purpose: To evaluate the field-oxide threshold voltage, VTFIELD. This method is applicable to both metal-gate and poly-gate transistors.

2.4.2.2.3 Geometrical Description: See Figure 2.4.2-8.

2.4.2.2.4 Test Procedures:

a. Circuit Diagram: See Figure 2.4.2-9.

b. Test Conditions: VTFIELD is the so-called 1-microampere field threshold voltage. The IDS current source is used to force 1 microampere through the channel for a particular VBS value. VTFIELD is measured with IDS = 1 microampere, VBS = 0 or 1 volt, and VGS = VDS. The last condition ensures that the transistor is operating in the saturation region.

c. Data Reduction Algorithm: VTFIELD is determined directly from the measured value.

2.4.2.3 Inverter

2.4.2.3.1 Test Module: CMOS-bulk inverter.

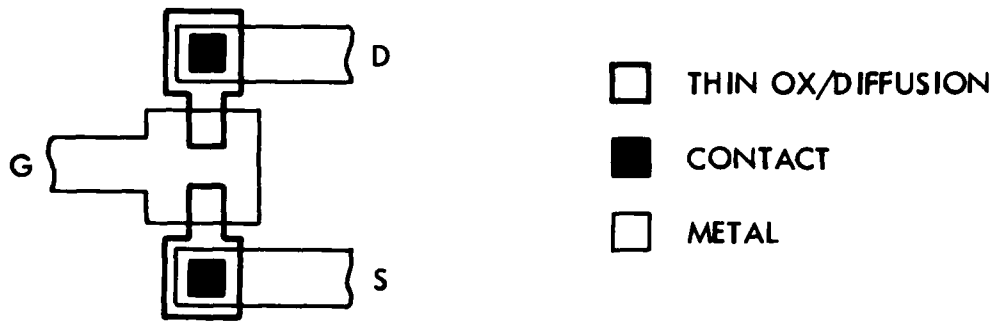


Figure 2.4.2-8. Layout of a four-terminal, field-oxide, metal-gate, n-channel MOS transistor (N - FMXT). The body contact is not shown here; it is included elsewhere on the test strip.

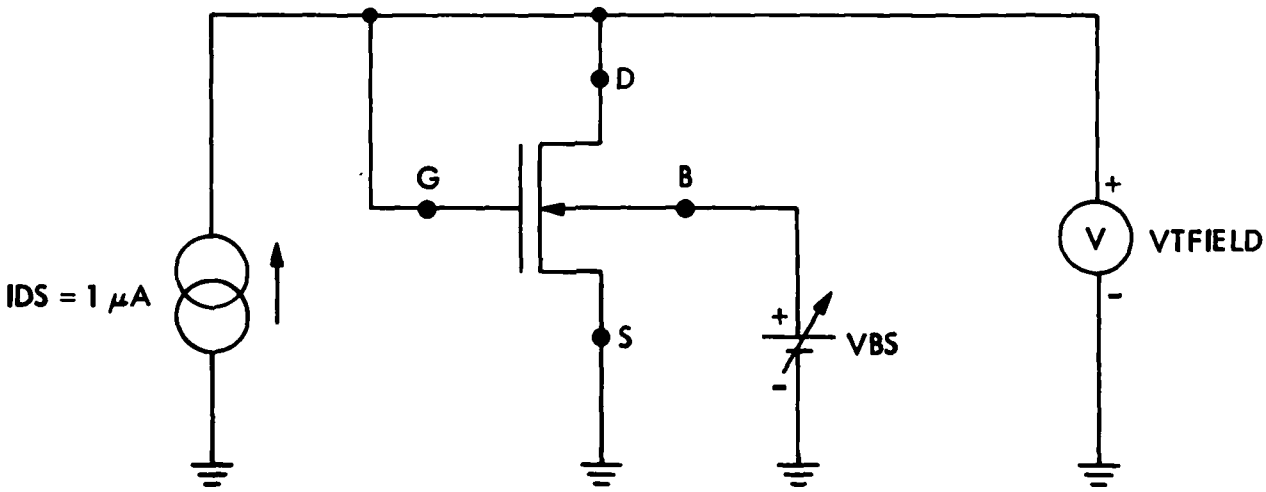


Figure 2.4.2-9. Field-oxide threshold voltage, VTFIELD, measurement circuit.

2.4.2.3.2 Purpose: To evaluate the inverter

a. Output voltage high, VOH

b. Output voltage low, VOL

c. Inverter threshold voltage, VTINV

d. Inverter gain, G

2.4.2.3.3 Geometrical Description: See Figure 2.4.2-10.

2.4.2.3.4 Test Procedures:

a. Output voltage high, VOH, and output voltage low, VOL:

1. Circuit Diagram: See Figure 2.4.2-11.

2. Test Conditions: The parameters VOH and VOL are measured for VI = 0 and VI = 5 volts, respectively.

3. Data Reduction Algorithm: VOH and VOL are determined directly from measured values.

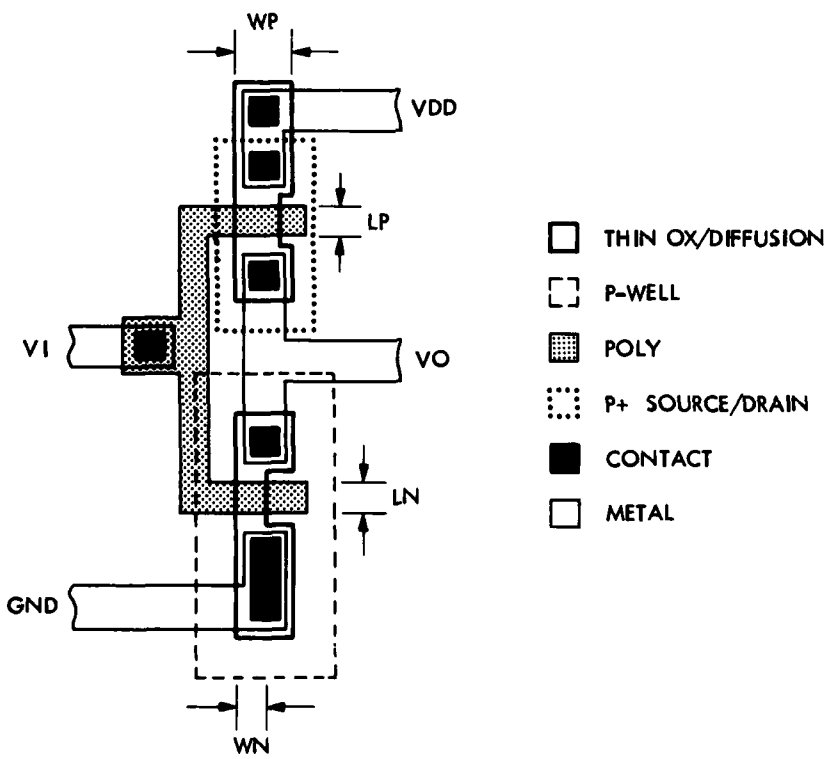


Figure 2.4.2-10. Layout of a CMOS-bulk inverter (INV), where WN = LN = LP = 5 micrometers and WP = 7.5 micrometers.

b. Inverter threshold voltage, VTINV:

1. Circuit Diagram: See Figure 2.4.2-12.
2. Test Conditions: VTINV is measured for VI = VO.

3. Data Reduction Algorithm: VTINV is determined directly from the measured value.

c. Inverter gain, G:

1. Circuit Diagram: See Figure 2.4.2-13.

2. Test Conditions: G is the magnitude of the slope of the inverter transfer curve measured at the inverter threshold voltage, VTINV. G is determined from a two-point method in which the inverter output voltage is measured at the inverter input voltages VTINV + DV and VTINV - DV, where DV = 25 millivolts.

3. Data Reduction Algorithm: The inverter gain is determined from:

$$G = [VO(VTINV - DV) - VO(VTINV + DV)]/2DV$$

The gain is expressed as a positive quantity.

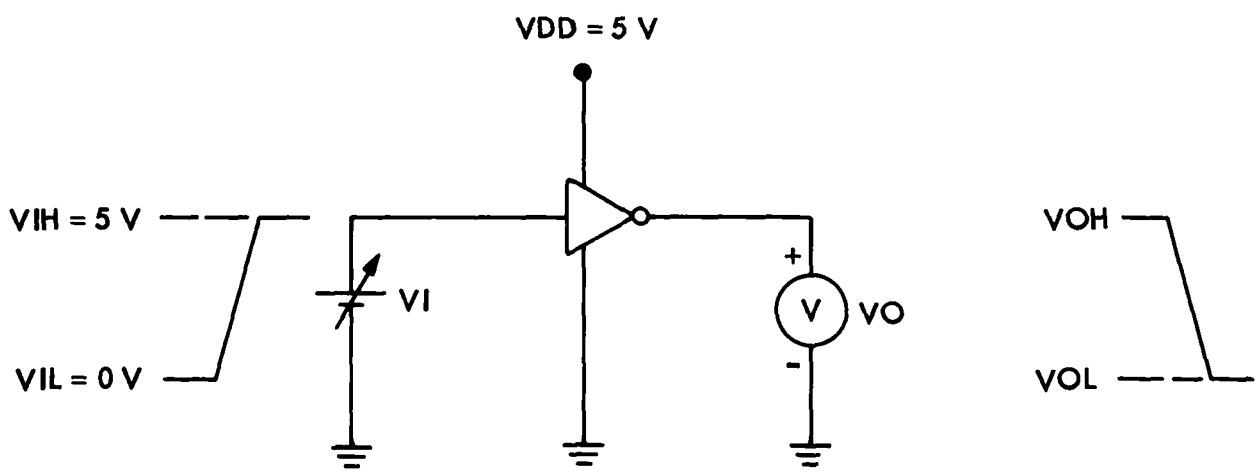


Figure 2.4.2-11. Inverter, VOH and VOL, measurement circuit.

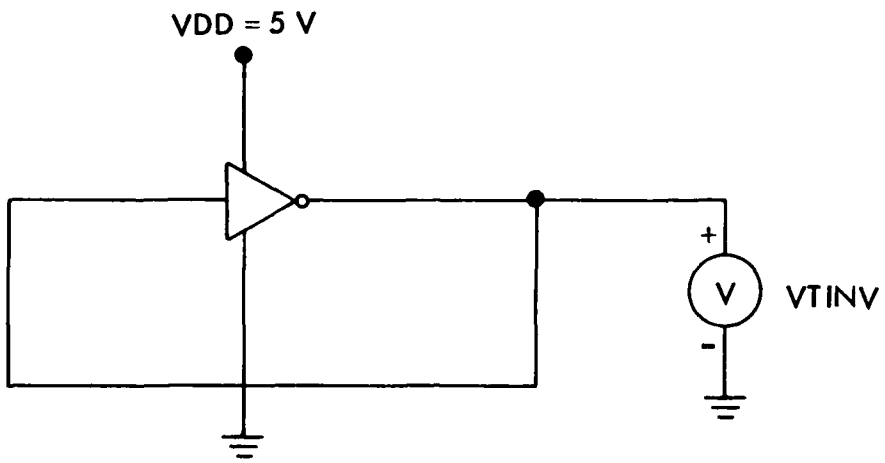


Figure 2.4.2-12. Inverter threshold voltage, VTINV, measurement circuit.

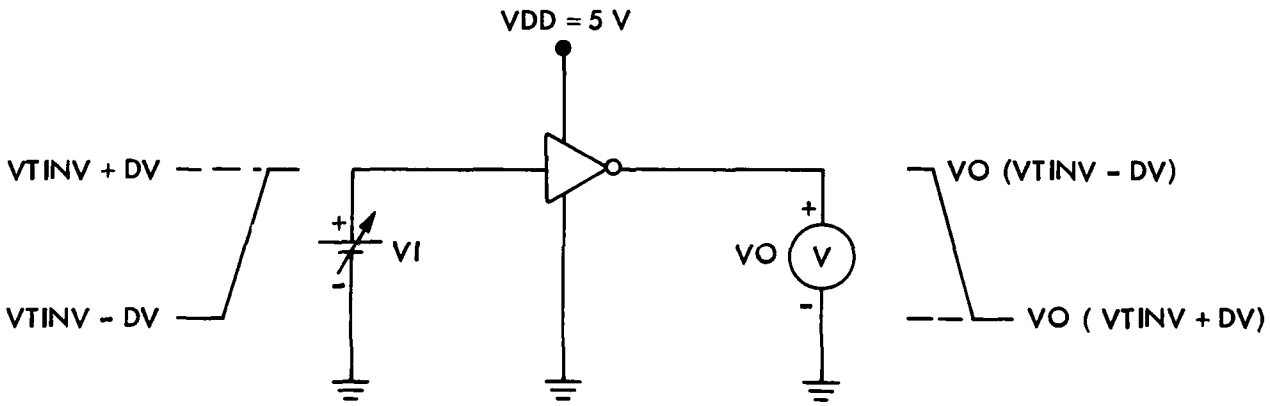


Figure 2.4.2-13. Inverter gain, G, measurement circuit.

2.4.2.4 Cross-Bridge Resistor

2.4.2.4.1 Test Module: Cross-bridge resistor.

2.4.2.4.2 Purpose: To evaluate the

a. Sheet resistance, RS

b. Linewidth, W, of metal, poly, and junction-isolated diffused layers

2.4.2.4.3 Geometrical Description: See Figure 2.4.2-14.

2.4.2.4.4 Test Procedures:

a. Sheet Resistance, RS:

1. Circuit Diagram: See Figure 2.4.2-15.

2. Test Conditions: RS is determined from the average of four resistance measurements. The resistances are determined by measuring two voltages between points 1 and 4 for current forced in both directions between points 2 and 5 and by measuring two additional voltages between points 2 and 1 for current forced in both directions between points 4 and 5.

If the magnitude of the current is the same for all four voltage measurements, then only the four measured voltages need be averaged. The magnitude of the current is adjusted so that the measured voltage is between 1 and 10 millivolts. This condition allows the voltage to be measured accurately with a digital voltmeter having a 1-microvolt resolution and prevents spurious current flow in junction-isolated layers due to junction forward biasing or junction breakdown.

3. Data Reduction Algorithm: RS is derived from the van der Pauw equation:

$$RS = (\pi/\ln 2)(\Sigma Vijkl/4I)$$

where

$$\Sigma Vijkl = V1425 - V1452 + V2154 - V2145$$

where i and j are the voltmeter connections and k and l are the current source connections. The positive terminal of the voltmeter is connected to i and the negative to j. Current is forced in both directions between k and l.

b. Linewidth, W:

1. Circuit Diagram: See Figure 2.4.2-16.

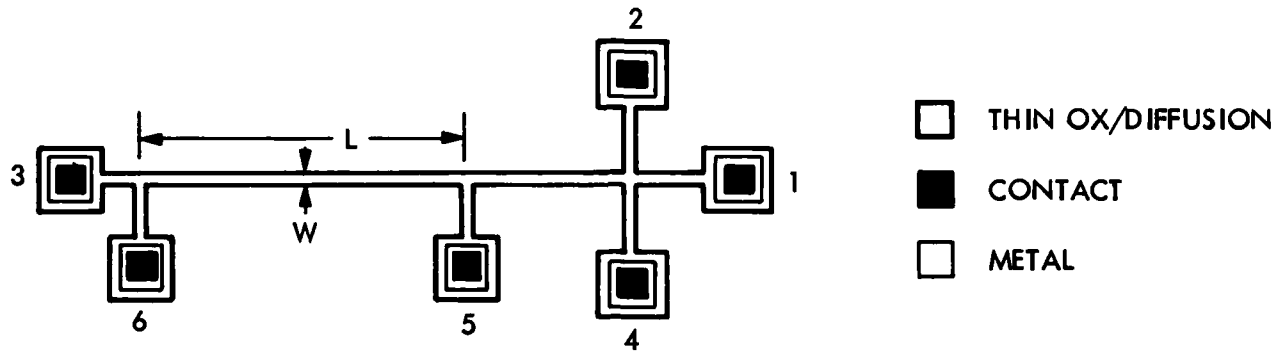


Figure 2.4.2-14. Layout of an n-diffused cross-bridge resistor (ND - XBR), where L = 160 micrometers and W = 5 micrometers.

2. Test Conditions: W is determined from the average of two resistance measurements. The resistances are determined by measuring two voltages between points 5 and 6 for current forced in both directions between points 2 and 3. If the magnitude of the current is the same for both voltage measurements, then only the two measured voltages need be averaged. The magnitude of the current is adjusted so that the measured voltage is between 1 and 10 millivolts.

3. Data Reduction Algorithm: W is derived from the following bridge equation:

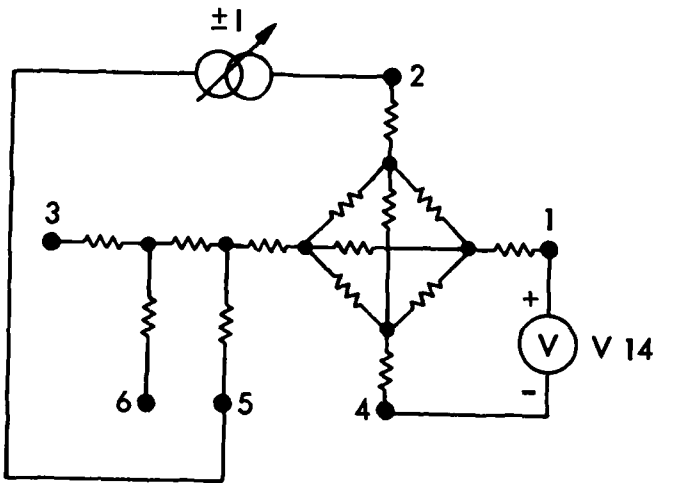
$$W = L * RS(2I/\Sigma Vijkl)$$

where

$$\Sigma Vijkl = V5623 - V5632$$

and where RS is obtained from the previous measurement and L is taken as the as-drawn dimension.

2.4.2.4.5 Reference: Buehler, M. G., S. D. Grant, and W. R. Thurber, "Bridge and van der Pauw Sheet Resistors for Characterizing the Linewidth of Conducting Layers," J. Electrochem. Soc., 125, 650-654 (1978).



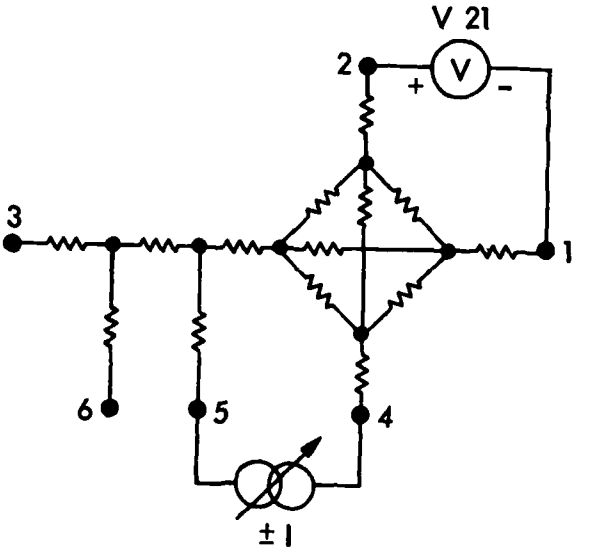


Figure 2.4.2-15. Sheet resistance, RS, measurement circuit.

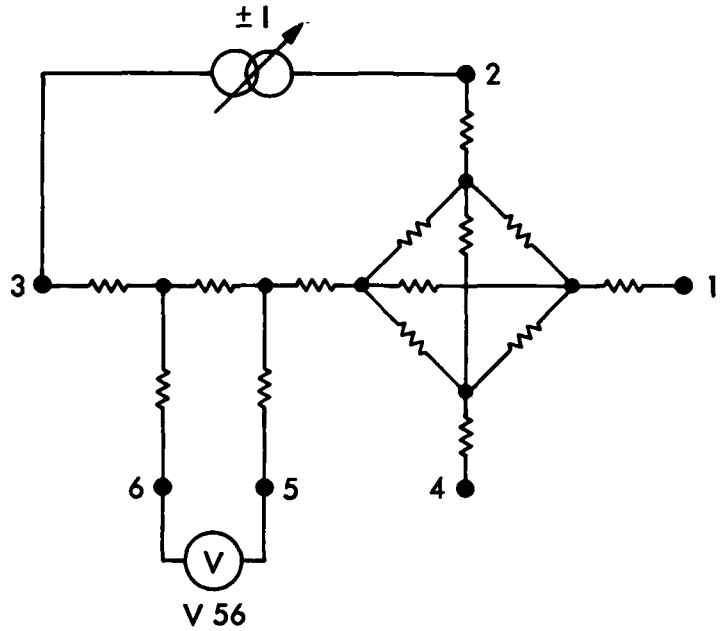


Figure 2.4.2-16. Linewidth, W, measurement circuit.

2.4.2.5 Contact Resistor

2.4.2.5.1 Test Module: Four-terminal contact resistor.

2.4.2.5.2 Purpose: To evaluate the contact resistance, RC, of metal to poly and metal to diffusion contacts.

2.4.2.5.3 Geometrical Description: See Figure 2.4.2-17.

2.4.2.5.4 Test Procedures:

a. Circuit Diagram: See Figure 2.4.2-18.

b. Test Conditions: RC is determined from the average of two resistance measurements. The resistances are determined by measuring two voltages between points 1 and 3 for current forced in both directions between points 2 and 4. If the current is the same in both the forward and reverse directions, then the resistance can be determined from the average of the voltage measurements. The magnitude of the current is adjusted so that the mea-

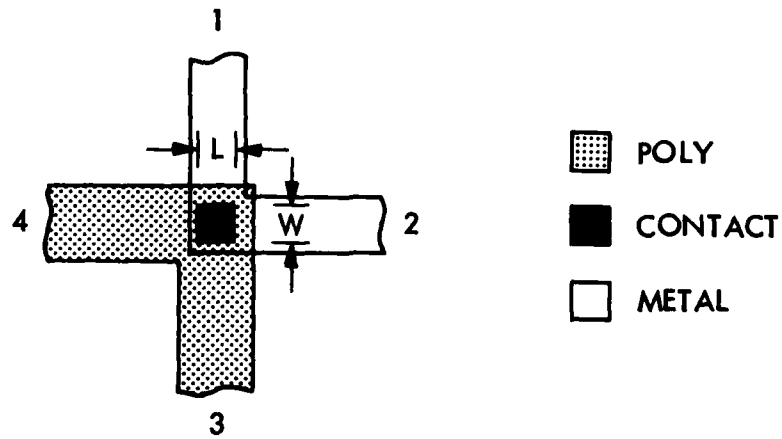


Figure 2.4.2-17. Layout of a metal n-poly contact resistor (M/NP - CR), where L = W = 5 micrometers.

sured voltage is between 1 and 10 millivolts, and this means that the contact resistance is measured at two values of current that differ by approximately an order of magnitude.

c. Data Reduction Algorithm: RC is derived from the following equation:

$$RC = \Sigma Vijkl/2I$$

where

$$\Sigma Vijkl = V1324 - V1342$$

where i and j are the voltmeter connections and k and l are the current source connections. The voltmeter's positive terminal is connected to i and the negative to j. Current is forced in both directions between k and l.

2.4.2.5.5 Reference: Proctor, S. S., and L. W. Linholm, "A Direct Measurement of Interfacial Contact Resistance," IEEE Electron Devices Letters, EDL-3, 294-296 (1982).

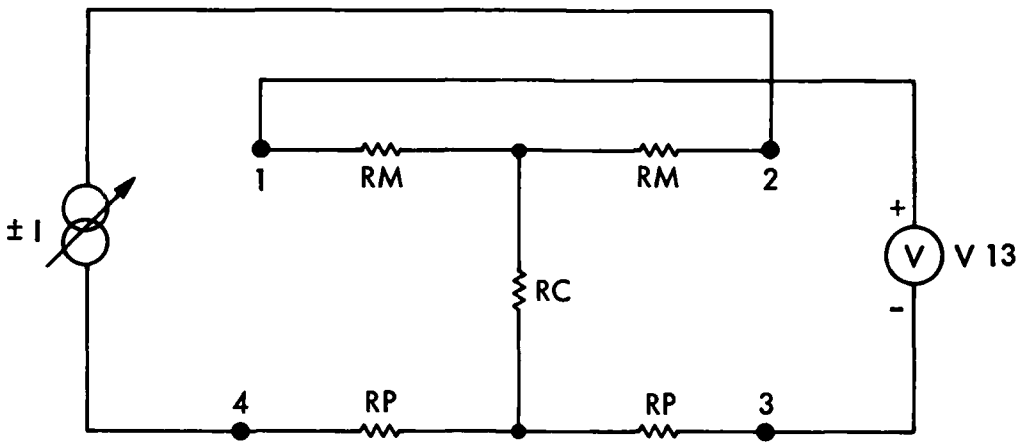


Figure 2.4.2-18. Contact resistance, RC, measurement circuit, where RM is the resistance of the metal leads and RP is the resistance of the poly (or diffusion) leads.

2.4.3 Split-Cross-Bridge Resistor

This test structure was developed to measure the sheet resistance, linewidth, line spacing, and line pitch of polycrystalline silicon, metal, and diffused source and drain layers where the linewidths or spaces exceed several micrometers. A detailed description for the design and measurement of this structure is included in this section. This concept is being extended to measure the linewidth of p-well layers and MOSFET channels.

The Split-Cross-Bridge Resistor for Measuring the Sheet Resistance, Linewidth, and Line Spacing of Conducting Layers*

Martin G. Buehler

Jet Propulsion Laboratory

California Institute of Technology

Pasadena, California 91109

ABSTRACT

A new test structure was developed for evaluating the line spacing between conductors on the same layer by using an electrical measurement technique. This compact structure can also be used to measure the sheet resistance, linewidth, and line pitch of the conducting layer. Using an integrated circuit fabrication process, this structure was fabricated in diffused, polycrystalline silicon, and metal layers. Test results confirm the structure's self-checking feature based on the line pitch. That is, a small difference between the measured and designed line pitch is used to validate sheet resistance, linewidth, and line spacing values. Rules for designing the test structure are presented in detail.

1. INTRODUCTION

This paper describes an approach to measuring the spacing between electrically conducting lines. The approach is an extension of cross-bridge resistor [1] linewidth measurements and is based on an electrical measurement technique in which the line spacing is determined from a specially designed cross-bridge resistor, termed the split-cross-bridge resistor. The need for such a measurement follows from the need to evaluate integrated circuit layout rules quickly and accurately. Alternative techniques for measuring line spacing are based on visual measurement techniques [2], which are time-consuming. The technique is illustrated by three test structures for determining the line spacing between metal layers, polycrystalline silicon layers, and diffused silicon layers. These layers are illustrated by the conducting layer shown in Figure 1. These structures were designed and fabricated using the layout rules for the linewidths and spacings listed in Table I.

---

*This effort was sponsored jointly by the National Aeronautics and Space Administration and the Defense Advanced Research Projects Agency.

To illustrate the measurement principle, consider the metal, split-cross-bridge resistor shown in Figure 1, which was designed to evaluate the line spacing between two metal lines. The split-cross-bridge resistor is a combination of three structures. The upper structure is a cross resistor, the middle structure is a bridge resistor, and the lower structure is a split-bridge resistor. The bridge resistor has a single conducting channel of width, $W_b = 2W + S$. The split-bridge resistor has two conducting channels, each with a width of W so that the effective width is $W_s = 2W$. The line spacing, S, is determined by subtracting the width of the split-bridge resistor ($W_s = 2W$) from the width of the bridge resistor ($W_b = 2W + S$).

2. MEASUREMENT TECHNIQUE

The technique requires three electrical measurements, which are illustrated by the following idealized relationships. These relationships are uncorrected for geometrical errors, which, as discussed in a later section, can be ignored if the structure is designed properly. In the analysis, the sheet resistance, R_s, of the layer is needed, and this is calculated from the cross resistor using the van der Pauw [3] equation:

$$R_s = (V_c/I_c)\ (\pi/\ln 2) \tag{1}$$

where the potential difference, V_c, is $V_1 - V_2$ for a current, I_c, passed into I_1 and out of I_2.

The width of the bridge resistor is determined from the idealized rectangular resistor expression:

$$W_b = 2W + S = R_s L_b I_b / V_b \tag{2}$$

where the potential difference, V_b, is $V_2 - V_3$ for a current, I_b, passed into I_1 and out of I_3. The distance between the voltage taps is L_b, which is the distance specified on the photomask. The sheet resistance, R_s, is determined from the cross resistor and Eq. (1). The width of the split-bridge resistor is determined from:

$$W_s = 2W = R_s L_s I_s / V_s \tag{3}$$

where the potential difference, V_s, is $V_4 - V_5$ for a current, I_s, passed into I_1 and out of I_3. The distance between the voltage taps is L_s. Note that the above equations require that no significant magnification occur so that L_b and L_s can be taken as their design values.

The general expression for line spacing is:

$$S = W_b - W_s = R_s(L_b I_b V_s - L_s I_s V_b)/(V_b V_s) \quad (4)$$

where the sheet resistance, R_s, is determined from the cross resistor and Eq. (1).

For $I_b = I_s = I$ and $L_b = L_s = L$, then:

$$S = R_s LI(V_s - V_b)/V_b V_s) \quad (5)$$

The linewidth is calculated by modifying Eq. (3) and assuming $I_s = I$ and $L_s = L$:

$$W = R_s LI/(2V_s)(6)$$

Finally the line pitch, P, between features is determined from:

$$P = W + S = W_b - (W_s/2) = R_s LI(2V_s - V_b)/(2V_b V_s) \quad (7)$$

Thus the split-cross-bridge resistor can be used to evaluate four critical parameters of a conducting layer, i.e., sheet resistance, linewidth, line spacing, and line pitch.

3. MEASUREMENT INTERPRETATION

A physical model was developed that compares the linewidths and spacings of photomasks or wafers features measured by either visual or electrical techniques. The size of a feature may expand or contract linearly due to a number of factors, such as the bloating and shrinking of features during photomask making or lateral etching, lateral diffusion, and the coating of features by deposited layers during wafer fabrication. The model requires the use of the following eight parameters:

W_p, S_p = linewidth and spacing of a feature observed visually on a photomask.

W_v, S_v = linewidth and spacing of a feature observed visually on a fabricated wafer.

W_e, S_e = linewidth and spacing of a layer measured electrically on a fabricated wafer.

V = difference in the location of a feature edge as determined from visual measurements made on photomasks and fabricated wafers.

X = difference in the location of a feature edge as determined from visual and electrical measurements on fabricated wafers.

The relation among these parameters is shown in Figures 2 and 3. Figure 2 illustrates the fabrication of a polycrystalline silicon (poly) layer where the width of the poly layer observed on the wafer (W_e or W_v) is smaller than the width (W_p) on the photomask. The visual width, W_v, is shown arbitrarily at the base of the oxide, which surrounds the poly layer. The electrical width, W_e, is shown at the midpoint in the length of the trapezoid describing the poly layer.

Figure 3 illustrates the formation of a diffused layer by either (a) a uniform oxidation, diffusion, oxidation process or by (b) a local oxidation, diffusion, oxidation process. As illustrated in Figure 3, the electrical width, W_e, includes not only the planar portion of the diffusion but a fraction of the lateral diffusion region as well.

The mathematical relationships among the eight parameters are illustrated in Table II. Note that the difference between the upper equations for metal and poly structures and the lower equations for diffused structures is the sign reversal of the V and X terms. The measured quantities W_b and W_s are shown, along with three derived quantities: $W_b - W_s$ (line spacing), $W_s/2$ (linewidth), and $W_b - W_s/2$ (line pitch). As can be seen in the table, W_e is larger or smaller than W_v or W_p depending on the nature of the layer being formed. For example, for a diffused layer, W_e is larger than W_p by $2V + 2X$. Similar comments apply to the line spacing. Such discrepancies are signficant in manufacturing processes.

The line pitch, $W_b - (W_s/2)$, listed in Table II deserves special discussion. From the table it can be seen that $W_e + S_e = W_v + S_v = W_p + S_p$. This means that the line pitch measured electrically is a direct measure of the features formed on the photomask. That is, $W_e + S_e$ are not affected by V and X because they are distances between features that have the same kind of edge. For example, $W_e + S_e$ is the distance between the left edges of two parallel lines, and this distance is not affected by linear changes in photomask features (bloating and shrinking) or by wafer fabrication processes, because both edges are affected in identical fashion, provided features are not magnified. If magnification can be ruled out, then $W_e + S_e$ can be used to verify that the design values for W and S were correctly implemented on the photomask and were measured correctly by the parametric test equipment. If magnification (or a change in pitch) occurs, then the design value cannot be used for L_b and L_s, and the distance between the voltage taps must be measured visually.

4. DESIGN CONSIDERATIONS

The design of the split-cross-bridge resistor follows from four geometrical design rules. If these rules are followed, then the equations given in Section 2 can be used directly to obtain results that are accurate within one percent.

The split-cross-bridge resistor was laid out so that it could be probed with a 2 by N probe array [4] where N is an arbitrary positive integer number. As illustrated in Figure 1, N = 4, and the probe pads are 80-micrometer squares separated by 80 micrometers. The distances L_b and L_s were designed to be the same length (135 micrometers) so that they can be easily verified using visual inspection techniques. The reason for specifying $L_b = L_s = 135$ micrometers is discussed below.

The cross resistor is constructed from two equal-width rectangles that intersect at right angles. Design rule #1 requires that the length, A, of each arm of the cross be at least twice the arm width, W, in order for the sheet resistance, R_s, to be calculated accurately from the idealized sheet resistance expression, Eq. (1). This rule follows from a detailed analysis of this structure [5]; the result of the analysis is shown in Figure 4. The R_s as calculated from Eq. (1) is slightly less than the true sheet resistance, R_{st}, as given by:

$$R_s = R_{st} \ (1 - E_1) \tag{8}$$

where E_1 is the geometrical error. By extrapolating the curve in Figure 4, the error for $A > 2W$ is seen to be negligible.

Design rule #1 requires that $A > 2W_b$ for the cross and bridge portion of the structure. The $A > 2W_b$ (rather than $A > W_b$) rule was chosen to include the design of diffused structures where lateral diffusion is a factor. If a structure has an excessively large lateral diffusion (such as a p-well in a CMOS process), the designer may have to calculate a suitable A/W_b ratio to minimize errors.

The following discussion applies to both the bridge resistor and the split-bridge resistor. Bridge resistors are constructed from a conducting channel that is tapped in two places along the length of the channel. There are three design rules to be considered. They include the distance between and width of the taps, the length of the taps, and the location of the taps relative to a change in the channel width.

Design rule #2 requires that the width, D, of the voltage taps be designed at the minimum width allowed and that the distance, L, between the taps be large enough so that the channel width can be calculated accurately from the uncorrected rectangular resistor expression ($W = R_sL/R$). This rule follows from a detailed analysis of the bridge resistor shown in Figure 5.

This analysis is based upon the results of conformal mapping described by Hall [6] where the length of the tap is much larger than its width. His result, Eq. (48), indicates that the resistance of a bridge between the voltage taps is given by:

$$R = R_{st}(L/W)\ (1 - E_2) \tag{9}$$

where the geometrical error, E_2, is:

$$E_2 = (2W/\pi L)[(D/W)\ \tan^{-1}\ (D/2W) - \ln((D^2/4W^2) + 1)] \tag{10a}$$

and the uncorrected channel width is:

$$W = (R_{st}/R)L \tag{10b}$$

L is the distance between the voltage taps, W is the width of the conducting channel, and D is the width of the tap. From this expression:

$$W = W_t/(1 - E_2) \tag{11}$$

This indicates that W, as calculated from the idealized rectangular resistor equation, Eq. (10b), is slightly larger than the true width, W_t.

The distance between and the width of the voltage taps of the bridge resistors were chosen to reduce the geometrical error to an acceptable amount. For the structures included in this study, the width of the voltage taps was taken as the minimum width allowed and L = 135 micrometers. The errors for E_2 are listed in Table I, where E_{2b} is the error for the bridge resistor and E_{2s} is the error for the split-bridge resistor. The errors are worse for the metal layer, but even for this worst case the error is less than one percent.

Design rule #3 requires that the length, G, of a tap from the current-carrying channel be at least twice the width, D, of the tap for the channel width, W, to be calculated accurately from the uncorrected rectangular resistor expression, ($W = R_sL/R$). This rule follows from a resistor model developed for the bridge as shown in Figure 6. Each major area of the resistor is assigned an equivalent resistor that is calculated from the rectangular resistor equation. The objective of the model is to determine the parameter "g," which indicates how much of the tap is effective in carrying current. The resistor model leads to:

$$R = R_{st}\ [(L/W) - E_3] \tag{12}$$

where:

$$E_3 = (D/W)/[1 + (W/gD)] \tag{13}$$

The model tap length is G* where G* = gD.

The parameter "g" was evaluated by finding a value for g that provides a good fit between a plot of E_3 versus D/W and an analytical expression derived from Hall's [6] Eq. (48):

$$E_4 = (L/W)E_2 = (2/\pi)[(D/W)\tan^{-1}(D/2W) - \ln((D^2/4W^2) + 1)] \quad (14)$$

This equation is applicable to the case where the tap is much longer than its width and was derived from conformal mapping theory. Eq. (13) is a very good approximation to Eq. (14) for g = 0.18. This result indicates that only a small fraction of the tap serves to shunt the current from the main channel. The design rule requires that the design tap length, G, be greater than or equal to 2D. This means that the design requirement for G/D is 11.1 times larger than required by the model ($G^*/D = g = 0.18$), which should minimize errors due to terminating the taps.

Design rule #4 requires that the edge of a voltage tap be located twice the channel width from a change in the channel width for the channel width, W, to be calculated accurately from the uncorrected rectangular resistor expression, ($W = R_sL/R$). As applied to the split-cross-bridge shown in Figure 1, this rule allows one to establish the distances H_b and H_s shown in the figure. This rule follows from a detailed analysis of current flow past a discontinuity in the width of the channel as shown in Figure 7. The analytical expression [6] for the current density, J(x), along the x-axis is:

$$J(x) = (1 + t)^{1/2}/(1 + f^2t)^{1/2} \quad (15)$$

where f is the ratio of the width of the smaller channel to the width of the larger channel. The distance along the x-axis is:

$$x = \frac{1}{\pi}\left[\ln\left(\frac{(1+t)^{1/2} + (1+f^2t)^{1/2}}{(1+t)^{1/2} - (1+f^2t)^{1/2}} - f\ln\frac{(1+f^2t)^{1/2} + f(1+t)^{1/2}}{(1+f^2t)^{1/2} - f(1+t)^{1/2}}\right)\right] \quad (16)$$

where the parameters, t and v, that link the above equations are given by:

$$t = \exp(-2v) \quad (17)$$

These expressions were derived from Hall's [6] Eqs. (40) and (41) using the transformation $2\tanh^{-1}z = \ln[(1 + z)/(1 - z)]$.

The results shown in Figure 7 indicate that the current density settles out in a distance that is less than twice the channel width from the discontinuity. This is true for both the small and the large channels. Notice that design rule #4 is conservative with respect to the large channel. That is, the W shown in Figure 7 is one-half the width, W_b, of the bridge resistor.

5. EXPERIMENTAL RESULTS

Measurement procedures for evaluating the sheet resistance and the linewidth are detailed elsewhere [1, 3, 7]. These procedures require the bridge voltages be measured for current flowing in both directions along the channel.

This should eliminate errors introduced by voltage offsets due to instrumentation errors and thermal voltages at switch relays. The elimination of such errors assumes that the magnitude of the current is the same in both directions. Also, the sheet resistance as determined from the cross resistor should result from the average of four resistance measurements. One measurement requires currents to be forced in both directions between points I_1 and I_2 and voltages measured between V_1 and V_2. The other measurement requires currents to be forced in both directions between points I_1 and V_1 and voltages measured between V_2 and I_2 [3, 7]. A minimum of eight resistance values is required to measure R_s, W, S, and P from the split-cross-bridge resistor. Additional measurements may be required to assure that the resistor is linear and that Joule heating is not a factor. Four different test structures were fabricated using a conventional NMOS process. The structures were fabricated in the metal layer, the n-source/drain layer, the poly layer on field oxide, and the poly layer on gate oxide. The structures were arranged on a test chip. Eight of these test chips were positioned randomly about a 100-millimeter diameter silicon wafer. The structures were measured using an automatic data acquisition system with a computer-controlled wafer-prober. The magnitude of the currents forced through the structure was adjusted so that the measured voltages were approximately 10 millivolts, except for the metal structure, where the smallest voltages were about 0.3 millivolts. The 10-millivolt limit was chosen to be large enough for easy, high-accuracy measurements but small enough so that junctions in diffused structures are not forward-biased. The 0.3-millivolt limit was chosen so that Joule heating in metal structures is minimized.

Results of the electrical measurements are listed in Table III. As seen in Table III, the measured linewidth and line spacing values deviate significantly from the design values W_o and S_o. This is especially true for the metal layer. Contrary to expectations, the diffusion linewidth is smaller than its design value and the poly linewidth is larger than its design value. These observations are most likely explained by bloat and shrink factors applied at the time the photomasks were generated to compensate for the etching of the diffusion and poly features. In spite of the deviation of the measured values from the design values, the measured line pitch values, P, were within one percent of the design line pitch values, $P_o = W_o + S_o$, for most cases, and all measured line pitch values were within four percent of design values. If the poly (thick oxide) data is ruled out, then all P values were within two percent of P_o. This verifies that P values can be used to identify faulty R_s, W, and S values. On other wafers, $P - P_o$ values were useful in flagging bad data.

6. DISCUSSION

In order to calculate R_s, S, W, and P from Eqs. (1), (5), (6), and (7), the split-cross-bridge resistor shown in Figure 1 must be designed according to the following layout rules. These rules are listed in Table IV and are reviewed below.

RULE #1: The length, A, of the arm of each cross must be at least twice the arm width, W_b, ($A > 2W_b$).

RULE #2: The width, D, of the bridge resistor's voltage taps be designed at the minimum width allowed and the distance, L_b and L_s, between the voltage taps must be large enough so as to minimize error E_2; see Figure 5.

RULE #3: The length, G, of the voltage taps from the current carrying channel must be at least twice the width, D, of the voltage taps ($G > 2D$).

RULE #4: The distance, H_b, and H_s, from the edge of a voltage tap to a change in the channel width must be at least twice the width of the channel ($H_b > 2W_b$, $H_s > 2W_s$).

The experimental results obtained for this structure are meant to demonstrate the "reduction to practice" of this structure. The results are not meant to assure the user that the structure will provide valid measurements for all possible linewidths and spaces. Some workers have suggested that small lines will be overetched more than large lines. If the etching is nonlinear in linewidth, the measured line pitch will not equal the design line pitch. This inequality will alert the user to the discrepancy. Additional experimental studies are needed to confirm the usefulness of this new test structure in measuring lines in the 1-micrometer and smaller range.

7. CONCLUSION

The split-cross-bridge resistor is a new compact test structure for measuring the sheet resistance, linewidth, line spacing, and line pitch between various kinds of conducting lines using electrical measurement techniques. A model was developed to relate the electrical measurements of linewidth and spacing to visual measurements. For the linewidth and spacing measurements, the electrical and visual techniques give different results due to lateral diffusions and the coating of layers. For diffused layers the optical linewidth is smaller than the electrical linewidth. For a coated layer, the optical linewidth is larger than the electrical linewidth by about twice the thickness of the coating. For pitch measurements the visual and electrical techniques measure the same quantity, provided features have not been magnified during the photomask and wafer fabrication processes. If magnification can be ruled out, then the electrical measurement of line pitch can be used to assure that R_s, W, and S values were measured correctly. Three kinds of test structures were designed to measure the sheet resistance, linewidth, line spacing, and line pitch of metal, poly, and diffused layers. The measured line pitch values were within one percent of the design values for most cases, and all values were within four percent of the design values.

ACKNOWLEDGEMENT

Several staff members assisted in the development of this project. R. Nixon laid out the test structures, and B. Blaes wrote the test programs and acquired the data. Technical discussions with T. Griswold of JPL, D. Perloff of Signetics, L. Linholm of NBS, and P. Losleben of DARPA, and the wafer fabrication services of the Information Sciences Institute are gratefully acknowledged.

REFERENCES

1. Buehler, M. G., S. D. Grant, and W. R. Thurber, "Bridge and van der Pauw Sheet Resistors for Characterizing the Linewidth of Conducting Layers," J. Electrochem. Soc., 125 (4), 650-654 (April 1978).

2. Nyyssonen, D., "Calibration of Optical Systems for Linewidth Measurements on Wafers," Proceedings of the Society of Photo-optical Instrumentation Engineers--Semiconductor Microlithography V, 221, 119-126 (1980).

3. van der Pauw, L. J., Philips Res. Rep., 13, 1 (1958); van der Pauw, L. J., Philips Tech. Rev., 20, 220 (1958).

4. Buehler, M. G., "Comprehensive Test Patterns with Modular Test Structures: The 2 by N Probe-Pad Array Approach," Solid-St. Technol., 89-94 (October 1979).

5. David, J. M., and M. G. Buehler, "A Numerical Analysis of Various Cross Sheet Resistor Test Structures," Solid-State Electronics, 20, 539-543 (1977).

6. Hall, P. M., "Resistance Calculations for Thin Film Patterns," Thin Solid Films, 1, 277-295 (1967/68).

7. Buehler, M. G., and W. R. Thurber, "An Experimental Study of Various Cross Sheet Resistor Test Structures," J. Electrochem. Soc., 125 (4), 645-650 (April 1978).

Table I. Split-Cross-Bridge Resistor Layout Rules and Voltage Tap Geometrical Errors (E_2)

Layer	Layout Rule Linewidth (μm)	Layout Rule Line Spacing (μm)	$\frac{W_b}{L_b}$	$\frac{D}{W_b}$	E_{2b}	$\frac{W}{L_s}$	$\frac{D}{W}$	E_{2s}
Metal	7.5	7.5	0.167	0.333	0.003	0.055	1.0	0.008
Polycrystalline Silicon	5.0	5.0	0.111	0.333	0.002	0.037	1.0	0.006
Diffusion (Thin Oxide)	5.0	7.5	0.130	0.286	0.002	0.037	1.0	0.006

$L_b = L_s = 135$ μm

Table II. Comparison of Linewidths and Spacings

Measured/Derived	Electrical	Visual	Photomask
Metal/Poly			
W_b	$= 2W_e + S_e$	$= 2\ W_v + S_v - 2X$	$= 2\ W_p + S_p - 2V - 2X$
W_s	$= 2W_e$	$= 2\ (W_v - 2X)$	$= 2\ (W_p - 2V - 2X)$
$W_b - W_s$	$= S_e$	$= S_v + 2X$	$= S_p + 2V + 2X$
$W_s/2$	$= W_e$	$= W_v - 2X$	$= W_p - 2V - 2X$
$W_b - (W_s/2)$	$= W_e + S_e$	$= W_v + S_v$	$= W_p + S_p$
Diffusion			
W_b	$= 2W_e + S_e$	$= 2W_v + S_v + 2X$	$= 2\ W_p + S_p + 2V + 2X$
W_s	$= 2W_e$	$= 2\ (W_v + 2X)$	$= 2\ (W_p + 2V + 2X)$
$W_b - W_s$	$= S_e$	$= S_v - 2X$	$= S_p - 2V - 2X$
$W_s/2$	$= W_e$	$= W_v + 2X$	$= W_p + 2V + 2X$
$W_b - (W_s/2)$	$= W_e + S_e$	$= W_v + S_v$	$= W_p + S_p$

Table III. Experimental Results from Several Split-Cross-Bridge Resistors

Layer	R_s Ω/□	$\bar{W}_e \pm W_{e\sigma}$ μm	W_o μm	$\bar{S}_e \pm S_{e\sigma}$ μm	S_o μm	$\bar{P}_e \pm P_{e\sigma}$ μm	$[(\bar{P}_e - P_o) \pm P_{e\sigma}]/P_o$ %
Metal	$(2.32 \pm 0.02) \times 10^{-2}$	5.52 ± 0.15	7.5	9.39 ± 0.16	7.5	14.90 ± 0.04	-0.65 ± 0.27
N-Diffusion	8.67 ± 0.12	4.54 ± 0.17	5.0	8.02 ± 0.13	7.5	12.57 ± 0.04	0.52 ± 0.34
Poly (thin ox)	23.34 ± 0.39	5.66 ± 0.07	5.0	4.39 ± 0.11	5.0	10.05 ± 0.10	0.50 ± 1.00
Poly (thick ox)	23.42 ± 0.39	5.96 ± 0.09	5.0	4.27 ± 0.09	5.0	10.22 ± 0.09	2.22 ± 0.85

$P_o = W_o + S_o$, $\bar{P}_e = \bar{W}_e + \bar{S}_e$

Table IV. Split-Cross-Bridge Resistor Layout Rules

Rule	Requirement
1	$A \geq 2W_b$
2	D = minimum linewidth
	L_b and L_s long enough to minimize E_2
3	$G \geq 2D$
4	$H_b \geq 2W_b$
	$H_s \geq 2W$

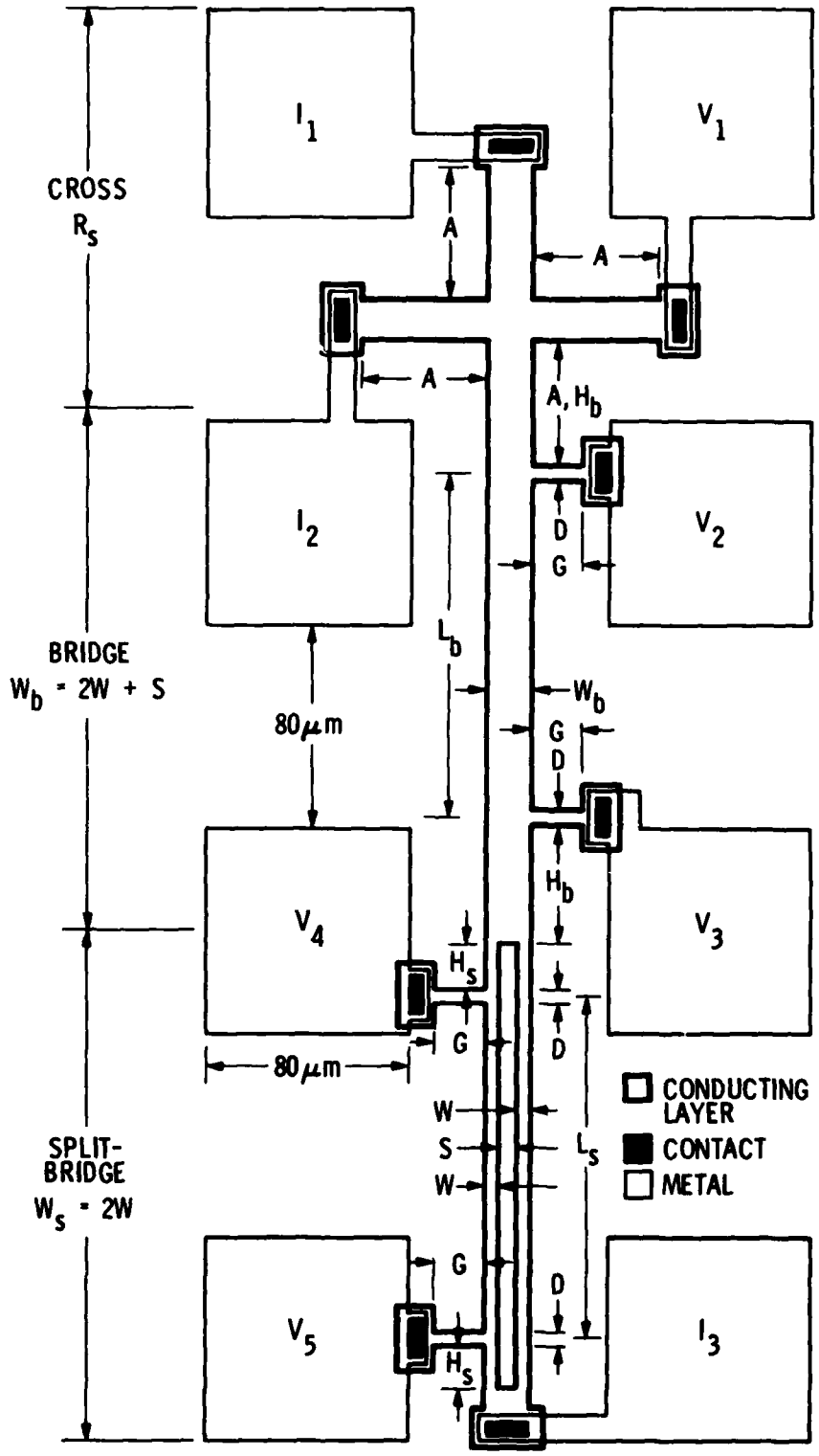


Figure 1. Split-cross-bridge resistor designed to evaluate the linewidth, W, line spacing, S, and sheet resistance of a conducting layer.

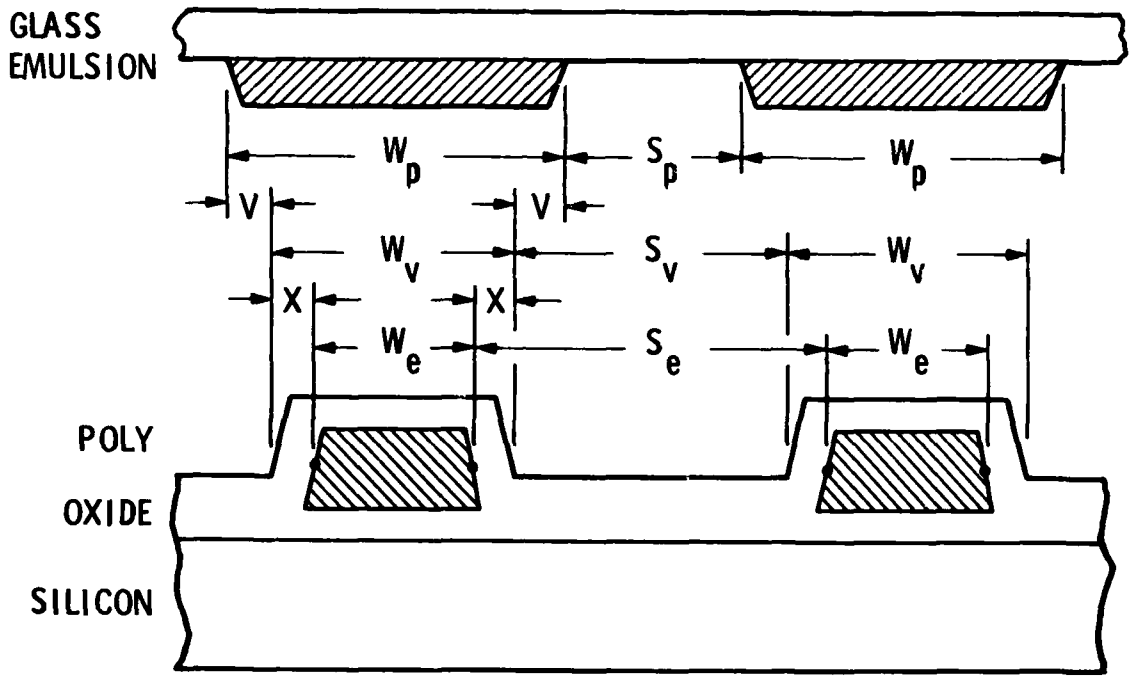


Figure 2. Cross-sectional diagrams of a polycrystalline (poly) silicon layer and a photomask where critical dimensions are indicated.

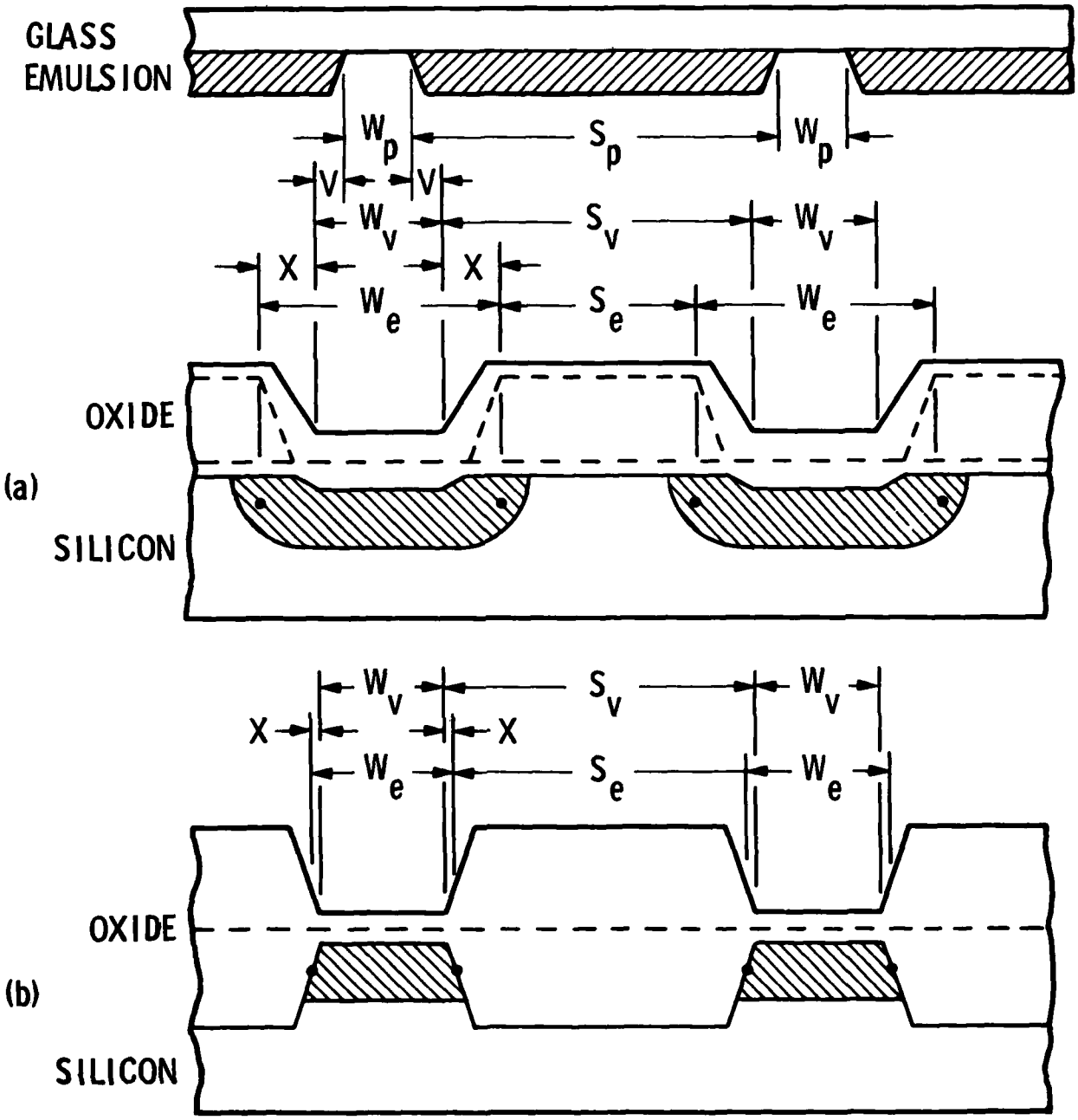


Figure 3. Cross-sectional diagrams of two diffused layers formed (a) by a uniform oxidation, diffusion, oxidation process and (b) by a local oxidation, diffusion, oxidation process where critical dimensions are indicated. Intermediate surfaces are shown by dashed lines.

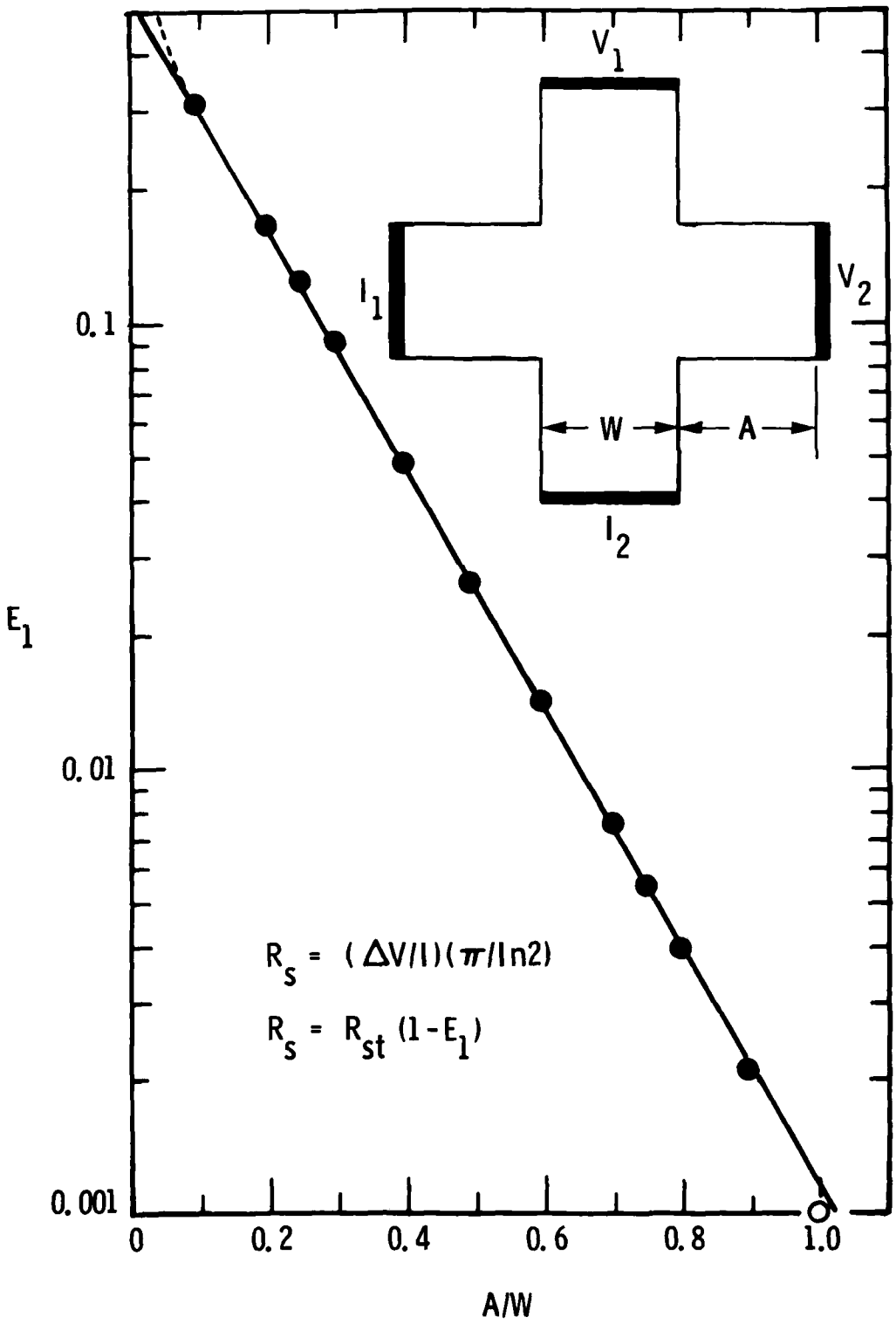


Figure 4. The error in measuring the sheet resistance from the Greek cross due to short arms. Keeping A ≥ 2W eliminates the geometrical error in calculating the sheet resistance from the van der Pauw equation, Eq. (1) [5].

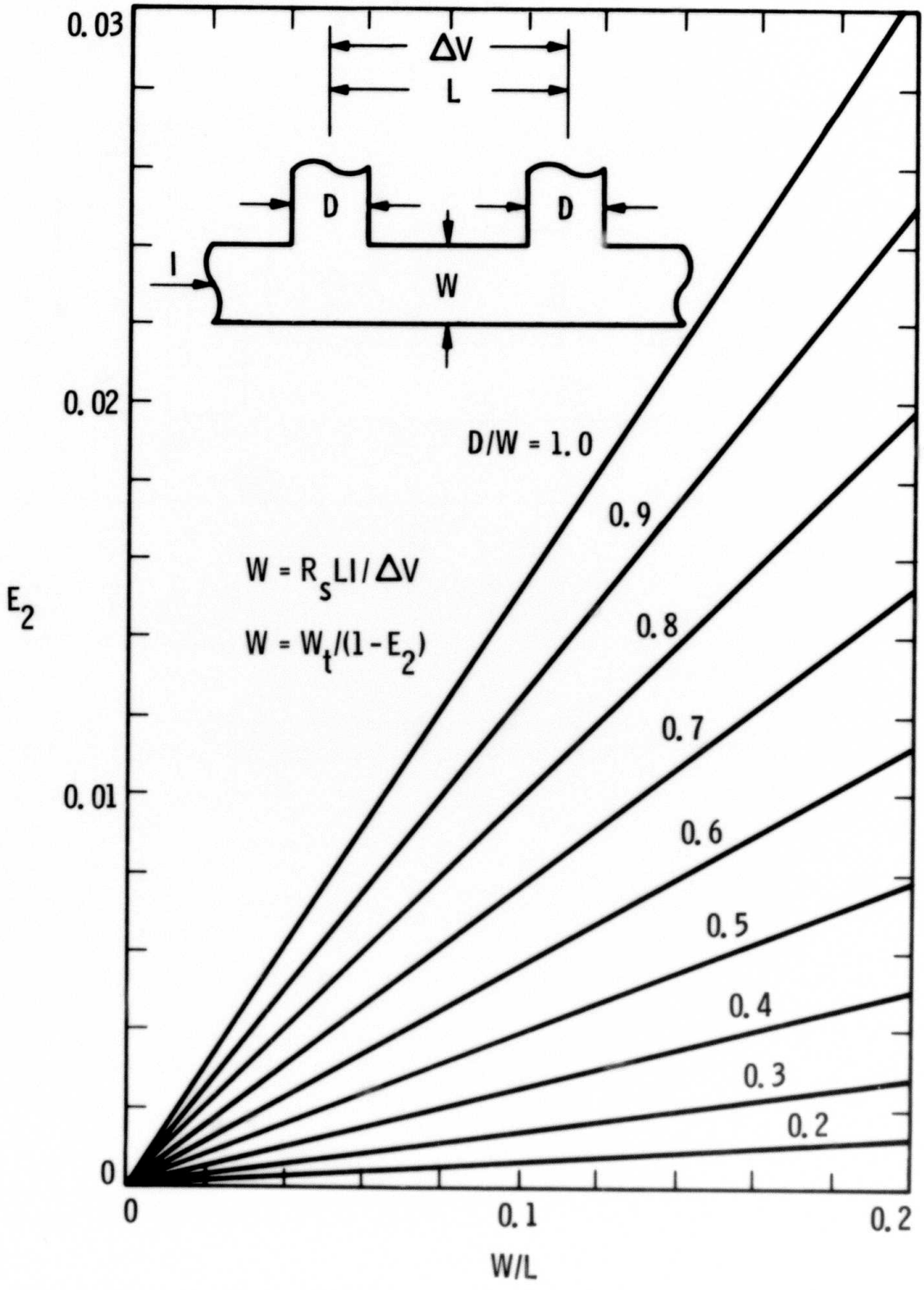


Figure 5. The error in measuring the linewidth from a bridge structure due to perturbations in the channel current flow at the voltage taps.

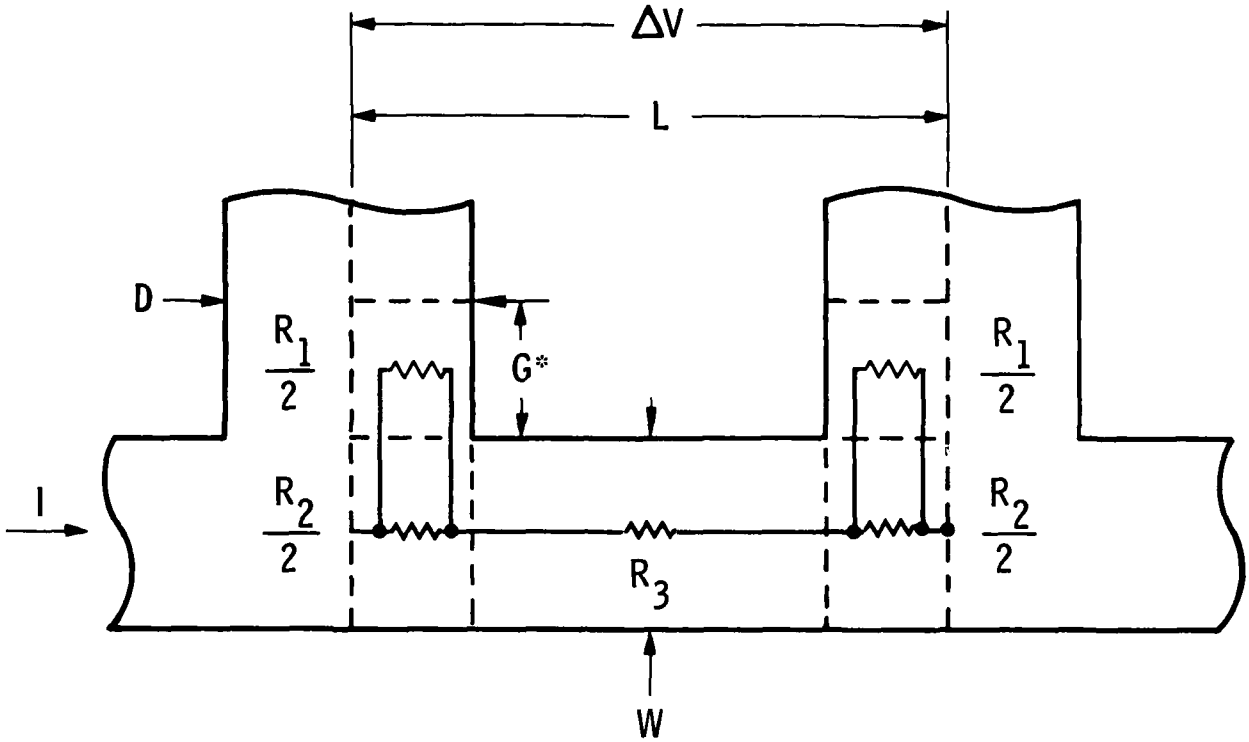


Figure 6. Resistor model of the bridge structure used to estimate the region of the tap that handles significant current flow.

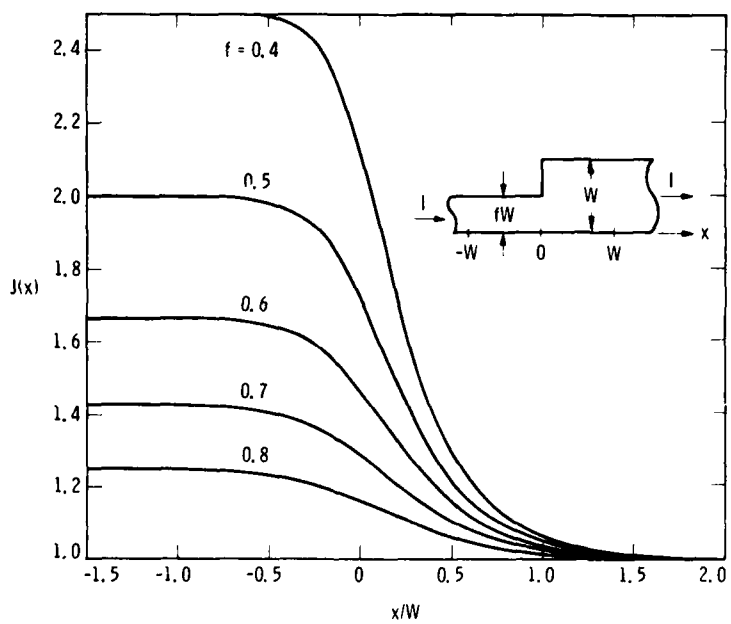


Figure 7. Current density along the bottom edge of a conducting layer that abruptly changes width. Keeping voltage taps a distance that is twice the width of the channel from the discontinuity, minimizes this error in calculating the linewidth from Eq. (6).

2.4.4 Yield Analysis Test Structures

Four types of yield analysis test structures were designed for the random fault test chips shown in Figures 2.3.2-3 and 2.3.2-6.

1. CAR (contact array resistor), N = 2,752
2. COR (comb resistor)
3. SCR (step-coverage resistor), N = 19,064
4. PAC (pinhole array capacitor), N = 7,616

where N is the number of elements (i.e., contacts, steps, or capacitors) in the largest subarray. Each of these structures is discussed in the following section.

The overall thrust of this methodology is based on the Poisson yield expression:

$$Y = \exp(-N/E)$$

where E is the number of elements per fault. Each test structure was divided into a number of subarrays, each with a different number of elements in each subarray. Each test structure is repeated a number of times across the wafer. From the yield for each subarray, the yield equation, given above, is solved for E to characterize the elements per fault.

The SCR was fabricated with 7.5-micrometer metal lines stepping over 5-micrometer poly lines where the largest number of steps was 19,064. Results from 111 SCR structures measured on nine wafers fabricated in two fabrication runs indicated no breaks in the metal lines.

The COR was fabricated on thick oxide with 7.5-micrometer metal lines separated by 7.5-micrometer spaces. Results from 111 COR structures measured on nine wafers from two fabrication runs indicated no shorts between the metal lines. That is, metal bridging was not observed in these structures. This is not surprising since cross-bridge resistor measurements indicated that the metal lines are heavily overetched. The average linewidth is about 4 micrometers, well under the as-designed width of 7.5 micrometers. Also, the COR was not designed under worse-case conditions. A redesign of the COR requires that the metal lines cross perpendicular to poly lines.

The CAR was studied in some detail and was found to be surprisingly difficult to analyze. Initially it was thought that broken contacts would be easy to detect, but this turned out not to be the case.

A schematic diagram and photomicrograph for the CAR is shown in Figure 2.4.4-1. This structure is an assemblage of seven strings with various numbers of contacts in each string. The total number of contacts in the array is 7,396. The array was inserted into a test chip that was repeated 66 times across the 100-millimeter diameter silicon wafer used in this study. One wafer

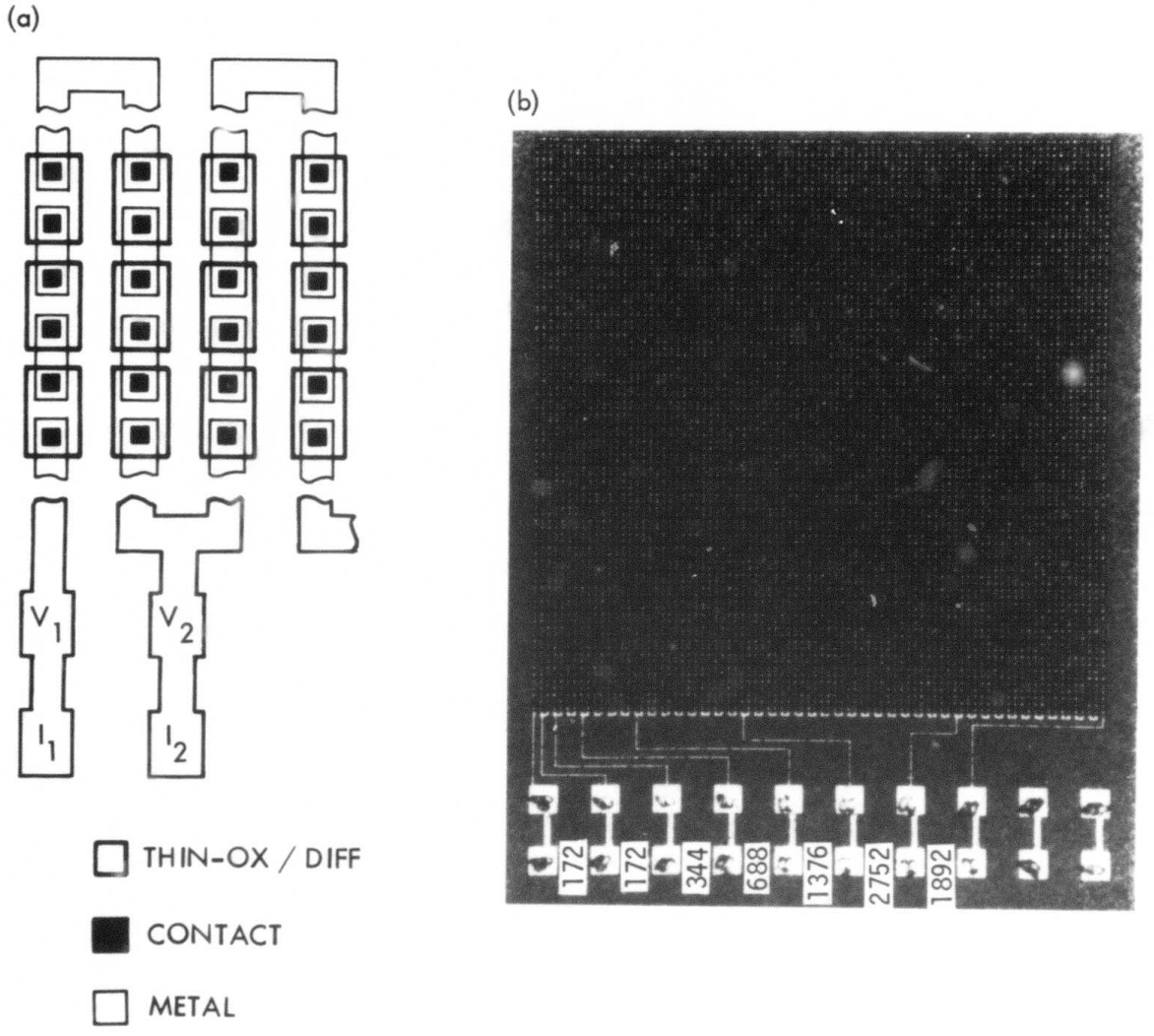


Figure 2.4.4-1. Contact array resistor (CAR) where (a) is a schematic diagram and (b) is a photomicrograph showing the number of contacts in each string. The total number of contacts in the entire array is 7,396.

then contains 488,136 contacts. Each test chip contains four different types of contacts:

1. Metal to p-type diffused silicon

2. Metal to n-type diffused silicon

3. Metal to p-type polycrystalline silicon

4. Metal to n-type polycrystalline silicon

(The metal used in this study was aluminium and it may or may not have been doped with silicon.) Since our experimental results indicate that the metal to p-diffusion contact is most prone to failure, results from this array will be used in the following discussion.

The resistance of the string is measured using four-terminal structures where current is forced into I_1 and out of I_2, and the voltage drop is measured between V_1 and V_2. The resistance of the string is divided by the number of contacts in the string and is expressed in units of ohms/contact, which includes the resistance of the contact and the resistance of the materials linking the contacts.

Before measuring the resistance of a string, a "probe-down" test was performed by requiring that the resistance between strapped pads (i.e., pads labeled V_1 and I_1 in Figure 2.4.4-1(a)) is less than 1,000 ohms. Results from contact resistor arrays are shown in Figure 2.4.4-2. Arrays which failed to pass a probe-down test are noted by colons. There are 113 contact strings that failed to pass the probe-down test. As shown in the figure, most of the strings have a resistance per contact element between 40 and 60 ohms. The resistance of a contact element includes contact resistance, diffusion resistance, and metal resistance. The numbers on the wafer maps represent a range of resistances where "0" represents 40 to 42 ohms, "1" represents 42 to 44 ohms, and so forth. The variation in the resistance values are due primarily to variations in the sheet resistance of the p-diffusion.

The resistor strings contain different numbers of contacts, and this number is noted above each array along with the array file name (e.g., RADPX). Each array was measured with a 1-microampere current source that was bounded at 1 volt to prevent voltage breakdown on the structure. It follows that the maximum measured resistance for a string must be less than 1 megohm. For the resistor strings shown in Figure 2.4.4-1(b) the maximum resistance of the longest string is about 140 kilohms, which is much less than the limit of 1 megohm. As seen in Figure 2.4.4-2, five open-contact strings were detected out of a total of (66 x 7 - 113) = 356 contact strings.

Before the contact array can be analyzed for randomly occurring defects, faults must be differentiated between local point defects (those due to a random break at the metal-silicon interface) and nonlocal global defects. For the contact array resistor, a nonlocal global defect is one that results in an open string due to a reason other than a randomly occurring open metal-to-silicon contact. As seen in Figure 2.4.4-2, open strings occurred at two loca-

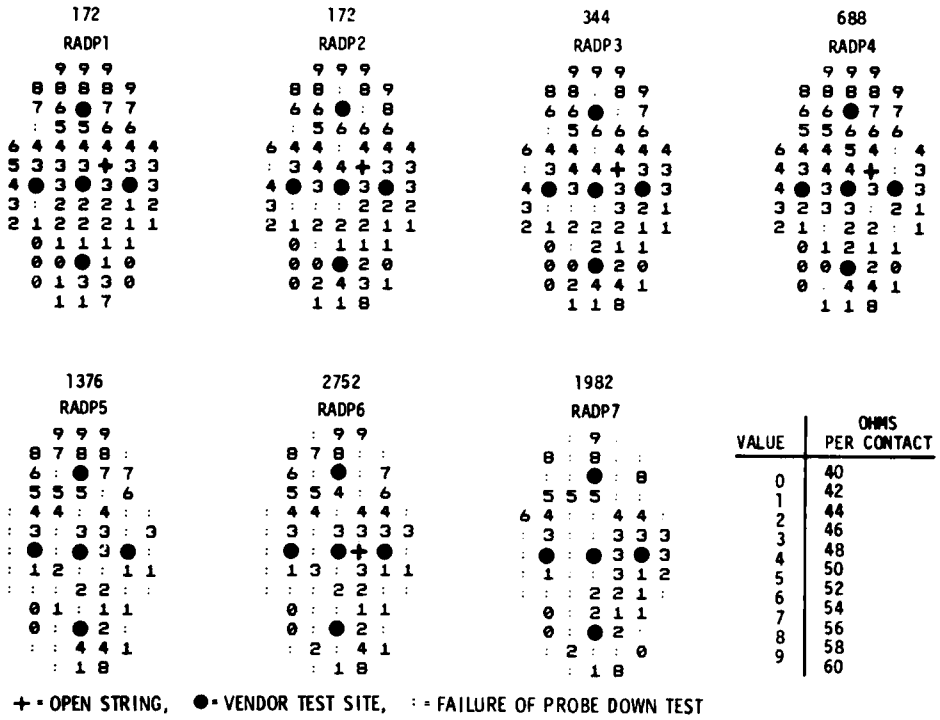


Figure 2.4.4-2. Wafer maps of the resistance per contact element of the seven resistor strings of the CAR (wafer 1213). The number of contacts in each string is noted above each wafer map.

tions: namely, row 8, column 5 and row 7, column 5. A photomicrograph of the row 8, column 5 site is shown in Figure 2.4.4-3. Here it is seen that a flaw has occurred in the metal layer that affects strings 1, 2, 3, and 4 and this corresponds to the open circuits observed electrically for these strings. A visual inspection of the row 7, column 5 site revealed no visual abnormality and so we concluded that a local defect was observed in this string.

Other nonlocal global defects were seen in various contact array resistors and these are shown in Figure 2.4.4-4. These defects are due to flaws in the photolithography, which can occur in any one of the layers.

An analysis of the data presented in Figure 2.4.4-2 indicates that the wafer contained at least 356 good contact strings with a total of 300,140 good metal to p-diffusion contacts. Of the 462 contact strings on the wafer, one string appeared to have a local defect, four strings had nonlocal defects, and 113 strings failed to pass the 1000-ohm probe-down test. (This test may be too stringent and a limit of 10,000 ohms may be more appropriate.) Because so few intended faults were observed it is difficult to quantify the results further.

The conclusions drawn from this work are:

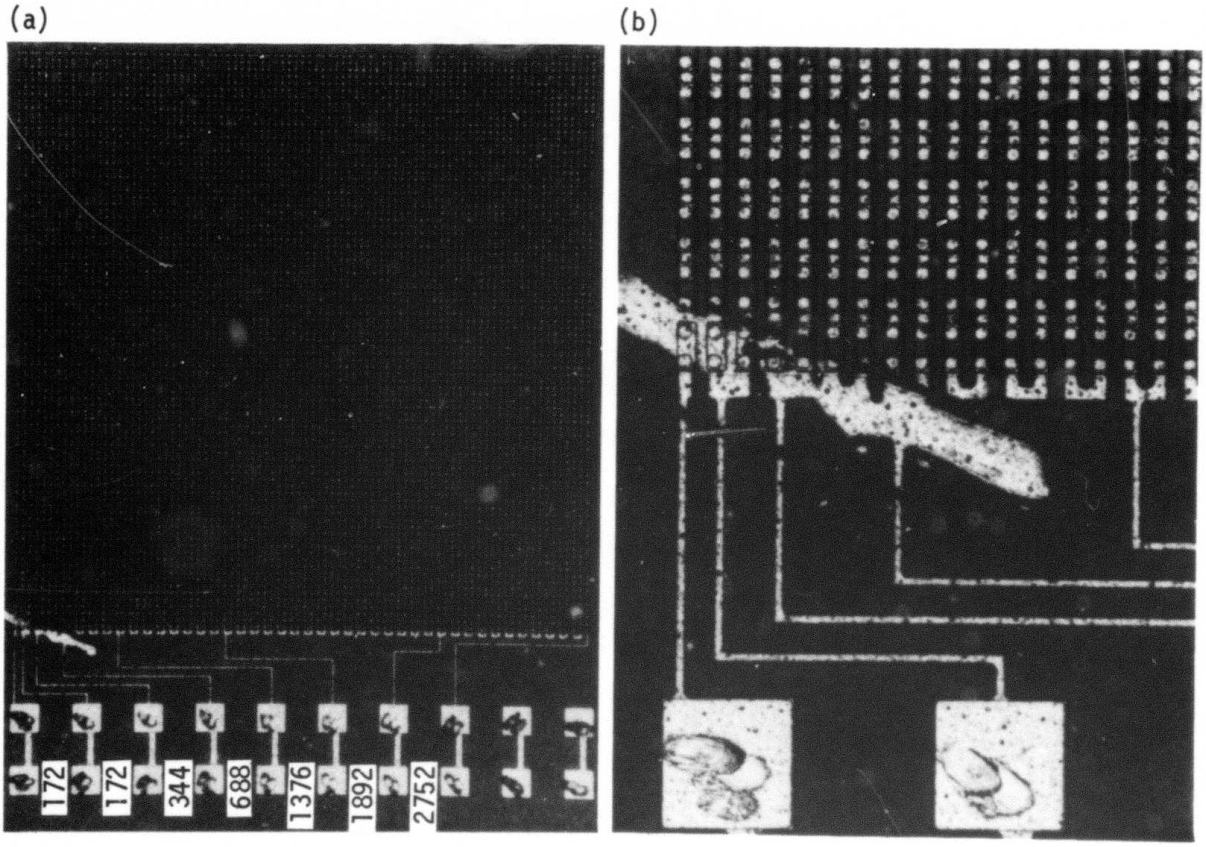


Figure 2.4.4-3. An example of the defect type that is excluded from the CAR analysis. This defective p-diffused CAR is located at row 8, column 5 of wafer 1213. (a) The overall array with a defect in the lower left-hand corner. (b) An expanded view of the defect in the metal layer.

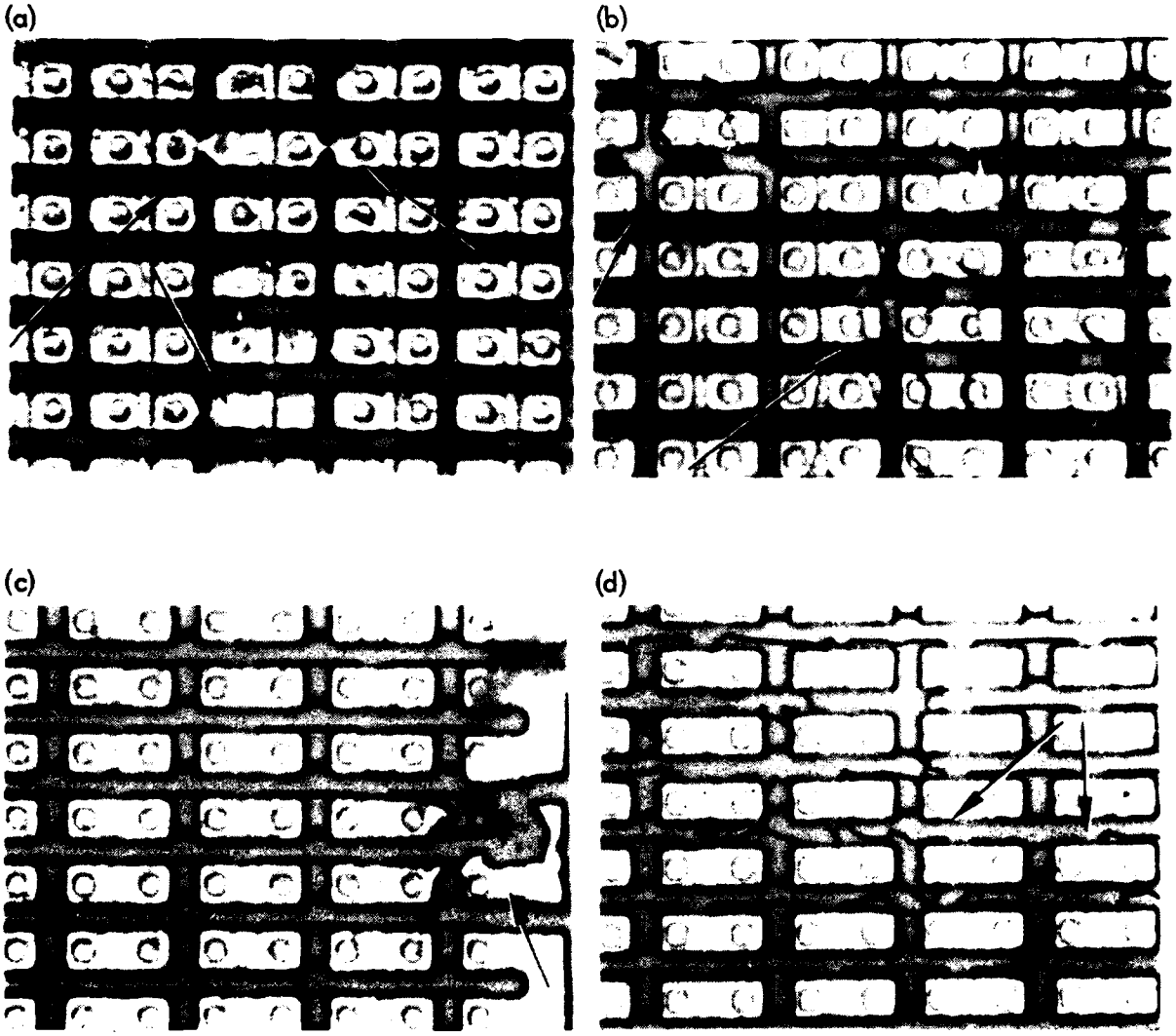


Figure 2.4.4-4. Photomicrographs of nonlocal global photolithographic-related flaws in contact resistors arrays with (a) shorts between metal links and missing contact windows, (b) poor definition of poly links, (c) open metal and misaligned contact windows, and (d) poor definition of diffused links.

1. The occurrence of local defects can be detected by examining the locations of open-circuit strings. If the open-circuit strings are located in adjacent strings in the same array, then this site should be visually inspected for the occurrence of a secondary fault.

2. Since we were unsuccessful in characterizing the local point defect frequency on a wafer with 66 CARs containing 7,396 contacts, we can conclude that the number of contacts on the wafer was too small to characterize faulty contacts.

3. The probe-down test is effective in eliminating probing errors from the data set. Further refinements require that a more suitable limit for the test and larger probe pads be designed into further test structures.

The pinhole array capacitor (PAC) was studied in some detail and was found to be relatively easy to analyze. The number of pinholes encountered for the wafers analyzed were easily detected.

A schematic diagram and equivalent circuit for the PAC are shown in Figure 2.4.4-5 and photomicrographs of the structure are shown in Figure

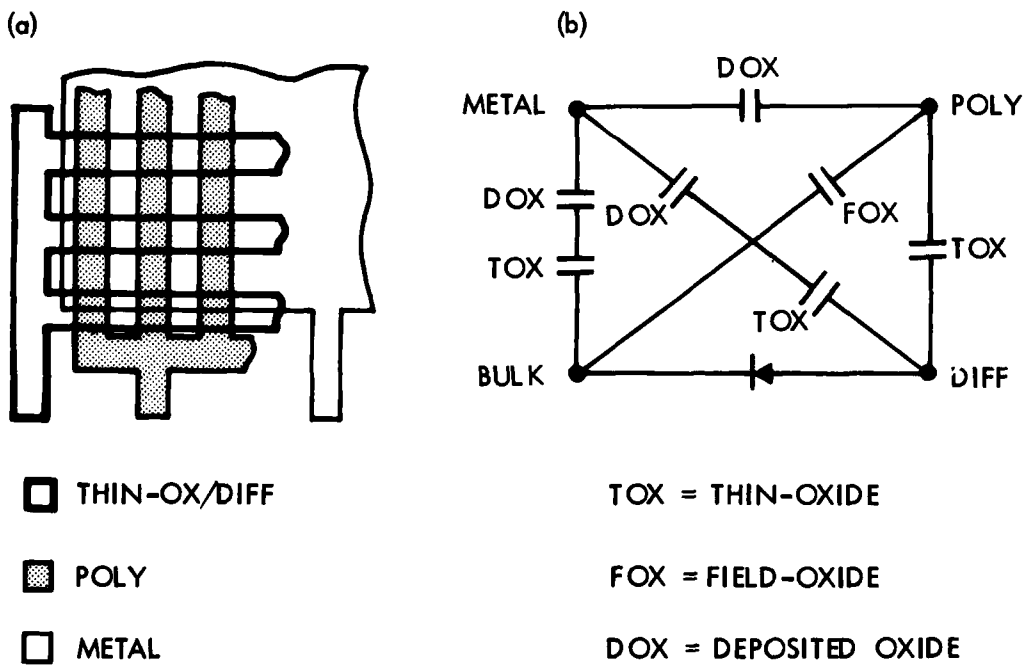


Figure 2.4.4-5. Pinhole array capacitor (PAC) (a) schematic diagram and (b) equivalent circuit.

2.4.4-6. The structure is composed of four conducting layers: bulk, diffusion, poly, and metal. The bulk and diffusion layers are isolated by a p-n junction space-charge region and the diffusion, poly, and metal layers are isolated by various oxide layers: thin oxide (or gate oxide), field oxide, and deposited oxide.

The diffused layer is a meshlike structure, as shown in Figure 2.4.4-6(c). This layer is formed from minimum geometry lines and spaces. The diffused layer is contacted once. At the top of the structure, as seen in Figure 2.4.4-6(a), the individual diffusion subarrays are connected together by a diffusion bus.

In this design the bulk is not contacted except through the back side of the wafer. The structure is composed of six subarrays with separate contacts for the poly and metal. The poly is formed from minimum geometry lines and spaces that cross the diffusion mesh at right angles. The number of poly-diffusion crossings forms the metric for the structure. Each subarray is capped by a metal layer.

Two types of structures were designed. One structure was designed for p-diffusion where the bulk material is n-type substrate. The other structure was designed for n-diffusion where the bulk material is p-well.

The test concept for the PAC requires that one detect the presence of shorts between the metal and poly layers, between the poly and diffusion layers, and between the metal and diffusion layers. The testing approach is slightly different for each layer.

The procedure for testing for shorts between the metal and poly layers requires that the measurement of the current for an applied voltage. In our measurement +5 volts was applied to the metal with respect to the poly.

The procedure for testing for shorts between the poly and diffusion layers requires that the poly be biased with respect to the diffusion such that channels are formed under the poly gates. This allows any poly to diffusion shorts to be "connected" to the diffusion bus. In our measurements +5 volts was applied to the poly with respect to the diffusion, irrespective of the diffusion type. This approach is satisfactory for n-diffusion. For p-diffusion the poly should be biased negatively with respect to the p-diffusion.

The procedure for testing for shorts between the metal and diffusion layers requires that the metal to poly layers be isolated. The poly is biased with respect to the diffusion so that channels are formed under the poly gates. The metal is then biased with both polarities and the current measured. In our measurements +5 volts was applied to the metal with respect to the diffusion and the poly was allowed to float.

For our instrumentation the open-circuit current values were less than 10^{-9} amperes. A pinhole resulted in a current of 10^{-4} to 10^{-3} amperes. This difference between the open and pinhole cases made the identification of pinholes a simple matter.

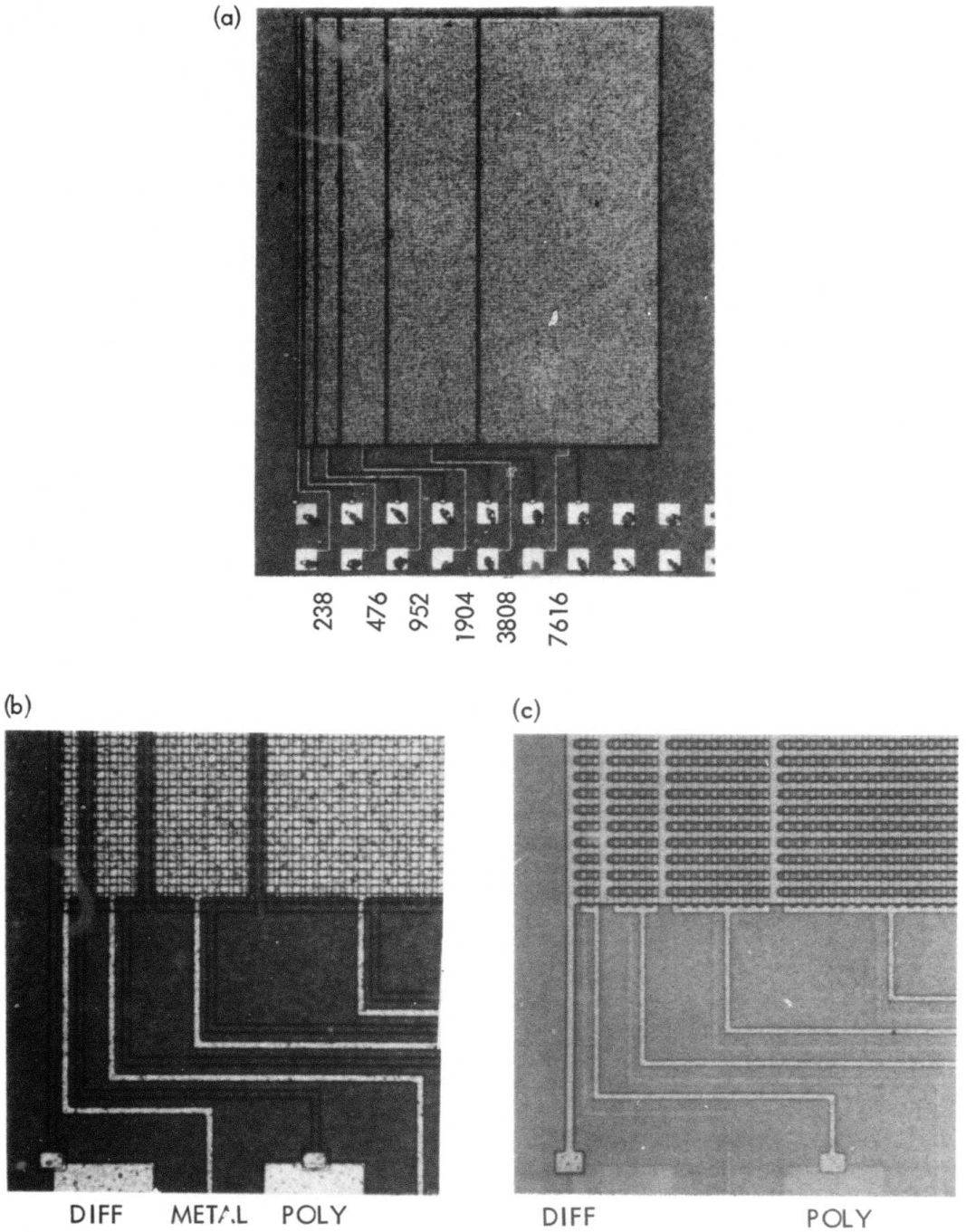


Figure 2.4.4-6. Photomicrographs of the PAC. (a) Entire test structure with the number of poly-diffusion crossings indicated. (b) Expanded view of the left-hand corner of the test structure. (c) Expanded view with the metal removed.

More pinholes were found between the metal and poly layers than between the poly and diffusion layers. A metal-poly short indicates that a pinhole has most likely occurred in the deposited oxide. However, as seen in the schematic diagram in Figure 2.4.4-5, there are three other possible conducting paths between the metal and the poly, but these require the unlikely occurrence of multiple pinholes in various layers.

Some pinholes were found between the poly and diffusion layers, which indicates that a pinhole has most likely occurred in the thin oxide. Some pinholes were found between the metal and diffusion layers, but when this occurred, pinholes also occurred between the metal and poly and between the poly and diffusion. In summary, these results allow the simplication of the rather complex equivalent circuit shown in Figure 2.4.4-5(b) into an equivalent circuit consisting of a metal to poly capacitor and a poly to diffusion capacitor.

The analysis of these results requires that shorts between the conducting layers be separated into randomly occurring, isolated pinholes and global defects that affect a large portion of the structure. To illustrate the analysis procedure, consider the wafer maps shown in Figure 2.4.4-7. If the measured current was less than 10^{-9} amperes, the site is noted by a colon. Otherwise, the site is marked by a plus sign. As seen in the figure, some plus signs are circled and these sites are excluded from the data set. The procedure followed in this analysis calls for the exclusion of data where a short occurs in adjacent arrays. Such faults are considered to be nonrandomly occurring faults.

A nonlocal global defect occurred at location row 5, column 7. As seen in Figure 2.4.4-7, a short was detected in all but the smallest subarrays. A photomicrograph of this site is shown in Figure 2.4.4-8 and reveals that a scratch has occurred after the poly was patterned and before the metal was deposited and so has caused a defect in the deposited oxide layer.

Once nonlocal global defects are identified and eliminated from the data set, the remaining data is analysed by using the exponential yield equation:

$$Y = \exp(-N/E)$$

where N is the number of elements (in this case the number of poly-diffusion crossings) and E is the number of elements per defect. The data represented by the wafer maps in Figure 2.4.4-7 are plotted in Figure 2.4.4-9, where the entire data set and the randomly occurring pinhole data are presented. (In the analysis the data point for the largest subarray was used to establish the E values.)

Certain conclusion can be drawn from this work:

1. This test structure allows a rapid and easy analysis of the pinhole density in oxide layers.

2. By inspecting for shorts in adjacent subarrays, one can often identify nonlocal global defects.

3. The design of this test structure must be improved by allowing for a probe-down test. Each layer that is contacted must have double probe pads to allow for a probe-down test. This is crucial, for in this structure a pinhole can go undetected if probes are not making good contact with the probe pads.

4. The design and/or testing of this test structure must be improved to allow currents in the diffusion layer to reach the diffusion bus. The poly layers must be biased so that channels are formed between the diffused regions to allow current to reach the diffusion bus.

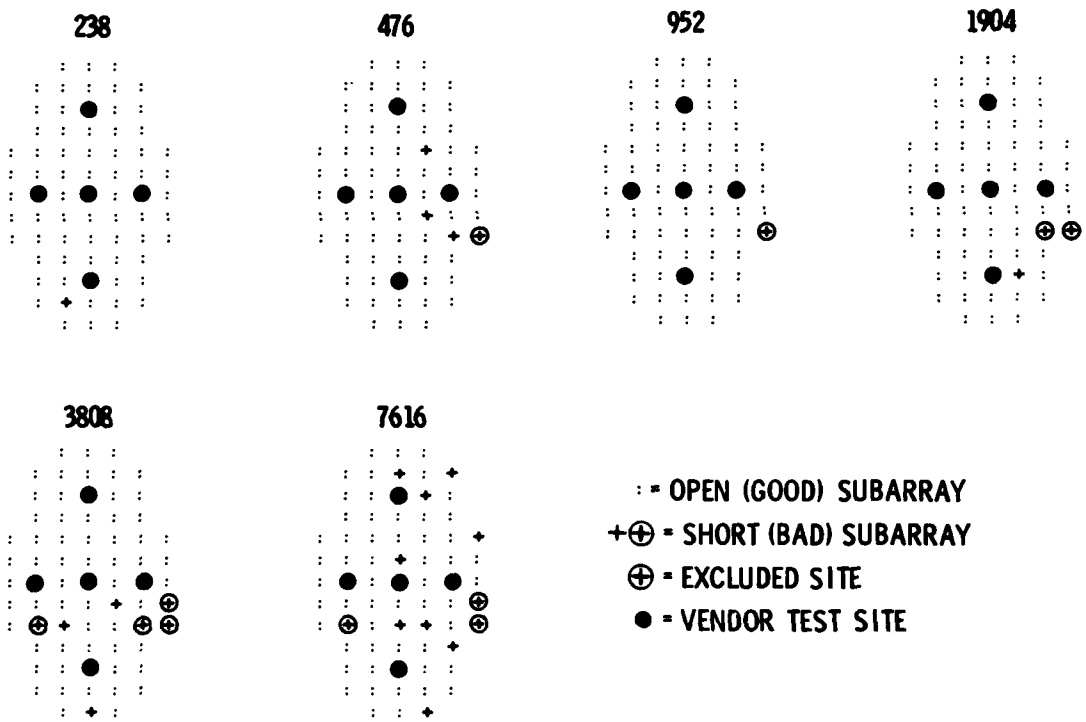


Figure 2.4.4-7. Wafer maps of the six subarrays found in the PAC, where shorts between the metal and n-poly are denoted by a plus (wafer 1104). The number of poly-diffusion crossings is noted above each wafer map.

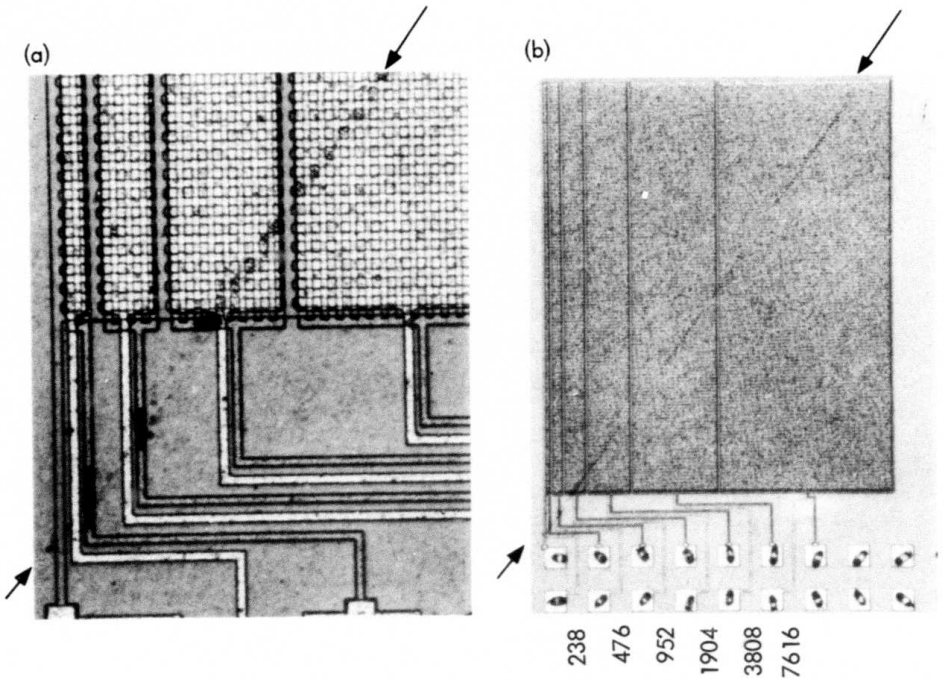


Figure 2.4.4-8. Photomicrograph of the PAC, showing a nonrandomly occurring defect that has caused a short between the metal and poly layers (site row 5, column 7, wafer 1104).

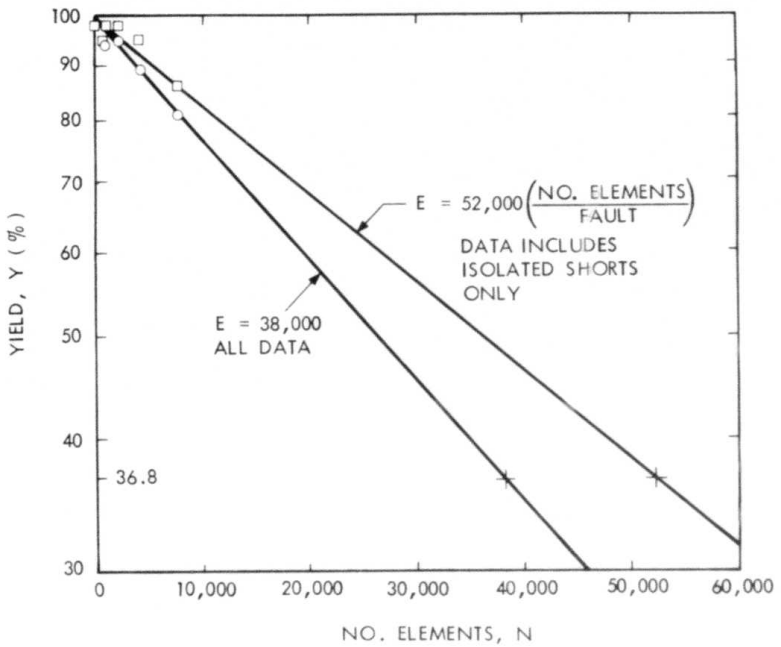


Figure 2.4.4-9. Yield analysis of the PAC data shown in Figure 2.4.4-7 (wafer 1104). The graph shows the 1/E point where N = E. This illustrates that the E value is extrapolated from high-yield numbers.

2.5 CMOS-BULK FOUNDRY RUNS

2.5.1 Foundry Wafers and Schedule

During the contract period two CMOS-bulk silicon foundry runs were fabricated and evaluated. Photomicrographs of the first two foundry runs are shown in Figure 2.5.1-1. The chips are 7-millimeter squares. The 100-millimeter diameter wafer contained 145 chips, and the 75-millimeter wafer contained 89 chips. Vendor 1 supplied five test chips whereas vendor 2 supplied no test chips. Since these runs were experimental, no specifications were placed on the vendor except to deliver a fixed number of fully fabricated, whole wafers. In future runs specifications will be developed for certain devices such as the p- and n-channel transistors.

During the contract period, runs 3 and 4 were initiated. The overall schedule is shown in Table 2.3.2-1. The turnaround time between the file close date and receipt of wafers was as short as 10 weeks and as long as 23 weeks.

2.5.2 Foundry Results and Conclusions

Results from two silicon foundry runs are listed in Tables 2.5.2-1 through 2.5.2-3. The first two tables list the parametric results from runs 1 and 2, respectively, and the third table lists the yield analysis results. In these tables D = diffusion, M = metal, N = n-type, P = poly or p-type.

The parametric results are expressed in the following format:

$$\mu \pm \sigma \left(\frac{\sigma}{\mu}\ \%\right) N_E/N$$

where μ is the mean value, σ is the standard deviation, $((\sigma/\mu)\%)$ is the percent standard deviation, N_B is the number of excluded values, and N is the total number of values acquired. The number of excluded values was determined by an empirical procedure that excluded outliers and invalid data from the main body of the data. These procedures have been automated, as explained in the next section. For the purposes of this discussion the number of excluded values is useful in identifying problem areas. The objective of this analysis is to determine the mean for the nominal part of the data set and to judge the acceptability of these values.

When examining the values one looks for certain nominal values. The linewidths used to form the cross bridge test structures were 5 micrometers for the poly and diffusion layers and 7.5 micrometers for the metal layer. As seen in the table, the poly layers are undersized, being in the 3-micrometer range, and the metal layer can be badly overetched, being in the 4-micrometer range.

Contact resistance values vary depending on the type of material being contacted, but the values listed are more than likely acceptable for CMOS-bulk because of the much higher impedance of the transistor on resistance.

The values for wafer 1102 listed in Table 2.5.2-1 indicate that this wafer has an excessively large number of excluded values, especially for those values labeled (a) - (d). As seen in the table the parameters affected are:

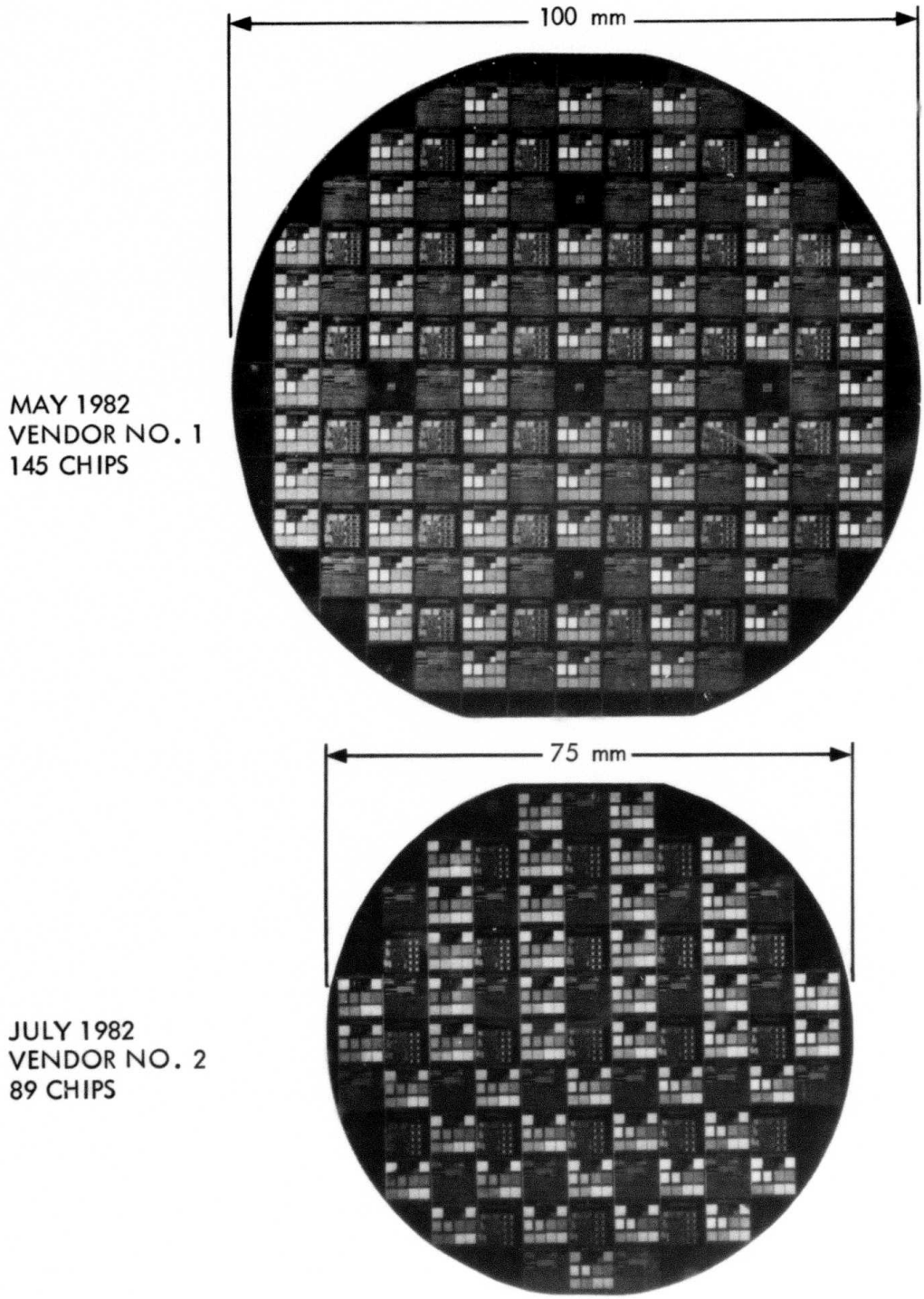


Figure 2.5.1-1. Silicon foundry CMOS-bulk wafers.

Table 2.5.2-1. CMOS-Bulk Foundry Parametric Characterization (Run 1)

WAFER	1102	1104	1113
LINEWIDTH (micrometers)			
ND (5)	5.34±0.17(3.2) 19/138	5.32±0.16(3.0) 17/138	5.34±0.03(0.56) 0/1
PD (5)	5.0±0.98(19.6) 56/138	4.80±0.16(3.3) 12/138	4.72±0.03(0.62) 0/1
NP (5)	3.63±0.21(5.8) 20/138	3.71±0.29(7.8) 9/138	3.49±0.16(4.6) 0/13
PP (5)			4.16±0.39(9.4) 0/38
M (7.5)	4.90±0.40(8.2) 20/138	4.85±0.17(3.5) 16/138	4.59±0.04(0.85) 0/1
SHEET RESISTANCE			
ND (ohm/sq.)	13.3±0.85(6.4) 17/138	13.3±0.7(5.3) 16/138	11.2±0.14(1.22) 3/1
PD (ohm/sq.) (a)	80.6±17.6(21.7) 70/138	66.9±7.7(11.5) 5/138	64.9±8.8(13.6) 2/13
NP (ohm/sq.)	15.96±0.50(3.1) 17/138	15.94±0.16(1.0) 9/138	15.70±0.57(3.7) 0/1
PP (ohm/sq.)			16.39±1.2(7.3) 0/38
M (milliohm/sq.)	21.9±0.37(1.68) 20/138	22.1±0.3(1.3) 15/138	22.7±0.83(3.7) 10/1
CONTACT RESISTANCE (ohm) (2 lambda X 2 lambda, I = 1 milliampere)			
M/ND	2.59±0.35(13.4) 17/138	2.58±0.32(12.4) 10/138	2.18±0.17(7.7) 11/1
M/PD (b)	10.3±2.3(22.2) 75/138	8.2±1.6 (14.5) 7/138	8.2±1.3(16.0) 6/138
M/NP	1.37±0.17(12.1) 24/138	2.01±0.49(24.4) 29/138	1.65±0.27(16.5) 34/1
M/PP			
3 TERM. PCH. TRANSISTOR (L=2 lambda, W=10 lambda)			
VT (volt)	0.91±0.04(4.3) 8/138	0.79±0.03(3.2) 6/138	0.90±0.027(3.0) 3/13
KPRIME (microampere/ volt^2) (c)	8.86+1.2(13.4) 74/138	10.1+0.97(9.6) 6/138	8.78+1.1(12.5) 3/138
IDSO (nanoampere)	5.13±1.29(25.1) 24/138	4.38±1.10(25.0) 8/138	3.92±0.91(23.3) 0/13
IGLEAK (nanoampere)	4.23±1.3(30.4) 21/138	0.37±0.10(28.8) 3/138	4.11±0.82(19.9) 56/1
VDBKDN (volt)	22.5±3.9(17.5) 39/138	14.6±3.89(26.7) 0/138	22.0±3.2(14.6) 3/138
4 TERM. NCH TRANSISTOR (L=2 lambda, W=10 lambda)			
VTO (volt)	1.14±0.06(5.1) 20/138	0.90±0.02(2.6) 9/138	0.897±0.086(9.6) 6/1
KPRIMEO (microampere/ volt^2)	15.3+0.55(3.6) 20/138	27.5+2.9(10.5) 10/138	25.9+3.0(11.7) 8/138
GAMMAO1 (volt$^{1/2}$)	1.57±0.14(8.6) 18/138	1.25±0.01(1.4) 12/138	1.28±0.18(14.1) 5/13
IDSO (picoampere)	0.4±0.2(48.6) 26/138	2.23±2.18(14.6) 0/138	2.37±0.54(22.7) 0/13
IDLEAK (nanoampere)	0.4±0.02(5.8) 2/138	0.39±0.02(5.7) 0/138	0.33±0.02(6.2) 0/138
IGLEAK (nanoampere)	0.4±0.01(3.2) 1/138	0.47±0.02(4.4) 0/138	0.43±0.009(2.0) 0/13
VDBKDN (volt)	22.3±0.57(2.6) 17/138	22.1±0.2(1.0) 2/138	22.2±1.14(5.1) 0/138
INVERTER (L = 2 lambda, pull down: W=2 lambda, pull up: W = 3 lambda)			
VHIGH (volt) (d)	4.92+0.4(8.2) 36/138	4.98+0.05(1.0) 1/138	4.99+0.02(0.38) 4/13
VLOW (millivolt)	0.17±0.012(7.1) 18/138	0.15±0.13(117) 1/138	0.16±0.007(4.4) 4/13
VINV (volt)	2.23±0.38(16.9) 36/138	2.58±0.14(5.4) 1/138	2.46±0.09(3.7) 5/138
GAIN	17.4±3.8(21.7) 39/138	13.9±1.5(10.5) 4/138	17.2±1.9(11.4) 5/138

Table 2.5.2-2. CMOS-Bulk Foundry Parametric Characterization (Run 2)

WAFER	2105	2107
LINEWIDTH (micrometers)		
ND (5)	4.93±0.42(8.5) 6/89	4.75±0.17(3.7) 42/89
PD (5)	4.36±0.62(14.1) 0/89	3.89±0.21(5.5) 0/89
NP (5)	3.87±0.41(10.6) 4/89	3.57±0.54(15.1) 2/89
M (7.5)	6.62±0.26(3.9) 0/89	6.59±0.11(1.7) 0/89
SHEET RESISTANCE		
ND (ohm/sq.)	11.6±1.13(9.7) 4/89	11.9±1.55(13.0) 28/89
PD (ohm/sq.)	80.27±5.5(6.9) 1/89	75.7±0.51(6.8) 4/89
NP (ohm/sq.)	19.24±1.2(6.3) 0/89	18.14±1.27(7.0) 3/89
M (milliohm/sq.)	25.5±1.2(4.7) 0/89	25.6±12.5(4.9) 0/89
CONTACT RESISTANCE (ohms) (2 lambda X 2 lambda, I = 1 milliampere)		
M/ND	3.4±0.65(18.9) 2/89	4.0±1.2(30.4) 33/89
M/PD	16.3±3.2(19.5) 1/89	21.5±6.3(29.3) 3/89
M/NP	3.37±0.98(29.1) 1/89	6.46±2.5(38.3) 3/89
M/PP		
3 TERM. PCH. TRANSISTOR (L=2 lambda, W=10 lambda)		
VT (volt)	0.75±0.15(20.2) 34/89	0.865±0.14(16.3) 44/89
KPRIME (microampere/ volt^2) (c)	4.75±1.3(27.7) 26/89	4.55±1.13(24.9) 37/89
IDSO (nanoampere)	16.1±2.7(16.5) 1/89	20.3±3.0(15.0) 32/89
IGLEAK (nonoampere)	15.7±2.5(16.0) 0/89	19.7±3.1(15.5) 32/89
VDBKDN (volt)	28.7±3.6(12.5) 1/89	29.04±0.56(1.9) 11/89
4 TERM. NCH TRANSISTOR (L=2 lambda, W=10 lambda)		
VTO (volt)	0.76±0.08(10.1) 2/89	0.69±0.10(15.2) 19/89
KPRIMEO (microampere/ volt^2)	32.2±4.1(12.8) 1/89	36.9±4.0(10.8) 32/89
GAMMA01 (volt$^{1/2}$)	1.18±0.06(5.07) 2/89	1.21±0.55(45.2) 20/89
IDSO (picoampere)	3.5±2.0(55.9) 3/89	5.68±4.07(71.7) 4/89
IDLEAK (nanoampere)	0.36±0.06(17.0) 1/89	0.37±0.03(8.7) 1/89
IGLEAK (nanoampere)	0.44±0.02(4.3) 1/89	0.41±0.1(24.4) 0/89
VDBKDN (volt)	19.1±1.5(7.8) 1/89	19.4±2.0(10.5) 4/89
INVERTER (L = 2 lambda, pull down: W=2 lambda, pull up: W = 3 lambda)		
VHIGH (volt) (d)	4.48±0.69(15.4) 6/89	4.22±1.1(26.3) 4/89
VLOW (millivolt)	170±254(150) 10/89	
VINV (volt)	2.53±0.84(33.2) 4/89	
GAIN	6.97±4.74(68.0) 12/89	

Table 2.5.2-3. CMOS-Bulk PAC Yield Characterization, E(elements/defect)

WAFER	1102	1103	1104	1108	1110	1114
M/ND	ND	2.2E5	4.5E5	2.2E5	4.5E5	ND
NP/ND	ND	2.2E5	4.5E5	2.2E5	2.2E5	ND
M/NP	1.1E5	3.1E4	3.1E4	3.1E4	2.5E4	1.8E4
M/PD	ND	ND	ND	4.5E5	4.5E5	2.2E5
PP/PD	1.1E5	2.2E5	1.5E5	4.5E5	2.2E5	2.2E5
M/PP	2.2E5	4.7E4	1.1E5	7.2E4	4.2E4	4.2E4

WAFER	2103	2109	2111
M/ND	3.3E5	1.1E5	3.3E5
NP/ND	1.1E5	6.4E4	1.7E5
M/NP	6.4E5	8.2E4	4.5E4
M/PD	3.3E5	ND	ND
PP/PD	1.1E5	ND	ND
M/PP	1.1E5	ND	8.2E4

*Values based on faults observed on largest array where N = 7,616.
ND = no faults detected.

(a) p-diffusion sheet resistance, (b) metal to p-diffusion contact resistance, (c) p-transistor KPRIME, and (d) inverter V_{high}. The problem was diagnosed by creating the wafer maps shown in Figure 2.5.2-1, where it is seen that the region where the inverters are stuck low correlates with the open metal to p-diffusion contacts. At this time the problem is felt to be a failure to properly dope the p-region. This explains why the p-diffusion sheet resistance and the p-transistor KPRIME values are also bad in this region.

The values listed in Table 2.5.2-1 indicate that wafer 1102 has inverter V_{high} values that are not pulling close to V_{DD} = 5 volts. Detailed wafer maps are shown in Figure 2.5.2-2. Regions where inverters fail to pull within 0.1 volts of the rails are shown within the outline. A similar analysis for run 2 is shown in Figure 2.5.2-3. The results shown in Figures 2.5.2-2 and 2.5.2-3 have led us to conclude that the toggling of the inverters is a good first test in qualifying wafers. Further studies are planned to identify specific reasons for inverters failing to pass the toggle test.

The yield analysis values for runs 1 and 2 are listed in Table 2.5.2-3. We attempted to evaluate our four yield analysis test structures: CAR, SCR, COR, PAC. As discussed in Section 2.4.4, we were successful in obtaining values only for the PAC. As seen in the table, the lowest E-values (most pinholes) generally occur between the metal and poly layers.

Tentative conclusions to be drawn from this work are:

1. Wafers have regions where parameters cluster into acceptable (nominal) and unacceptable (excluded) values.

2. Knowledge gained on one wafer cannot be transferred to other wafers within the same lot.

3. Data acquired from wafer probe measurements can have poor integrity due to probing problems and due to photolithographically flawed structures.

4. A significant amount of unacceptable data must be sorted from the total data set.

5. Inverters are potentially useful as a first-level screen for wafer evaluation.

2.5.3 Data Analysis Tool

The empirical procedures used to evaluate the first two foundry runs were unsatisfactory. A rigorous approach was needed and the STAT2 [1] data analysis program developed at NBS was chosen. The program commands for the JPL version of STAT2 are listed in Table 2.5.3-1. The most frequently used commands are the XGT, XLT, XOL, and PLT commands.

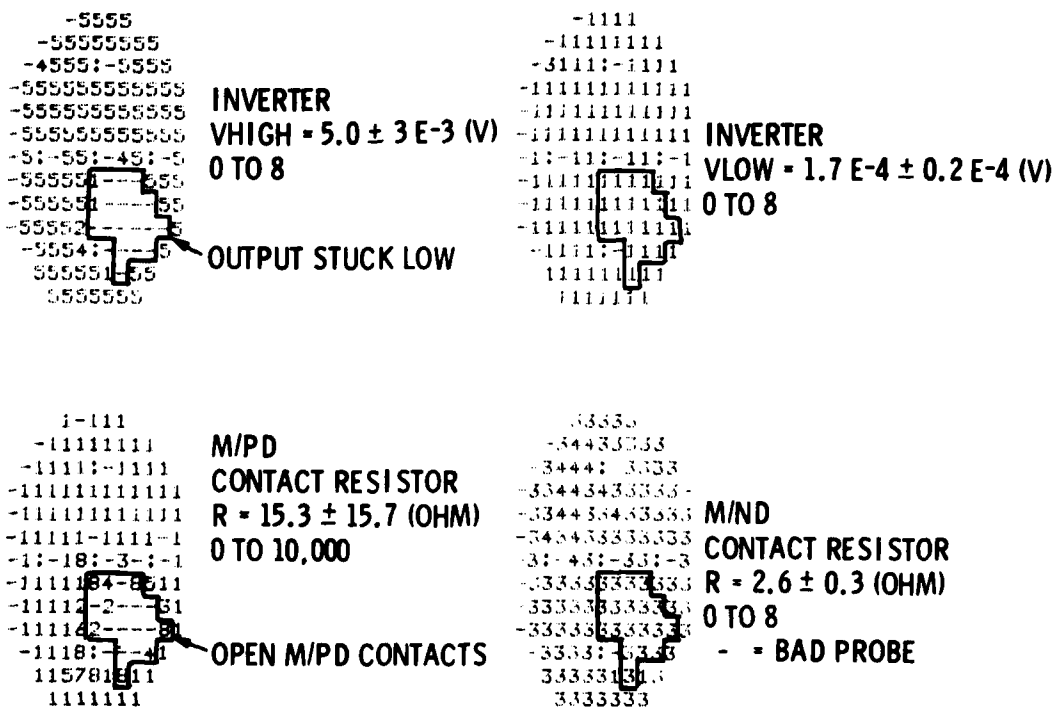


Figure 2.5.2-1. Wafer maps of CMOS-bulk inverters and contact resistors showing that some inverter outputs are stuck low due to open metal to p-diffusion contacts (wafer 1102).

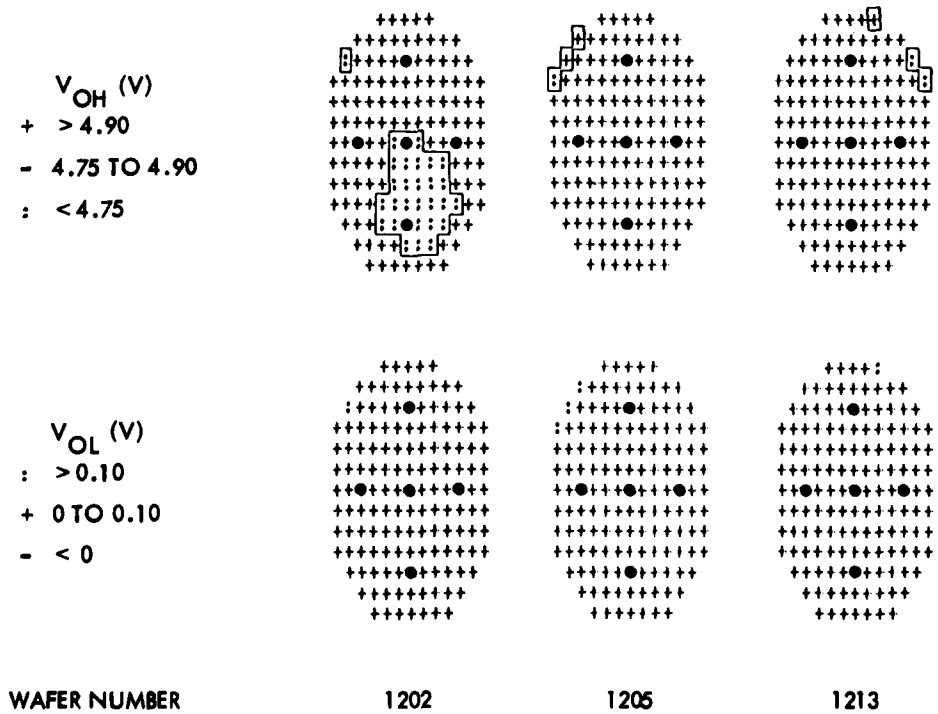


Figure 2.5.2-2. Wafer maps of CMOS-bulk inverters from run 1 showing the regions where the inverters failed to toggle properly.

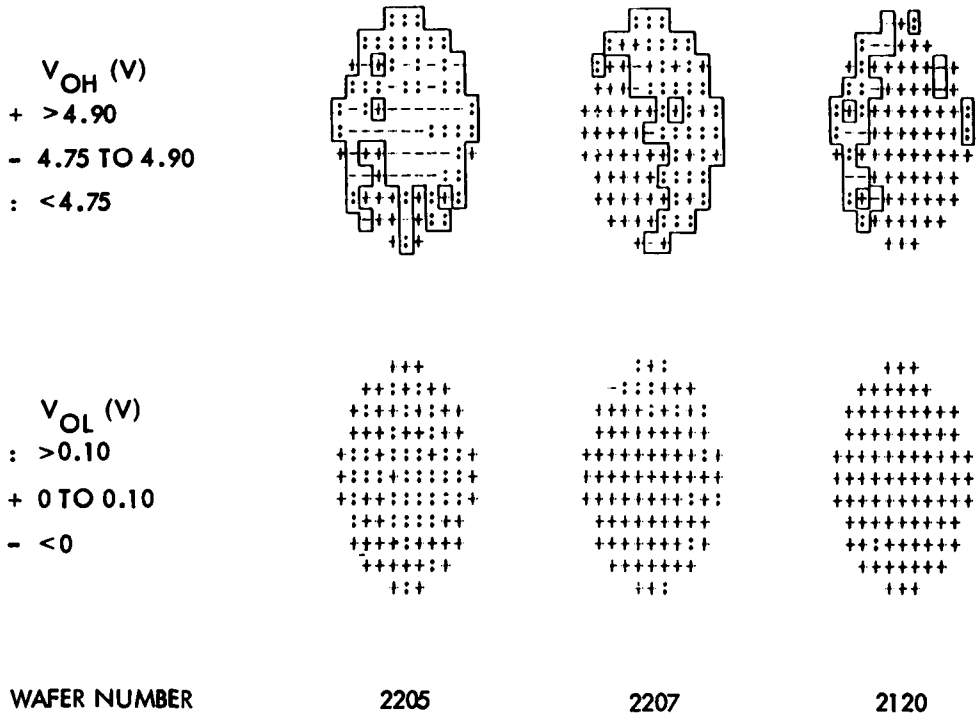


Figure 2.5.2-3. Wafer maps of CMOS-bulk inverters from run 2 showing the regions where the inverters failed to toggle properly.

Table 2.5.3-1. Statistical Analysis/Wafer Map Program Commands

```
END Exit
RESTO Restore all points to included status
ENN Enter value (N) for N*sigma
PRSTAT Print statistics
HISTOG Draw histogram
LIP List point (ROW, COL)
IIP Include point (ROW, COL)
XIP Exclude point (ROW, COL)
LXP List excluded points
XOL Exclude outliers (DELTA = 0.05 - 0.90)
XPP Exclude peripheral points
XNS Exclude points GT N*sigma from mean
XGT Exclude points GT than (VALUE)
XLT Exclude points LT (VALUE)
PLT Wafer map (MODE[1,0,-1],AMIN,AMAX)
CRUREAD Read CRUNCH file (FILENAME)
CRULST List current CRUNCH file
CRUSEL Select CRUNCH array (# or NAME)
AXP Alter excluded points
AIP Alter individual point (ROW, COL, VALUE)
NEG Negate data values
MAP Wafer map (W,H,SC,LBL,LO,HI)

>
```

A new format was chosen for data presentation:

$$\mu \pm \left(\sigma \; \frac{\sigma}{\mu} \; \%\right) N_N, \; N_O, \; N_I$$

where μ is the mean value, σ is the standard deviation, $((\sigma/\mu)\%)$ is the percent standard deviation, N_N is the number of nominal values, N_O is the number of outlier values, and N_I is the number of invalid values.

Invalid data are identified as illustrated in the flow diagram shown in Figure 2.5.3-1. It is assumed that the test system is calibrated. Invalid data are first identified by instrumentation validation tests that consist of a probe-down test and an instrumentation bound test. The probe-down test consists of a two-terminal resistance check. The test structure validation tests consist of a model check, data stability test, and double-check. The model check validates the structure under test. That is, if the structure is supposed to be a resistor, then the current-voltage characteristic must be linear. The data stability test requires that the same parameter be measured twice, perhaps after the structure has been operated in some other condition, and that the difference in values be within a given tolerance. For example, the parameter might be transistor threshold voltage measured before and after a gate voltage stress. The double-check test can take two forms. Repeated measurements of a parameter can constitute a double-check test. Verification of the pitch measurements using the split-cross-bridge resistor is a double-check validating the results.

Once invalid data are identified, the nominal values are sought by excluding outliers. The procedures are separated between continuous parameters (such as transistor threshold voltage and sheet resistance) and random parameters (such as pinholes and broken wires). The data analysis procedures are shown in Figure 2.5.3-2. As shown in the figure, the outlier exclusion is based on either (a) expected values, (b) $K\sigma$ elimination, or (c) correlations. Expected value exclusion is based on a preset tolerance for a parameter. For example, the output voltage of an inverter must pull within a 0.1 volt of the rail or be excluded from the data set. The $K\sigma$ elimination is executed by the XOL command in STAT2. This command eliminates outlier values that are more than a certain number of standard deviations from the mean value with a specified probability that good data will be excluded. Correlation exclusion is based on the proposition that randomly occurring events are unlikely to occur in adjacent subarrays of random fault test structures such as the PAC (pinhole array capacitor). An example of the correlation exclusion procedure is given in Section 2.4.4.

A example of the use of the data analysis procedure is shown in Figures 2.5.3-3 and 2.5.3-4. The data set was derived from n-channel transistor threshold voltage measurements. The entire data set includes 143 values and of these, five values represent the vendor test chip with the value zero. In addition, five other measured values are negative; clearly these are invalid data that must be excluded from the data set. The first step is to exclude data that are less than a small positive value; this is done with the XLT IE-9 command. The data were then plotted with a lower bound of 0 and an upper bound of 1 volt with the command PLT 0, 0, 1. The result is shown in Figure 2.5.3-3.

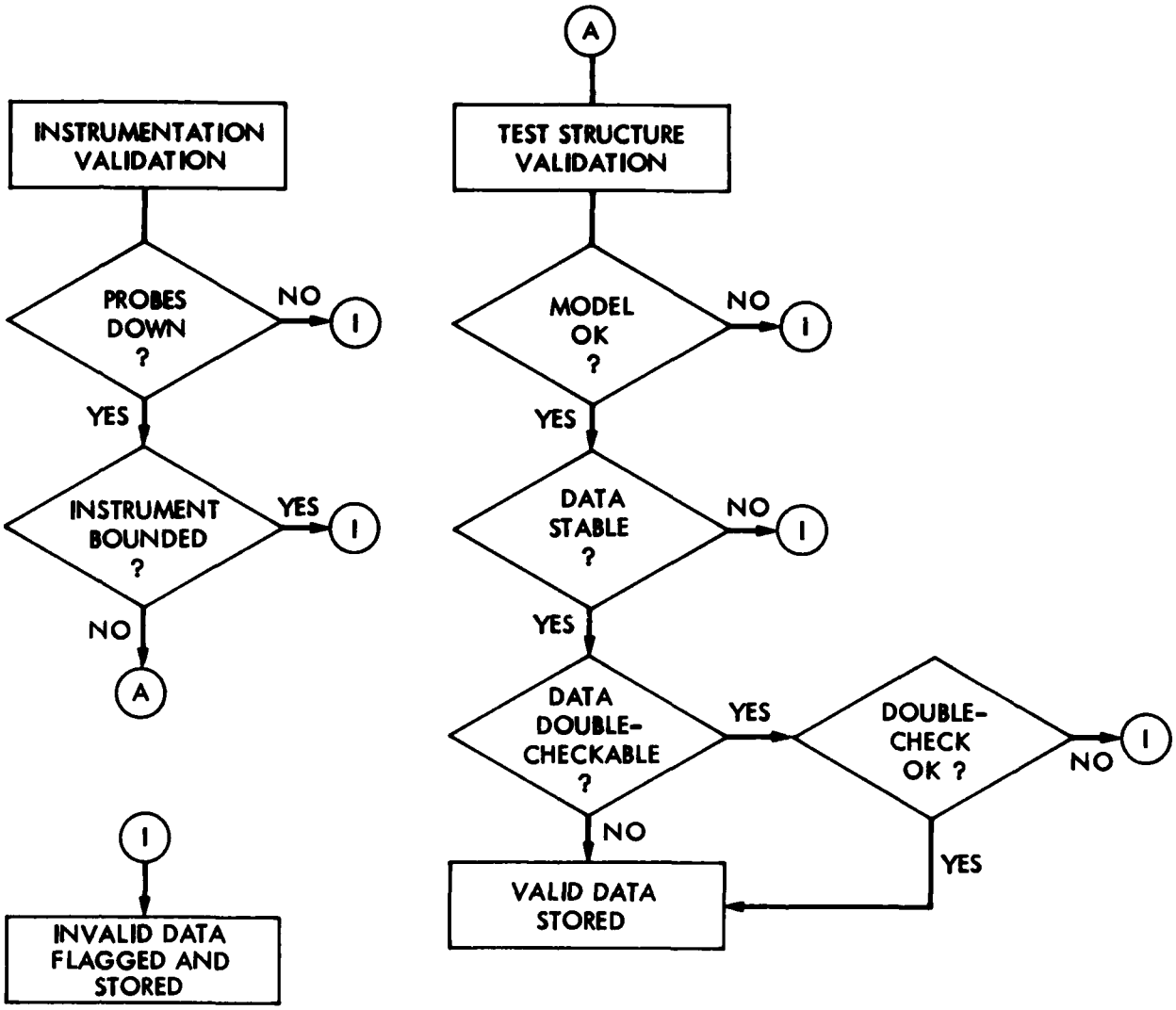
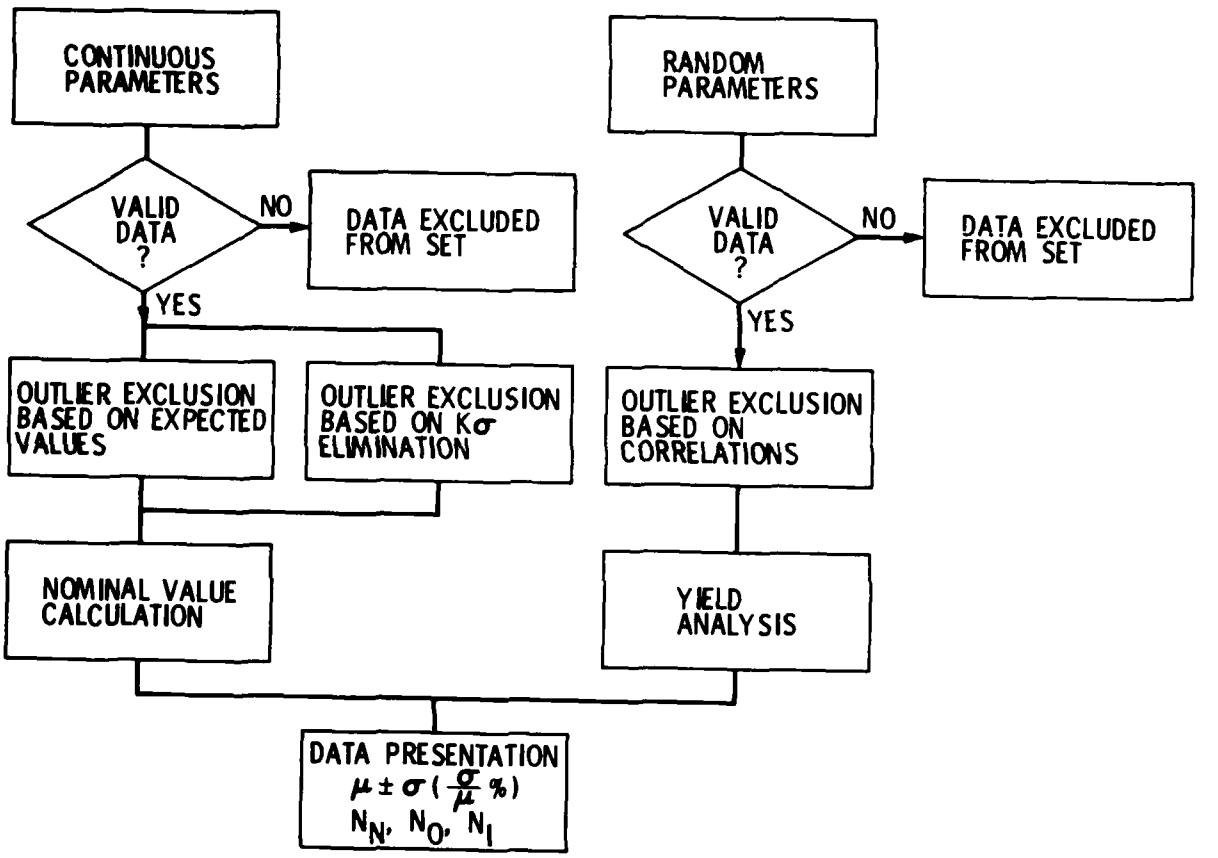


Figure 2.5.3-1. Data integrity assurance scheme.



μ = MEAN VALUE
σ = STANDARD DEVIATION
N_N = NO. NOMINAL VALUES
N_O = NO. OUTLIER VALUES
N_I = NO. INVALID VALUES

Figure 2.5.3-2. Data analysis procedure.

```
File:   CZSA13.107        Array:   VT0

RANGE   1.626E+00 /   1.143E-01   MEAN   8.910E-01   MEDIAN   9.001E-01
SIGMA   1.096E-01 (    12.30 %)    1*SIGMA   1.096E-01

          87888
        888888888
       88888@88888
      888888888888:
      888888899999:
      :88898989994:
      89@899@99909:
      888999999999+
      8889999899+91
      8888999999999
       89999@99999
        899999999
         8999999

UB   1.000E+00    -   0   1   2   3   4   5   6   7   8   9   +   XCL
LB   0.000E+00    0   0   1   0   0   1   0   0   1  63  65   2   10

>                                          @ = VENDOR TEST CHIPS =  -5

                                           INVALID DATA =            5
```

Figure 2.5.3-3. N-transistor threshold voltage wafer map and histogram in which data less than 10^{-9} volts have been excluded from the data set. (The data values are shown in Figure 2.1.3-1.)

```
File:   CZSA13.107        Array:   VT0

RANGE   9.289E-01 /   8.391E-01   MEAN   8.959E-01   MEDIAN   9.003E-01
SIGMA   1.920E-02 (     2.14 %)    3*SIGMA   5.760E-02

          0::00
        ::2012122
       :0333@33344
      :14555665555:
      355646667776:
      :6657676777::
      66@567@777@8:
      456777788888:
      5657677678:8:
      6566678888997
       68777@88877
        578678998
         6787788

UB   9.289E-01    -   0   1   2   3   4   5   6   7   8   9   +   XCL
LB   8.391E-01    0   5   3   4   7   5  16  25  32  22   4   0   20
>                                          @ = VENDOR TEST CHIPS =  -5

                                                                    15

                                           INVALID DATA =           -5

                                           OUTLIERS =               10
```

Figure 2.5.3-4. N-transistor threshold voltage wafer map and histogram in which invalid data (data less than 10^{-9} volts) and outlier data have been excluded from the data set.

The XCL = 10 value indicates that the five vendor test-chip values and the negative values were excluded from the data set. The threshold voltage data statistics are:

$$V_{T\mu} \pm V_{T\sigma}\ (V_{T\sigma}\%)\ N_N,\ N_O,\ N_I$$

$$0.891 \pm 0.11\ (12.3)\ 133,\ 0,\ 5$$

which indicates that the percent standard deviation is 12.3%.

It is clear from the histogram in Figure 2.5.3-3 that outliers must be excluded from the data, for the data range from 0.11 to 1.6 volts. The exclusion is done by using the XOL 0.2 command. Using PLT 1, 1, 1 the data are plotted using the auto scale feature and the result is shown in Figure 2.5.3-4. As seen in the figure, an additional 10 data points were eliminated from the data set. The threshold voltage statistics are:

$$V_{T\mu} \pm V_{T\sigma}\ (V_{T\sigma}\%)\ N_N,\ N_O,\ N_I$$

$$0.89 \pm 0.019\ (2.1)\ 123,\ 10,\ 5$$

Now the percent standard deviation is a more realistic 2.1%.

An examination of the wafer map shown in Figure 2.5.3-4 indicates the location of the invalid and outlier data sites. Most of them are located on the periphery of the wafer where manufacturing flaws are most likely to occur. Also evident in the figure is the variation in threshold voltage across the wafer. This variation is most likely due to the ion implantation step which adjusts the threshold voltage of the transistors.

2.5.4 Reference

1. Mattis, R. L., L. J. Till, and R. C. Frisch, "A Computer Program for Analysis of Data from Microelectronic Test Structures," National Bureau of Standards Internal Report 82-2492 (June 1982).

LOGICAL MODELS OF PHYSICAL FAILURES*

C. Timoc, M. Buehler, T. Griswold, C. Pina, F. Stott and L. Hess

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

Abstract

This paper identified some of the most frequent physical failures in bipolar and MOS technologies and modeled their logical behavior. The prevalent failures in the bipolar technology are: open connections, shorts between connections, and "piped" transistors with excessive collector to emitter leakage current. A selected bipolar technology was used as an example for modeling failures. The open connection behaved as a single stuck-at, the short as a wired-AND, the "piped" output transistor as a single stuck-at, and the "piped" input transistor as a multiple stuck-at fault.

The most common failures in the MOS technology are: open connections, shorts between connections, and defects affecting the transistor's gate oxide. The logical model of an open connection, previously modeled as a stuck-open fault, was shown in this paper to behave as a single stuck fault at a transistor's gate; the logic network was modeled at the transistor level, and the line capacitance was modeled as a memory element.

I. Introduction

The single stuck-at fault is presently the most common assumption in test generation and fault simulation because it is considered to model the logical behavior of the most frequently occurring physical failures. However, it has been shown by many writers that certain physical failures are not detected by a test set generated under the assumption of the single stuck-at fault and some solutions for solving this problem have been proposed. The main disadvantage of these solutions is that they require non-trivial changes in the ways test generation and fault simulation are presently implemented, thus increasing excessively the memory space and computation time.

The terminology used in this paper is that proposed by Avizienis [1]. The fault-tolerant model of an information processing system could be viewed as comprising four universes: physical, logical, informational, and external. An undesired event in the physical universe is called a failure. Physical failures can cause undesired events in the logical universe where they are commonly called faults. A stuck-at fault is such an example; it models the behavior of many different types of physical failures. The undesired event that might be caused by a fault in the informational universe is an error. Finally, an error could cause a crash in the external universe. This paper focuses on the cause-effect relationship between the physical and logical universes; more specifically, it concentrates on the logical models of some of the most common physical failures.

Certain failures behave differently from the permanent single stuck-at faults and therefore are not necessarily detected by a test set generated under the assumption of a single stuck-at fault. Wadsack [2] describes such physical failures occurring frequently in the CMOS technology which, when present, transform a combinational network to a sequential network. To model such failures it is proposed [2] to replace each faulty gate by a sequential network comprising six gates and a latch. As a result, a test set developed under the single stuck-at fault assumption for the sequential network detects all single stuck-at and stuck-open faults in the original combinational network.

The main disadvantage of the solution proposed in the previous work is the need of many changes in the model of a logic network in order to simulate the failures described above. These changes are inefficient especially in the fault simulators which keep track of all faults in one simulation pass; every change in the network model increases memory space and the computation time.

A class of failures common to all technologies are the accidental short circuits between signal connections; their logical model is called by Mei [3] a bridging fault. The logical behavior of a short circuit is modeled in the previous work as either a "symmetrical" wired-AND or a wired-OR function. However, the logical behavior of short circuits in the MOS technology cannot always be modeled as proposed in [3].

An objective of this paper is to identify the most common failures occurring in the bipolar and MOS technologies. Another objective is to

This work was performed for the Jet Propulsion Laboratory, California Institute of Technology sponsored by the National Aeronautics and Space Administration under Contract NAS7-100.

*To be presented at the 1983 Test Conference (October 1983).

model the logical behavior of these failures in terms of single or multiple stuck-at faults. The results of this work are significant because it is shown that presently used fault simulators and test generators capable of simulating effectively stuck-at faults can be enhanced to cover the most frequently found failures in the current technologies.

II. Frequent Failures in the Bipolar Technology and Their Logical Models

A recent literature survey performed by the authors shows that the most prevalent failures in the bipolar technology are: (1) open connections, (2) shorts between connections, and (3) "piped" transistors with excessive emitter to collector leakage currents. The open connections are caused by either improper manufacturing or by environmental and electrical stress while a device is operating. The former failures are due to over-etching of metal lines, sharp step coverage, and defective via contacts. The latter are mostly the results of electromigration and electrocorrosion. Although there are different bipolar technologies such as TTL, ECL, IIL, STL, etc., we will limit our discussion only to a TTL technology used for implementing large computers [4]. An example of a small TTL network with a fanout net is illustrated in Fig. 1.

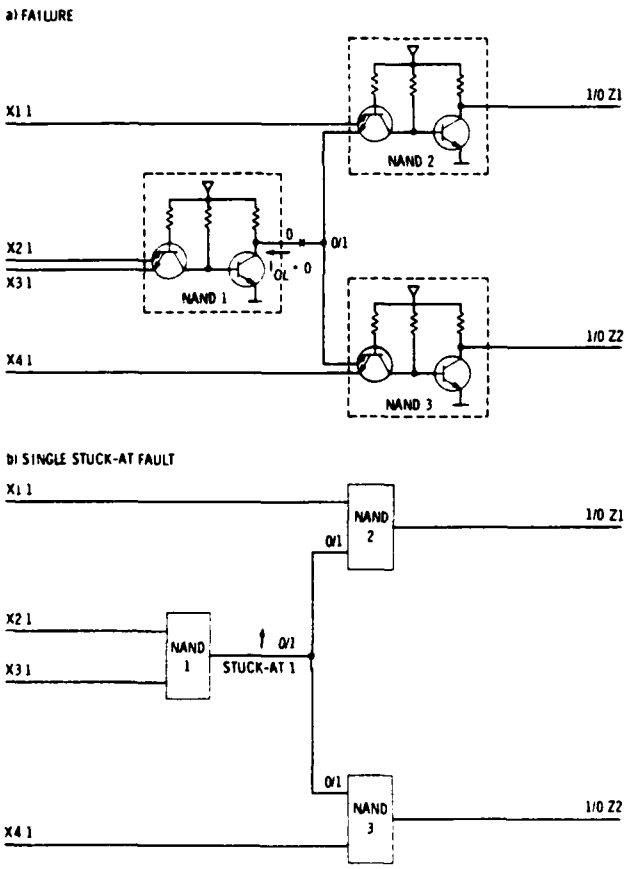


Fig. 1 An open fanout stem physical failure and the single stuck-at fault logical model

Let us assume first that the logical network appearing in Fig. 1a is free of faults. For the input vector shown in the figure, the output transistor of NAND1 is conducting, therefore, current coming from the inputs of NAND2 and NAND3 will be flowing into the output of NAND1. This current is commonly called output sink current or current flowing into the output when the output logical value is low (I_{OL}). The output sink current of NAND1 is the sum of the input source currents from the lower input of NAND2 and the upper input of NAND3. Thus, a logical 0 at the fanout branches is electrically equivalent to the inputs sourcing current. If an input of a TTL gate does not source any current, then the logical value at that input is 1. In the network shown in Fig. 1a, an open in the fanout stem stops the sourcing of current from the inputs connected to the fanout branches. Consequently, the logical value at the fanout branches, in the presence of an open fanout stem, is 1. From the above analysis, it can be concluded that an open fanout stem can be modeled as a stuck-at fault. Note that the stuck-at fault involves two operations: (1) the fanout stem is disconnected from the output of the gate, and (2) the fanout stem is permanently forced (stuck) to a logical value. Usually the disconnection is not explicitly shown; only the permanent forcing of the fanout stem is marked with an arrow pointing up or down for a stuck-at-1 or a stuck-at-0, respectively. The logical model of the open fanout stem failure is illustrated in the gate level diagram of Fig. 1b by an arrow pointing up for the fanout stem stuck-at-1. (The logical values of the lines which change in the presence of a fault are shown as good network slash faulty network.) Note that the effect of NAND1 output stuck-at-1 is detected at both primary outputs Z1 and Z2 for the input vector ($X_1 = 1$, $X_2 = 1$, $X_3 = 1$, $X_4 = 1$).

A second example of an open connection is depicted in Fig. 2. The presence of this open cuts off the flow of input source current from the upper input of NAND3 (Fig. 2a); thus the open input takes the logical value 1. The logical model of the open fanout branch is illustrated in Fig. 2b by an arrow pointing up, for a stuck-at-1, at the faulty fanout branch. Note that only the affected fanout branch takes the value 1 in the presence of an open connection, not the entire fanout net. The presence of the stuck-at-1 at the input of NAND3 is detected only at the primary output Z2 in the presence of input vector (1,1,1,1).

A failure affecting all bipolar transistors is called a "pipe" by Barson et al. [5]; it is identified by other authors [6-9] as probably the major yield detractor of current bipolar technology. A pipe is a microscopic short penetrating through the base from the emitter to the collector. Pipes appear electrically as an excessive leakage current from the emitter to the collector. Thermal oxidation of silicon is a frequent processing step in the bipolar technology which has an undesired effect causing dislocations in the crystalline structure of the silicon. During the fabrication step of the emitter some impurities diffuse through the shallow base region along these dislocations, therefore, establishing a low resistance path between the emitter and the collector. The

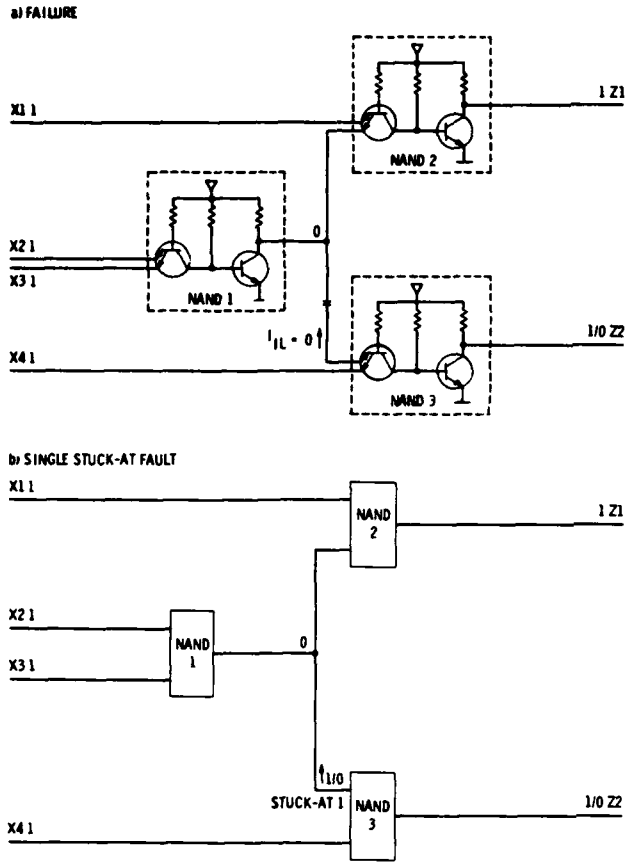


Fig. 2 An open fanout branch physical failure and its single stuck-at fault logical model

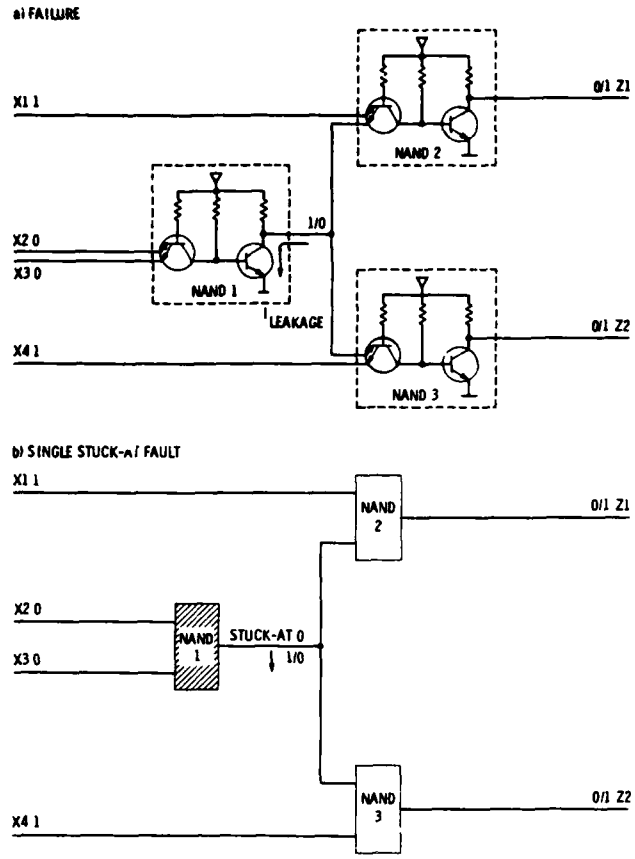


Fig. 3 Excessive current leakage in the output transistor caused by "pipes" and the corresponding single stuck-at fault logical model

effect of a piped output transistor is shown in Fig. 3. The excessive leakage in the piped output transistor of NAND1 (Fig. 3a) creates a sink current into the output of NAND1 although the piped transistor is off. As a result, the output of NAND1 can appear as a logical 0 instead of logical 1. The effect of the piped output transistor is observed at Z1 and Z2 in the presence of the input vector (1,0,0,1). The logical model of the piped transistor (Fig. 3b) is the output of NAND1 stuck-at-0.

A piped multi-emitter input transistor is shown in Fig. 4. In the presence of an excessive leakage current into the upper input of NAND3 (Fig. 3a), current flows from the lower input of NAND2 and from the output of NAND1 into the upper input of NAND3. As seen in the figure, the piped input transistor has no effect on the output of NAND3 because the lower input sources current thus keeping the output of NAND3 to 1 independent of the leakage current into the upper input. However, the lower input of NAND2 sources current in the presence of this failure; thus, if the leakage current is excessive it forces the lower input of NAND2 to a logical value 0. As a result, the logical effect of the piped input transistor of NAND3 is not observable at its own output but at the output of another gate (NAND2). (This behavior is similar to that of the "shorted input diode" described by Chang [10] in connection with a failure in the DTL technology.) The aforementioned pipe is detected only by the vector (1,0,0,0) at the primary output Z1. Our attempts failed to model this pipe by a single stuck-at fault. A possible solution is to model it by a multiple stuck-at fault: one at the output of NAND1 and another one at the lower input of NAND3 as illustrated in Fig. 4b. It is worth observing that the test pattern generated for the output of NAND1 stuck-at-0 by using the D-algorithm [11], requires that both paths through NAND2 and NAND3 to primary outputs Z1 and Z2 be sensitized, thus, assigning logical 1 to the top input of NAND2 and the lower input of NAND3. Under these conditions neither piped input transistors of NAND2 or NAND3 could be detected. However, if single path sensitization [12] is employed to generate a test vector for NAND1 output stuck-at-0 and if a path is purposely "desensitized" through NAND3, the lower input of NAND3 is at logical 0, thus, creating the right conditions for detecting not only the input piped transistor in NAND3 but also the output piped transistor of NAND1. This is a possible enhancement to the test generation algorithms to detect input or output piped transistors.

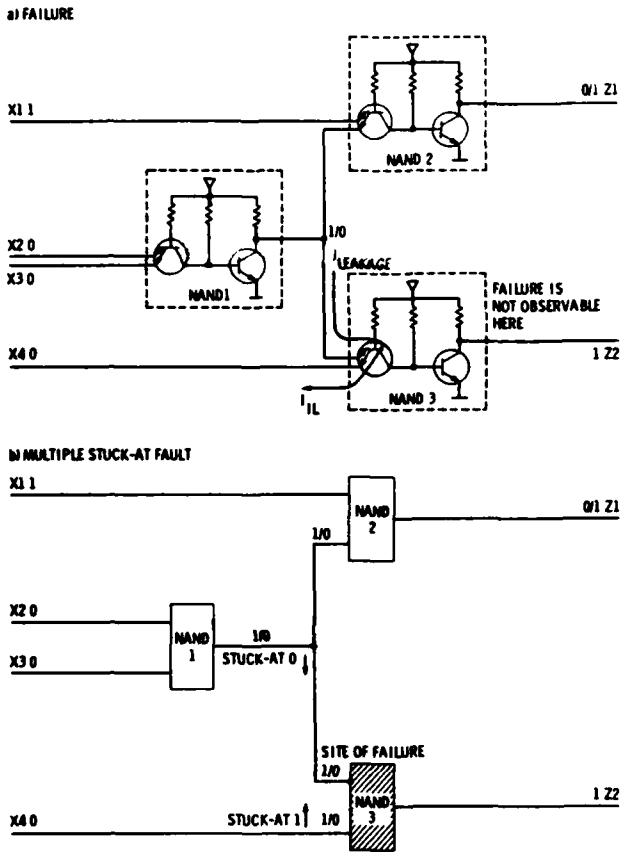


Fig. 4 A "piped" input transistor with excessive current leakage modeled as a multiple stuck-at fault

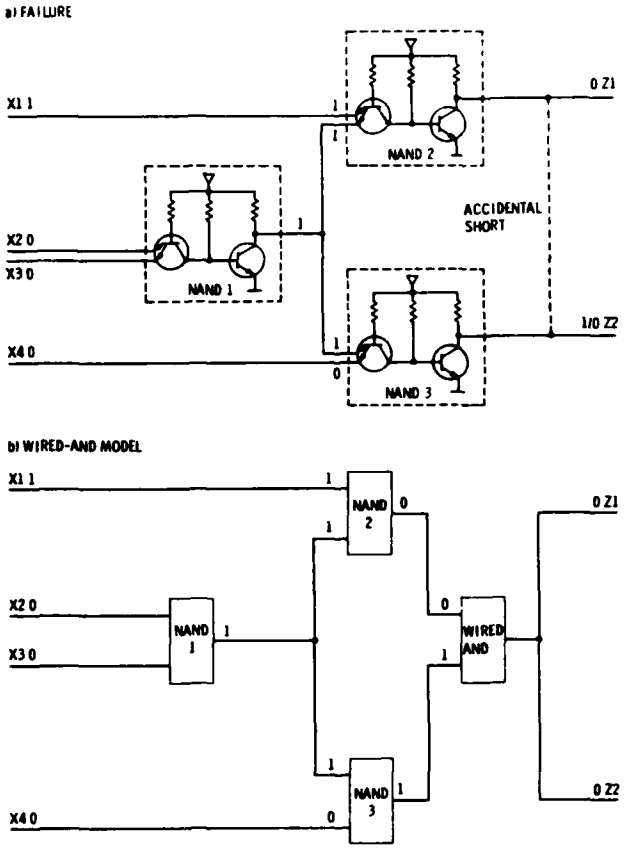


Fig. 5 A signal wire to signal wire short circuit failure and its wired-AND logical model

A short circuit failure between signal wires is illustrated in Fig. 5. For this particular bipolar technology the short circuit between the outputs of two gates behaves logically as a wired-AND. In other words, if one of the gates has its output transistor conducting, it pulls down the accidental short, or if both output transistors are not conducting the shorted outputs are pulled up.

III. Logical Models of MOS Failures

Failures of connections and of active components in MOS technology are caused both by faulty manufacturing steps and by electrical and environmental stress during the operation of a device. Connection failures of interest are either open signal lines or short circuits between signal lines. Open connections are the result of several manufacturing steps such as over-etching, poor contacts (metal to metal, metal to polysilicon, etc.) and sharp step coverage. Overstress induced open connections could be explained by electromigration, electrocorrosion, and electrostatic discharge. An open connection to the gate of a transistor has an unpredicted logical behavior unless additional factors are considered. An incomplete open can be modeled as an RC network (Fig. 6) in series with the connection to the gate which slows down the switching on and off of the affected transistor. This delay fault could become a static fault if we consider that the interface between the silicon and the gate oxide ($Si - SiO_2$) is highly unstable. Positive charges could be trapped in the gate oxide during fabrication or while the device is operational. A frequent positive charge is the sodium ion which in the early MOS technologies was one of the most widespread failure mechanisms. Several reliability studies performed almost a decade ago [13-16] show that more than 60% of MOS chip failures could be attributed to the positive charges, probably sodium ions, trapped in the gate oxide. Under high positive voltage and elevated temperature, these ions migrate to the $Si - SiO_2$ interface decreasing the threshold voltage of n-transistors and increasing the threshold voltage of p-transistors to the extent that they remain permanently on or off, respectively. Current MOS technologies appear to have eliminated the sodium ion contamination problems, however, the hot-electrons [17] and hot-holes [18] injection from the channel, source, or drain into the gate oxide are reported as affecting the reliability of current MOS transistors. An open gate connection in the presence of hot-holes in the gate oxide, or other positive charges, may permanently turn on an n-transistor as illustrated in Fig. 7. In

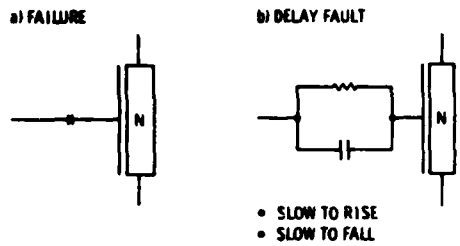


Fig. 6 An incomplete connection to a transistor's gate and its corresponding delay fault

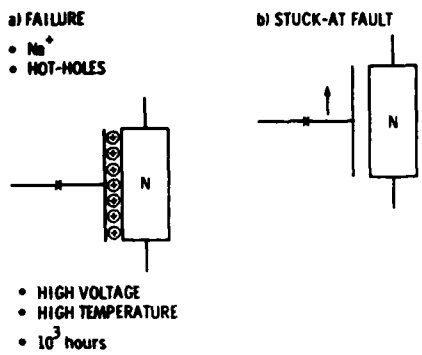


Fig. 7 An open connection to a gate of a transistor in the presence of positive charges in the gate oxide is modeled as a stuck-at fault

the presence of hot-electrons injected in the gate oxide, an n-transistor could be permanently turned off as depicted in Fig. 8. Ortner and Clemens [19] performed a reliability study of a RAM and found open gate connections due to incomplete etching of metal to polysilicon contact windows. Although the transistors had open gates, they operated correctly with the signal received through the capacitance of the faulty contact window until the hot-electrons were injected into the gate oxide permanently turning those transistors off. In summary, an open signal connection to the gate of an MOS transistor can be modeled

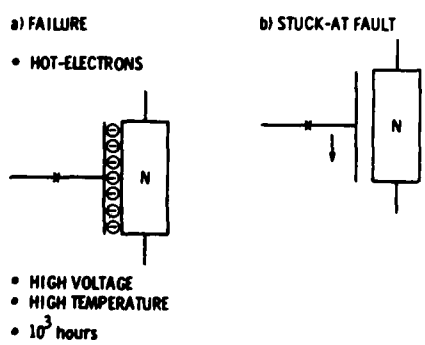


Fig. 8 An open connection in the presence of negative charges in the gate oxide and its stuck-at fault model

either as a delay fault (slow on, slow off), a stuck-at-0, or a stuck-at-1 fault at the transistor's gate.

Two other failures likely to occur in VLSI are shorts between drain and source due to punch-through or avalanche breakdown in narrow channels [20], and shorts between gate and channels [21,22] through a transistor's thin gate oxide. The former failure can be modeled logically as a stuck-at fault which keeps the transistor permanently conducting independent of the voltage applied at the gate (Fig. 9). The latter also behaves as a stuck-at fault forcing an enhancement mode transistor to turn off because the voltage at the gate is the same as that of the channel (Fig. 10). Note that the model of the short through the gate oxide forces an entire fanout net to the voltage level of the channel; it is modeled for the particular case illustrated in Fig. 10 as the fanout stem stuck-at-0.

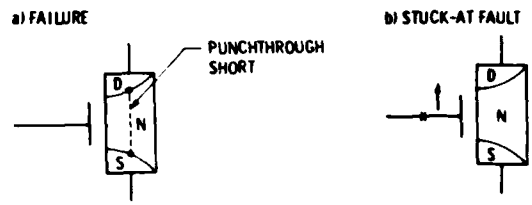


Fig. 9 A short circuit failure between the source and the drain of a transistor is modeled as a stuck-at fault

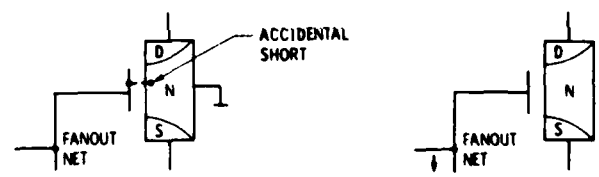


Fig. 10 An accidental short circuit between the gate and the channel of a transistor and its corresponding stuck-at fault model

A fault, called by Wadsack [2] stuck-open, not detected by a test pattern developed under the assumption of the single stuck-at fault is described in reference to Fig. 11. If the p-transistor connected to input X had an open connection to its gate, in the presence of positive charges in the gate oxide (Fig. 11a) it would be permanently off. The output capacitance in the presence of such a failure acts as a memory element and transforms the depicted simple combinational (NAND gate) logic to a sequential circuit. Note that such a failure can be detected only by the following unique sequence of two test vectors. The input vector (X = 1, Y = 1) turns on both n-transistors thus discharging C_L to logical 0. The subsequent vector (0,1), attempts to charge C_L but the output capacitor remains discharged because the faulty p-transistor does not turn on. Wadsack [2] modeled this failure with a sequential network comprising six gates and a data

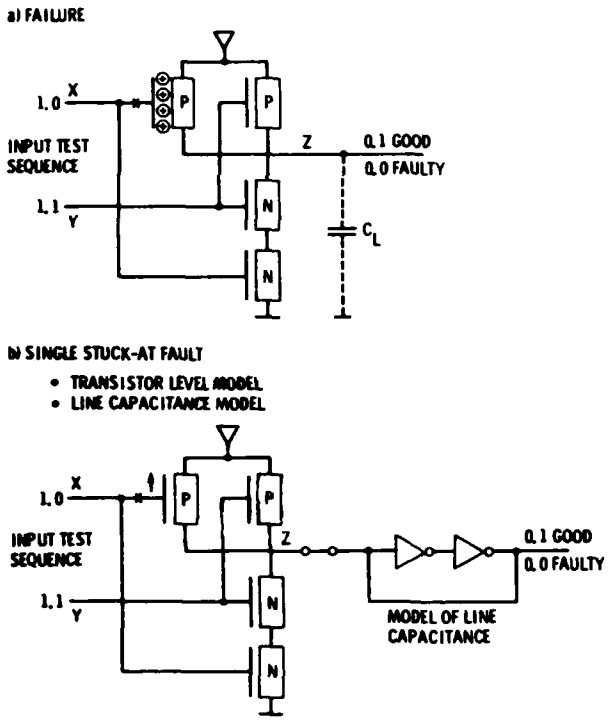


Fig. 11 A failure causing a transistor to remain permanently off could be modeled as a single stuck-at fault in a transistor network with the line capacitance modeled as a zero delay memory cell

latch such that current test generators and fault simulators capable of handling single stuck-at faults in logic gate level models could be used for developing tests for these kinds of failures. However, by taking into consideration present trends of modeling MOS logic at the transistor level [23,24] we propose the model of Fig. 11b for a NAND gate and its associated output capacitance. The latch at the output models the load capacitance; it is similar to those widely used in RAMs and could be assumed for our purposes as having zero delay. The advantage of such a model is that the single stuck-at fault assumption remains valid for covering faulty transistors permanently turned off. For example, a stuck-at-1 fault at the p-transistor connected to input X is necessary and sufficient to model the behavior of that transistor permanently off. As a result, current test generators and fault simulators capable of modeling logic networks at the transistor level could generate test vectors which detect such defective transistors under the assumption of single stuck-at faults. The model of the line capacitance can be eliminated if a transistor level network is simulated with special values which "remember" the previous logic level. Such an approach was proposed by Levendel et al [25] where Z_0 and Z_1 are additional simulation values for a floating wire which "remembers" the logic values 0 and 1, respectively.

Short circuits between two signal lines are frequent failures in MOS technology. We described previously in reference to the shorts in bipolar technology, that a short circuit can be modeled as a wired-AND function. An example of a short circuit in CMOS technology appears in Fig. 12, where the outputs of a NAND and a NOR gate are accidentally connected. The logical behavior of such a short is not as well defined as that in the aforementioned bipolar technology. The physical dimensions of the p and n transistors, together with the input vectors determined the logical value at the shorted output. Consider for example, that in Fig. 12 the input vectors are (X1 = 1, Y1 = 1) and X2 = 0, Y2 = 0). In the absence of the accidental short, the output Z1 = 0 because both n-transistors of the NAND gate are off and the output Z2 = 1 because both p-transistors of the NOR gate are on. If the accidental short is present, then it is possible that Z1 = 0 will dictate the logical value of the short because the resistance of the two conducting n-transistor in series is usually smaller than the resistance of the two conducting p-transistors in series. (However, it is also possible that the voltage of the short lies in between the logical levels). Another case in which Z2 = 0 dictates the logical value of the short could occur when (Z1 = 0, Y1 = 1) and X2 = 1, Y2 = 1). In this case, Z1 is pulled up only by a p-transistor while Z2 is pulled down by two n-transistors in parallel, thus, the probable logic value of the shorted outputs is 0. The following situations summarize the possible outcomes with a determined logic behavior of the short in our example:

1) If Z1 is 0, then it forces Z2 to 0
2) If Z1 is 1, then it forces Z2 to 1
3) If Z2 is 0, then it forces Z1 to 0
4) If Z2 is 1, then it forces Z1 to 1

Accordingly, a short circuit failure can, in general, be modeled in four different ways, as illustrated in Fig. 13. The accidental short in Fig. 13a could occur between the outputs of any two gates. Fig. 13b, depicts the model of

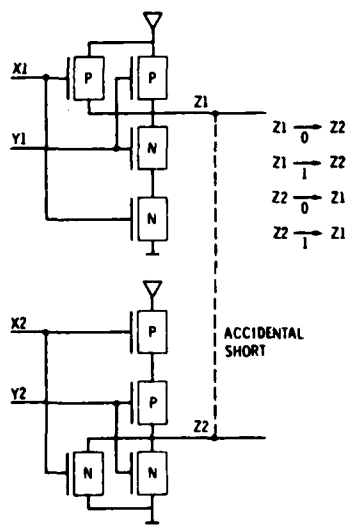


Fig. 12 A short circuit failure between the outputs of two CMOS gates

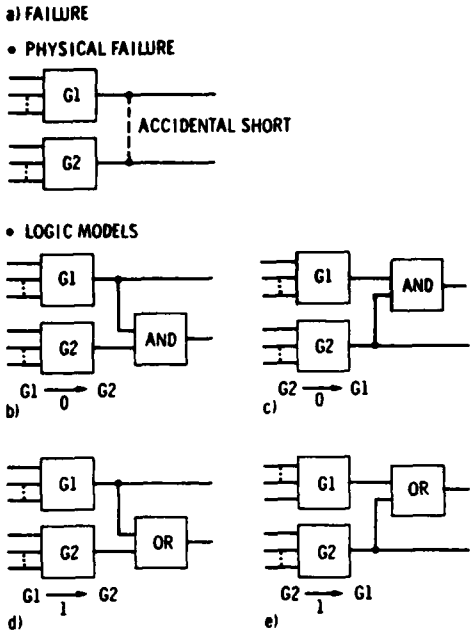


Fig. 13 A signal-wire to signal-wire short circuit failure and its logical models

a short in which G1 output at logic 0 forces the output of G2. It is also possible that when G2 is 0 it could force the output of G1 to 0 as shown in Fig. 13c. These two models are not the same as the wired-AND model of a short in the aforementioned TTL technology because only one output changes its behavior in the presence of the short, not both as in the case of the wired-AND. We propose to call the models of Figs. 13b and c asymmetrical wired-AND and the model Fig. 5b asymmetrical wired-AND. Two other possible models of a short in the MOS technology are the asymmetrical wired-OR illustrated in Figs. 13d and e. In summary, an accidental short circuit failure between any two MOS gates could be modeled in four different ways as asymmetrical wired-AND and wired-OR functions.

IV. Conclusions

This paper shows that most physical failures are detected by a test set developed under the assumption of single and multiple stuck-at faults. The logical network could be modeled at gate level for certain bipolar technologies and it should be modeled at transistor level for MOS technologies. The memory effect of the latter technologies could be modeled either explicitly with a memory element or with additional simulation values which remember the previous state of a connection line. Short circuits between signal wires are the only physical failures that cannot be modeled as permanent stuck-at faults. It was shown that MOS shorts cannot be modeled as symmetrical wired-AND or wired-OR. A more general model of a short was introduced as an asymmetrical wired-AND and wired-OR.

The failures modeled in a particular TTL technology were: open connection, short circuit between two signal wires, and piped transistor with excessive current leakage. The model of an open is a single stuck-at fault, the short behaves as a symmetrical wired-AND, the output piped transistor as a single stuck-at fault, and the piped input transistor as a multiple (two) stuck-at fault. It is significant to note that piped transistors in a network with fanout are tested for maximum leakage current if it is possible to sensitize only one path from a fanout stem to a primary output and desensitize all other paths from that stem.

A literature survey done by these writers identified some of the most prevalent failures in the MOS technologies as open connections, short circuits between source and drain. The open connection failure of an MOS transistor's gate taken alone has an undetermined behavior. However, it was shown that such an open in the presence of positive or negative charges in the gate oxide could be modeled as a stuck-at fault at the transistor's gate. The failures of an MOS transistor such as hot-electrons, hot-holes, positive charges in the gate oxide (Na^+), punch-through or avalanche breakdown induced shorts between source and drain, and shorts of the gate to the channel are all modeled as stuck-at faults at the faulty transistor's gate.

A failure which turns off permanently an MOS transistor was modeled in previous works as a special fault called stuck-open. It is shown here that such a failure could be modeled by a single stuck-at fault if the logic network is described at transistor level and the memory effect of the line capacitance is modeled either as a memory element or with special simulation values. These simulation values remembers the value of a line when the output of a gate is switched to high impedance.

The results show that many frequently occurring failures in current technologies behave as single and multiple stuck-at faults. Consequently, the test generators and fault simulators presently in use and capable of handling single stuck-at faults could be enhanced to simulate multiple stuck-at faults. By contrast, previous works recommended complex models.

VII. Acknowledgement

The authors wish to acknowledge the survey of papers performed with the help of R. Corder. This work was performed under the management of W. R. Scott of JPL, H. Sonemman of NASA, and P. Losleben of DARPA.

VIII. List of Acronyms

CMOS: Complementary MOS
DTL: Diode-Transistor Logic
ECL: Emitter Coupled Logic
IIL: Integrated Injection Logic
ISL: Injection Shotky Logic
MOS: Metal-Oxide-Semiconductor
NMOS: N-Transistor MOS
RAM: Random Access Memory
TTL: Transistor-Transistor Logic

IX. References

[1] A. Avizienis, "The Four-Universe Information System Model for the Study of Fault-Tolerance," Proc. 12th International Symposium on Fault Tolerant Computing, pp. 6-13, 1982.

[2] R. L. Wadsack, "Fault Modeling and Logic Simulation of CMOS and MOS Integrated Circuits," The Bell System Technical Journal, pp. 1449-1474, May-June 1978.

[3] K. C. Y. Mei, "Bridging and Stuck At Faults," IEEE Trans. on Computers, Vol. C-23, No. 7, pp. 720-727, July 1974.

[4] A. H. Dansky, "Bipolar Circuit Design for VLSI Gate Arrays," International Conference on Circuits and Computers, pp. 674-677, Oct. 1980.

[5] F. Barson, "Emitter-Collector Shorts in Bipolar Devices," IEEE Journal of Solid State Circuits, Vol. SC-11, No. 4, pp. 505-510, Aug. 1976.

[6] McD. Robinson et al., "Orientation and Implantation Effects on Stacking Faults During Silicon Buried Layer Processing," Journal of the Electrochemical Society, Vol. 128, No. 4, pp. 926-928, April 1981.

[7] J. A. Topich, "Reduction of Defects in Ion Implanted Bipolar Transistors by Argon Back Side Damage," Journal of the Electrochemical Society, Vol. 128, No. 4, pp. 866-870, April 1981.

[8] M. C. Chen and V. J. Silvestri, "Pre- and Post-Epitaxial Gettering of Oxidation and Epitaxial Stacking Faults in Silicon," Journal of the Electrochemical Society, Vol. 128, No. 2, pp. 389-395, Feb. 1981.

[9] J. F. Casey et al., "Piped Emitter Delineation Using the Wright Etch," Journal of the Electrochemical Society, Vol. 129, No. 2, pp. 354-357, Feb. 1982.

[10] H. Y. Chang, "A Method for Digitally Simulating Shorted Input Diode Failures," The Bell System Technical Journal, pp. 1957-1966, July-Aug. 1969.

[11] J. P. Roth, "Diagnosis of Automatic Failures: A Calculus and a Method," IBM Journal of Research and Development, Vol. 10, pp. 278-291, July 1966.

[12] D. B. Armstrong, "On Finding a Nearly Minimal Set of Fault Detection Tests for Combinational Logic Nets," IEEE Trans. on Electronic Computers, Vol. EC-15, pp. 66-73, Feb. 1966.

[13] G. L. Schnable et al., "Reliability of CMOS Integrated Circuits," Computer, pp. 6-17, Oct. 1978.

[14] R. C. Maurer, "CMOS Life Suitability Evaluation Program," Report No. RADC-TR-79-167, Oct. 1979.

[15] H. C. Rickers, "Microcircuit Device Reliability: Memory/LSI Data," DoD Information Analysis Center, Reliability Analysis Center Report No. MDR-7, Winter 77-78.

[16] G. M. Johnson, "Voltage Stress Effects on Microcircuit Accelerated Life Test Failure Rates," NASA Contract No. NAS8-3117, Aug. 1976.

[17] N. Shiono and C. Hashimoto, "Threshold-Voltage Instability of n-Channel MOSFETs Under Bias-Temperature Aging," IEEE Trans. on Electron Devices, Vol. ED-29, No. 3, pp. 361-368, Mar. 1982.

[18] H. Gesch et al., "Generation of Interface States by Hot Hole Injection in MOSFETs," IEEE Trans. on Electron Devices, Vol. ED-29, No. 5, pp. 913-918, May 1982.

[19] W. R. Ortner and J. T. Clemens, "A Reliability Failure Mode in Dynamic MOS Circuits - the Unopened Metal to Polysilicon Contact Window and the Floating Gate Transistor," Proc. 1977 Reliability Physics Symposium, 1977.

[20] B. Eitan and D. Frohman-Bentchkowsky, "Surface Conduction in Short-Channel MOS Devices as a Limitation to VLSI Scaling," IEEE Trans. on Electron Devices, Vol. ED-29, No. 2, pp. 254-266, Feb. 1982.

[21] C. A. Goodwin and J. W. Brossman, "MOS Gate Oxide Defects Related to Treatment of Silicon Nitride Coated Wafers Prior to Local Oxidation," Journal of the Electrochemical Society, Vol. 129, No. 5, pp. 1066-1070, May 1982.

[22] C. W. Green, "PMOS Dynamic RAM Reliability - A Case Study," Proc. 17th Reliability Physics Symposium, pp. 213-219, April 1979.

[23] R. E. Bryant, "MOSSIM: A Switch Level Simulator for MOS LSI," Proc. 18th Design Automation Conf., pp. 786-790, June 1981.

[24] A. K. Bose et al., "A Fault Simulator for MOS-LSI Circuits," Proc. 19th Design Automation Conference, pp. 400-409, 1982.

[25] Y. L. Levendel et al., "Accurate Logic Simulation Models for TTL Totempole and MOS Gates and Tristate Devices," The Bell System Technical Journal, Vol. 60, No. 7, pp. 1271-1287, Sept. 1981.

2.7 PROCUREMENT TECHNOLOGY

The procedures for procuring reliable, custom circuits from silicon foundries may require the modification of current circuit procurement practices. It is clear that the traditional approaches involving circuit burn-in and accelerated life testing must be supplemented with other approaches such as test-chip evaluation. The integration of the traditional and the test-chip approach is illustrated in Figure 2.7.-1. As seen at the very bottom of the figure, the circuit qualification is achieved not only after the circuit passes packaged circuit tests but also after circuit burn-in, accelerated-life test, and test-chip analysis. Some or all of the functions shown in the figure could be carried out by the foundry.

The test chip can serve a number of functions. If the foundry is responsible for developing and merging a wide variety of circuits for its wafer production, test chips can serve as the vehicle for internal process control or for providing product assurance information to the purchasing quality assurance activities, supplementing the circuit performance verification tests that may be required by the specifications. In the event that the foundry is used strictly to manufacture a range of circuits designed out-of-house and merged into a wafer fabrication process, the test chips can validate the process. This mode of operation would divorce the responsibility for proper performance of the circuit design from the foundry.

The sequence of events begins with the definition of the functional and environmental requirements.

Circuit design evaluation tools such as:

1. Layout rule checker
2. Circuit simulator (timing)
3. Logic simulator
   a. Functionality
   b. Fault simulation
   c. Test vector generation
4. Logical-to-physical check

Circuit design constraints such as:

1. Layout rules
2. Timing
3. Noise margins
4. Power dissipation
5. Testability
6. Reliability (physical)

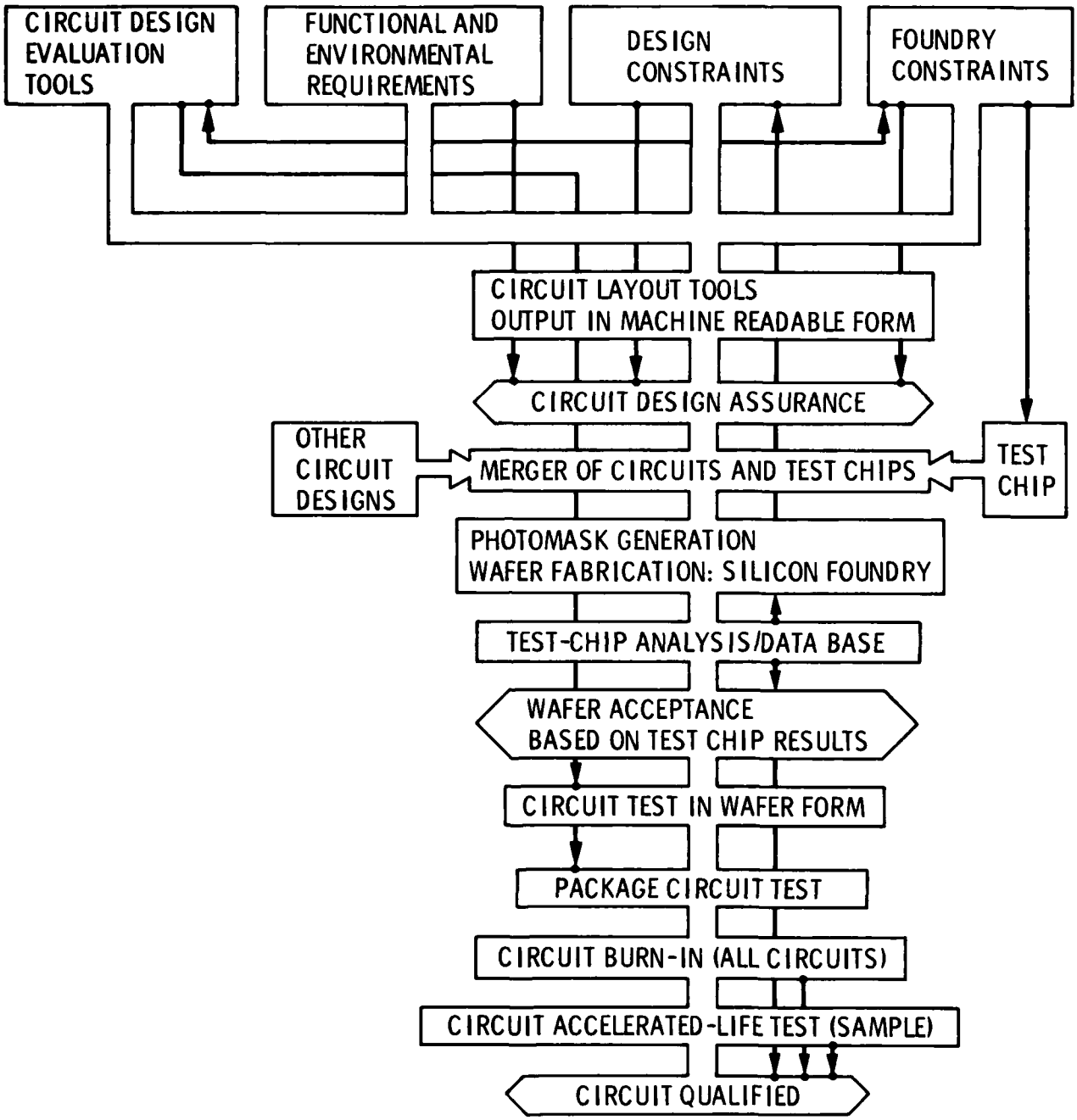


Figure 2.7-1. Flow diagram describing the procurement of reliable, custom circuits from silicon foundries.

7. Reliability (fault models)

8. Radiation

Foundry constraints such as:

1. Technology

2. Mask levels

3. Layout rules

4. Electrical parameters

As noted in Figure 2.7-1, some of the parameters needed to define the above elements must be derived from test chips fabricated on a previous run. This brings up a crucial point. Certain critical parameters, which are called out in the wafer fabrication specifications, must not change from run to run or circuits may not perform to specification. At this time the critical parameters have not been identified.

The circuit layout process interacts with the above elements and the output is in a machine-readable form. For JPL this means CIF (Caltech Intermediate Form).

A critical step involves circuit design assurance. As indicated in Figure 2.7-1, three of the starting elements are involved. Several critical steps are the logical-to-physical checker and design for testability verification.

At this point the circuit is ready to be submitted to the silicon broker, who performs the merge of the circuits and test chips. The merge consists of the following steps:

1. Chip placement algorithm (based on yield)

2. Creation of nonCIF layers

3. Feature bloating and shrinking

4. Housekeeping (sign of levels, labels)

The output of the silicon broker is in a machine-readable form suitable for photomask generation. Once the photomasks are created (currently with an electron-beam pattern generator), they are inspected for geometry control and defect density.

The photomasks are then submitted to the silicon foundry for wafer fabrication. Based on results obtained from test chips, the wafer fabrication process is accepted or rejected. At present, the specifications consist of a few transistor parameters measured at a few (approximately five) sites on each wafer. As work on the test chip methodology progresses, the specification will be more extensive and will include parameters that evaluate the reliability of the wafer fabrication process. Note that results from the test-chip analysis

go into a data base that is used in the final circuit qualification and is used in the next circuit design effort.

Circuits are given a preliminary electrical test in wafer form to identify nonfunctional circuits and save the cost of packaging. Circuits are given a more rigorous electrical test once they are packaged. All the circuits undergo circuit burn-in, and a sample of the circuits receive accelerated-life tests. Circuits are then qualified from the results obtained from the test-chip analysis, circuit burn-in, and accelerated-life test steps. It remains for future efforts to decide how much of the qualification process can be assumed by the test chip.

2.8 TECHNOLOGY TRANSFER

Conference Presentations:

1. Buehler, M. G., T. W. Griswold, C. A . Pina, and C. Timoc, "Microelectronic Test Chips for Obtaining Reliable, Custom Integrated Circuits," Microelectronic Measurement Technology Seminar, San Jose, CA (March 23, 1982). Also presented at Stanford University, Electrical Engineering Colloquium November 5, 1982.

2. Buehler, M. G., "Process Controls," Computer-Aided Manufacture for Semiconductor Fabrication Applications Short Course, sponsored by U. C. Berkeley, Palo Alto, CA (January 18, 1983).

3. Buehler, M. G., "Microelectronic Test Chips for VLSI Electronics," Semiconductor Material and Device Characterization Short Course, sponsored by Arizona State University, Tempe, AZ (February 18, 1983).

Publications:

1. Buehler, M. G., T. W. Griswold, C. A. Pina, C. Timoc, "Test Chips for Custom ICs: Six Kinds of Test Chips," Circuits Manufacturing, 22, 36-42 (June 1982).

2. Griswold, T. W., "Portable Design Rules for Bulk CMOS," VLSI Design, III, 62-67 (September/October 1982).

In addition to the above formal transfer of this work, we have presented seminars on test chips to local industry and have provided tours of our test system to numerous engineers and scientists.

SECTION 3

DISCUSSION AND FUTURE OUTLOOK

At the end of the contract period (September 30, 1982) our test-chip tester had been operational for about six months. During this period we concentrated on writing the software to test the CMOS-bulk wafers received during the summer of 1982. In the future we will expand the capability of the test-chip tester in a number of directions:

1. System calibration. Procedures will be established for periodic instrument calibration, and techniques will be developed for wafer-level calibration through the use of known components.

2. Wafer-level reliability assessment. Wafer-probing equipment will be modified (e.g., hot chucks and large count multiprobes) to allow the voltage and temperature stressing of test structures in wafer form.

3. Automated test program development. The preparation of test software is a labor-intensive activity. The preparation of test software will be made more efficient by taking advantage of the modularity of test structures and the pad location information contained in the test-chip layout software.

The test structures needed to evaluate a CMOS-bulk process are in various states of development. We have had good success in evaluating the sheet resistance, linewidth, and line spacing of conducting layers with the split-cross-bridge resistor. Also, pinhole density and conductor bridging evaluations can be done with confidence using the pinhole array capacitor and the comb resistor test structures, respectively.

But a host of other measurements are poorly developed. For instance, we can accurately measure transistors threshold voltages using the max-slope approach, but this technique is time-consuming. Certain layout rules cannot be measured with the split-cross-bridge resistor and collision resistors must be developed to handle these cases. Also, the characterization of the n-substrate resistivity and the metal to n-substrate contact resistance requires a special test structure. The peripheral capacitance of diffused regions seems to be easy to characterize using structures with different periphery to area geometries. But the measurement of the gate overlap capacitance appears to require more advanced techniques.

We have analyzed two CMOS-bulk wafer fabrication runs and have discovered that results from one wafer can be very different from those from the next wafer. These results have important implications concerning the placement of test strips and test chips on wafers. Also, the toggling of inverters appears to a good first structure to screen the wafers. Also, the nonprime sites, located around the periphery of the wafers, should be considered for the placement of yield analysis test structures, which require a lot of area. In the future we will continue to evaluate CMOS-bulk silicon foundry wafers to provide definitive answers to test strip and test chip placement questions.

Results from our fault modeling studies indicate that a good number of faults fall into the stuck-at fault category. This is good news because stuck fault testing is relatively easy to do. But certain faults, e.g., the stuck-open pull-up transistor, is much more difficult to evaluate and characterize. We intend to evaluate this fault using an experimental approach involving a matrix of faulted NAND gates.

SECTION 4

APPENDIX

Chapter 9

Microelectronic Test Chips for VLSI Electronics

MARTIN G. BUEHLER

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

I. Historical Perspective 529
II. Introduction 535
A. Test-Chip Measurement Environment 537
B. Objections to the Use of Test Chips 538
III. Types of Test Structures 538
A. Test Structures for Device Parameter Extraction 540
B. Test Structures for Layout Rule Checking 545
C. Test Structures for Process Parameter Extraction 547
D. Test Structures for Random Fault Analysis 549
E. Test Structures for Reliability Analysis 551
F. Test Structures for Circuit Parameter Extraction 554
IV. Test-Chip Organization and Test-Structure Design 557
A. Test-Chip Organization 557
B. Test-Structure Design Rules 561
V. Test-Chip Testers and Advanced Test Structures 565
VI. Future Directions 569
A. Data Acquisition 570
B. Data Reduction 570
C. Vendor Transactions 572
Appendix 572
References 574

I. HISTORICAL PERSPECTIVE

The microelectronic test chip is an ancillary test device that is manufactured along with product circuits on wafers. It is composed of numerous devicelike test structures that are measured by a variety of means to obtain

Copyright © 1983 by Academic Press, Inc.
All rights of reproduction in any form reserved.
ISBN 0-12-234106-6

information that is difficult, if not impossible, to obtain from product circuits. The test structures are designed to provide a rapid analysis of a specific portion of the wafer fabrication process. Traditional failure analysis techniques, when applied to a product circuit, are time consuming and hence expensive tasks. The use of test chips is expected to enhance the reliability of the final product and reduce its cost.

Test chips have been used by the integrated circuit industry since its beginnings in the early 1960s. In general, test chips have been used for component characterization, manufacturing process control, equipment and operator performance evaluation, look-ahead indicators of circuit yield, and new wafer fabrication process evaluation. In the early days test chips were rarely used in the purchase or rejection of circuits or manufacturing equipment. As explained later, this situation is changing.

An early test chip was described in the 1968 work of Barone and Myers [1]. They developed a test chip (1.3 mm × 3.0 mm) with 44 probe pads to aid in the evaluation of a bipolar 8-bit adder circuit that contained 448 components. The test chip consisted of the following eleven test structures: a bipolar transistor, base resistor, base-under-emitter resistor, step-coverage resistor, metal-sheet resistor, base-collector diode with buried layer, buried-layer resistor, metal–semiconductor contact resistor array, metal–metal contact resistor array, multiemitter transistor, and MOS capacitor. These test structures were used in device characterization, process control, and circuit reliability evaluation. The number of elements (100) in the contact resistor array structures was remarkably large compared with other contemporary test structures where the number was generally much less.

In this same time period, two other test chips are noteworthy. In 1969 Schnable and Keen [2] describe a test chip for monitoring LSI reliability life aging. Their chip was designed to allow the measurement of first-level to second-level metal contact resistance, dielectric pinhole density and breakdown, metal step coverage, metal sheet resistance, and the resistance of an array of diffused resistors. In 1970 Sahni [3] described a chip for evaluating the reliability of bipolar integrated circuits. His structures examined the leakage current of transistors, integrity in the metallization, moisture resistance of the passivation layer, integrity of the bonds, and resistance of the metallization. Both test chips [2, 3] were packaged and subjected to thermal stress tests.

A test chip that is representative of this era is shown in Fig. 1. It was described in 1972 by Penney and Lau [4] and was developed for aluminum-gate PMOS integrated circuits having about 6000 individual transistors. The test chip (~0.9 mm × 1.3 mm) had 200 probe pads and was designed as a reliability evaluation device. It contains seven electrically testable test structures as listed in Table I.

Back then Penney and Lau [4] were rather optimistic about the use of test chips.

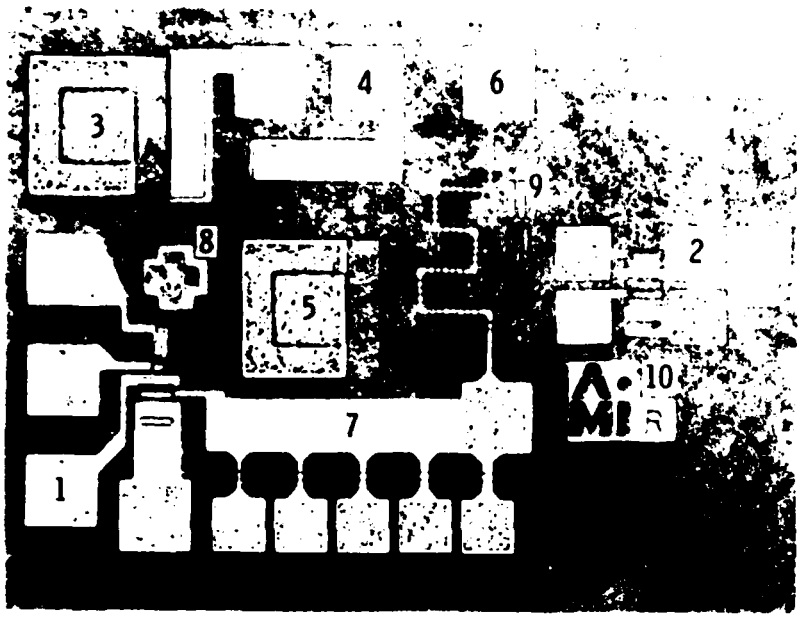


Fig. 1. An early test chip (0.9 mm × 1.3 mm) for use in evaluating a PMOS circuit wafer fabrication process [4]. Test structures are listed in Table I. (Used with permission.)

TABLE I

Test Structures for the Test Chip Shown in Fig. 1

Number	Test structure	Comment
1	Inverter	
2	Input transistor and field threshold transistor	
3	Large P–N junction diode with enhancement gate	
4	MOS capacitor	Identical in size to the enhancement gate of structure 3
5	Large P–N junction diode	Identical in size to structure 3 but without enhancement gate
6	Metal step-coverage resistor	Number of steps = 40
7	Metal resistors with different linewidths	Used to evaluate metal thickness
8	Alignment marker	
9	Etch control structure	
10	Logo	

> If such a device (test chip) is included with production circuits, it is valid to conclude that good performance of the test chip implies good working circuits. There is merit to being able to evaluate a standard circuit (test chip) on a daily basis rather than to randomly evaluate different types of circuits where data comparison is meaningless.

These views are just now beginning to be accepted by the semiconductor industry.

The above test chips were intended for internal use by chip manufacturers. The first test chip used to accept or reject wafers was described in 1974 by Reynolds *et al.* [5]. Their chip (2.54 mm × 2.54 mm) contained 35 test structures and 35 probe pads. It was intended to validate the layout rules, wafer-fabrication process, and reliability of aluminum-gate PMOS integrated circuits. The purpose of the test chip was not to control fabrication practices but merely to assure that the process was under control.

Early test chips had limited usefulness, for they were not comprehensive and were not designed for automatic wafer probing. In addition, they were usually designed as an afterthought to a circuit design effort. The situation is summed up nicely by Tingley and Johnson [6].

> Process development (test) chips usually result from hurried efforts by process development engineers with limited experience in actual circuit design or layout. The resulting designs tend to be highly idiosyncratic, with numerous omissions which any experienced circuit or layout designer, MOS physicist, or reliability specialist would readily notice. Usually poorly designed for automatic testing, process development chips often tie up one or more highly paid technical people for weeks doing manual probing, testing, and data reduction.

Recently a comprehensive test structure was developed by Ham [7] for characterizing SOS technology. His chip (6.6 mm × 6.6 mm) contained 175 test structures and 1250 probe pads. This chip has a very large number of probe pads and requires 21 probe cards to access all the electrical test structures.

Another present day test chip developed by Mitchell and Linholm [8] is shown in Fig. 2. This test chip (5.1 mm × 5.1 mm) has 216 probe pads and 40 test structures, which can be accessed by one probe card. These structures are listed in Table II. A comparison of this list with the structures listed in Table I provides a measure for the progress that has been made in the past 10 years. For instance, the structures shown in Fig. 2 have been arranged to allow for automatic data acquisition. The pattern is probe-pad intensive so that test structures are electrically isolated from each other. Finally, the chip area is 25 times larger so that a statistically significant number of steps can be included in the step-coverage resistors (structures 1–3). The number of steps (115,200) included in the structures on the NBS-28A test chip is con-

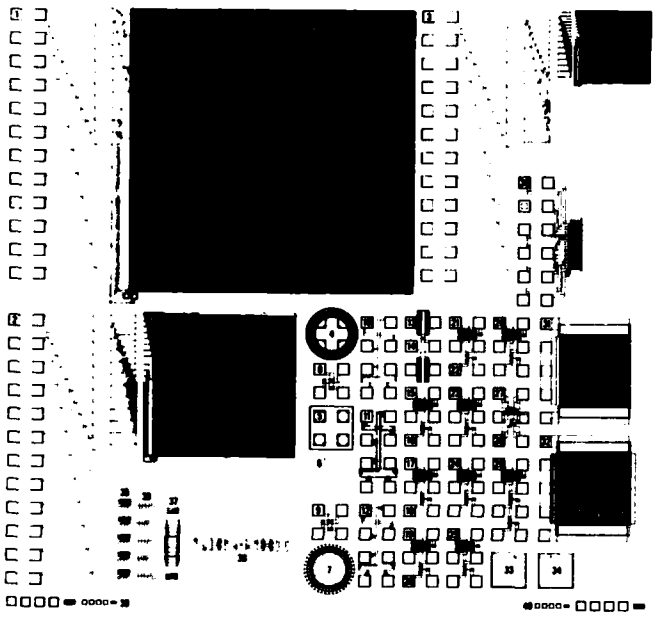


Fig. 2. Comprehensive test chip (5.1 mm × 5.1 mm) NBS-28A for analyzing either an NMOS or PMOS self-aligned polysilicon gate, junction isolated process [8]. Test structures are listed in Table II.

siderably more than the 40 steps seen in structure number 6 of Fig. 1. Such large test structures are necessary in order to provide more than a go/no-go test and to properly emulate today's complex circuits which can have one-half million transistors [9].

In recent years there has been a growing awareness of the vital role test chips can play in the manufacture of electronic devices. This awareness can be seen in the commercial availability of multifunction parametric testers [10], which greatly facilitate the acquisition of test-chip data. Professional societies have devoted special sessions to this topic such as the process/device monitors session at the Semiconductor Silicon 1981 Symposium [11]. Finally, since 1974 [12] the National Bureau of Standards has made and is making significant contributions to test-chip methodology [8].

To date, test chips have been most extensively used within circuit fabrication facilities. Typically a wafer found in a high-volume circuit manufacturing facility will appear as seen in Fig. 3 where 8 test chips are interspersed among 550 identical circuits.

Test chips are now being used more frequently in the acceptance or rejec-

TABLE II

Test Structures for Test Chip NBS-28A [8] Shown in Fig. 2[a]

Number (#)	Test structure	Comments
1, 2, 3	Metal step-coverage serpentine resistor over polysilicon grid	Linewidth and spacing: #1 = 4 μm, #2 = 2 μm, #3 = 1 μm. Number of steps = 115,200.
4, 5	Polysilicon-gate oxide–silicon capacitor	Area = 829 μm². #4 = round. #5 = square.
6, 7	Metal-field oxide–silicon capacitor	Area = 829 μm². #6 = square. #7 = round.
8, 9	Contact resistor	Contact area = 64 μm² (square). #8 metal-implant, #9 = metal–polysilicon.
10, 11, 12	Cross-bridge resistors	Linewidth = 16 μm. #10 = metal, #11 = polysilicon, #12 = implant.
13–22, 27–28	Field effect transistors (3-terminals) #14 = metal gate, field oxide. All others = polysilicon gate, gate oxide. #27 and 28 have all dimensions reduced, whereas #17–#22 have only W and L reduced relative to #15 and #16.	# / W (μm) / L (μm) / Reduction 13 / 160 / 28 / — 14 / 160 / 26 / — 15 / 120 / 8 / — 16 / 12 / 8 / — 17 / 60 / 4 / ½ 18 / 6 / 4 / ½ 19 / 30 / 2 / ½ 20 / 3 / 2 / ½ 21 / 15 / 1 / ½ 22 / 1.5 / 1 / ½ 27 / 15 / 1 / ½ 28 / 1.5 / 1 / ½
23–26	Inverter	#23 = #15 + #16 (L = 8 μm). #24 = #17 + #18 (L = 4 μm). #25 = #19 + #20 (L = 2 μm). #26 = #21 + #22 (L = 1 μm).
29	NAND gate	#29 = #15 + 2(#16) (L = 8 μm).
30	Multigate field-effect transistor (polysilicon gate)	L = 1, 1.5, 2, 2.5, 3, 3.5, 4, 6, 8 μm.
31, 32	Comb field-effect transistor	W = 14.5 mm, L = 8 μm, #31 + metal gate, #32 = polysilicon gate.
33, 34	Secondary ion mass spectrometry (SIMS) analysis area	#33 = implant, #34 = polysilicon. Area is a square 240 μm on a side.
35	Resolution structure for each photomask	Linewidth = 1, 2, 4, 8, 16 μm.
36	Etch control structure for each photomask.	Spacing = 0, 1, 2, 4, μm. Also 8 μm for window, contact, and passivation levels.
37	Surface profilometer structure	
38	Chip identification number	
39, 40	Photoresist mask alignment marks	#39 for positive photoresist and #40 for negative photoresist

[a] Photomask levels are implant, polysilicon, contact, metal, and passivation.

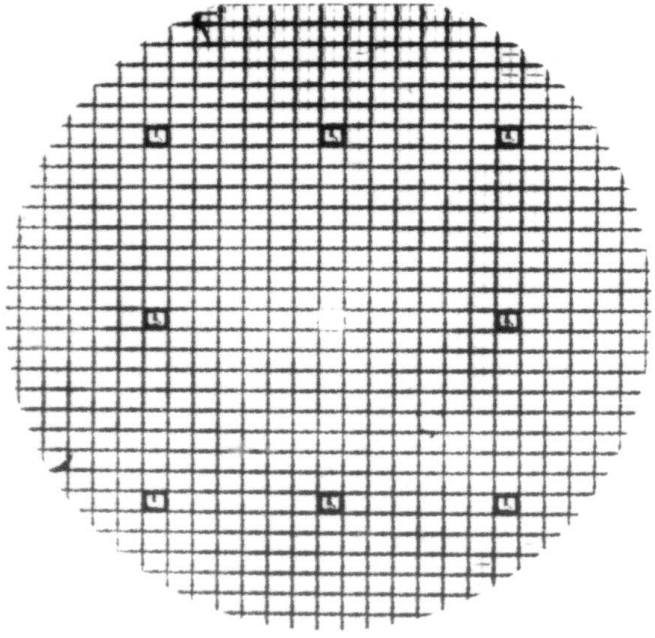


Fig. 3. High-volume integrated circuit wafer.

tion of wafer lots of custom-integrated circuits produced by "silicon foundries" [13] that specialize in the manufacture of custom circuits. This is not a new role for test chips, for the pioneering work of Reynolds *et al.* [5] used the same principle. What is new is the number of circuits found on a wafer. This is illustrated in Fig. 4 by a wafer that consists of 50 different circuits and 5 different test chips. In this multicircuit environment, it is no longer possible to accept or reject wafer lots based on circuit performance. Instead, lot acceptance is based on results obtained from test chips. In this environment it is essential that a test-chip methodology be perfected that will facilitate the buying and selling of custom circuits. This chapter describes the current state of the test-chip art and indicates where improvements are needed.

II. INTRODUCTION

Over the years a number of alternate names have been given to the test chip. Many of the names are indigenous to a particular company. The names

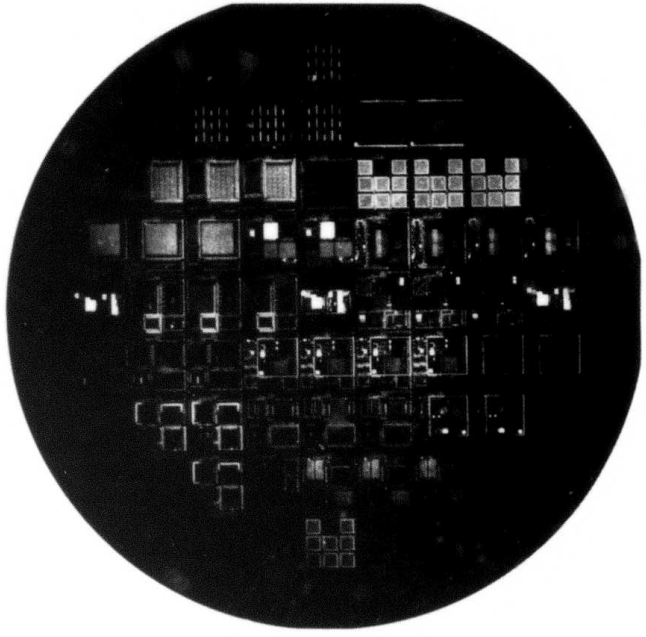


Fig. 4. Custom multicircuit wafer.

vary from test pattern, test coupon, test pellet, reliability test site, reliability evaluation device, process control bar, PCM (process control monitor), to a most descriptive name, "panic." The name panic is derived from the use of the test chip as a diagnostic in emergency situations. The name test pattern is used quite frequently, but this name conflicts with the meaning assigned by a test engineer who refers to a test pattern as a train of electrical pulses used to test a circuit. In recent years test chip seems to be the most universally accepted term.

The test structures discussed in this chapter are those that are measured by electrical methods at the end of the wafer-fabrication process. Test chips usually contain other large-area test structures that are measured by a variety of techniques. For example, this text does not discuss in-process tests where mechanical probes are used to measure bipolar transistor characteristics on partially fabricated wafers. Nor do we discuss the use of four-point probes to measure the sheet resistance of implanted layers, or optical techniques to measure the thickness of dielectric layers. In addition, various physical analysis techniques such as Auger, SEM, SIMS [14], or angle lap

and stain are not discussed. Many of these techniques are described in a special issue of *IEEE Transactions on Electron Devices* [15] and in the other chapters of this book.

A. Test-Chip Measurement Environment

The electrical test methods described in this chapter are restricted to those methods where test structures are measured in the dark in wafer form with an automatic multifunction tester in a near-room ambient. This restriction eliminates very low-current and high-frequency measurements; it eliminates measurements made far from room temperature and humidity; and it restricts the use of light to photon flooding where a burst of light is used to quickly pull a device out of deep depletion. The restriction is imposed in order to focus attention on the need to make rapid and statistically meaningful measurements in a production environment. Instead of having a stand-alone measurement capability dedicated to the measurement of a single quantity, the thrust of this approach is to measure as many quantities as quickly as possible with a single set of computer-controlled instrumentation.

The goal is to gather as much information as economically as possible while preserving the integrity of the data. The restrictions mentioned above have the following important implications for the test technology:

(1) *Measurements must be rapid.* The measurement time to evaluate a single parameter must be less than a few seconds. This means that simple procedures are needed for both data acquisition and on-line data reduction. Results from these stripped-down measurements must be carefully correlated with results from more accurate measurements to ensure their validity.

(2) *Test chips must be measurable in wafer form with the use of a multifunction tester in a near-room ambient.* The measurement environment is rather noisy, especially for low-level signals. To aid the measurement process it may be necessary to build in buffer amplifiers on-chip to increase signal levels.

(3) *Test chips must be organized so that all test structures can be accessed by a single probe array.* As will be shown, a scheme has been developed that uses a 2-by-N probe array to access all test structures.

(4) *Test structures must be designed so they are modular,* can be stored in a computer cell library, and can be rapidly assembled into a new test chip. This goal is achieved by designing test structures so they can be probed by a 2-by-N probe pad array.

(5) *Test structures must be designed to allow the characterization of very small regions,* since many of the electrical properties of VLSI devices are dominated by peripheral rather than bulk transport. Because microelectronic test structures are fabricated from the very elements that compose the

integrated circuits, they are uniquely qualified to be used as analytical tools in evaluating VLSI circuits.

B. Objections to the Use of Test Chips

As described in Section I, test chips have been used for a number of years; however, their utility is often questioned. An analysis of the objections to their use reveals their strengths and weaknesses. Some of the objections follow.

(1) *Test chips take up the space of functional circuits* and so are an automatic yield loss. The space taken up by a test chip must be justified by the return received in improved process control and diagnostic capability. These are difficult to justify on an absolute basis.

(2) *Test results from test chips reveal proprietary information* about the wafer fabrication process. Test chips greatly assist the "reverse engineering" of a wafer fabrication process. When test chips are used to transact business, the results often must be kept proprietary.

(3) *Test chips divert attention* away from making circuits perform properly to making the test chips perform. This objection is not relevant in the "silicon foundry" business where fabricated wafer lots are accepted based on test-chip results rather than on circuit performance.

(4) *The measurement of test chips requires an excessive amount of engineering effort* to develop test software. The impact of this objection can be minimized through careful planning aimed at minimizing software costs. Currently no standards exist for test software that will allow the transfer of test programs between different test equipment.

(5) *Advanced test chips cannot be measured with a digital integrated circuit tester.* To take full advantage of the diagnostic capabilities of a test chip, a special tester, a multifunction parametric tester, must be used.

The use of test chips in the semiconductor industry is becoming increasingly important. As Potter has observed [16], test chips are essential in many commercial transactions. Scott *et al.* [17] indicate that test chips offer an effective tool to speed up the LSI fabrication, assembly, and test processes and to enhance the reliability of the final product—all at a substantial cost savings. Currently test chips are supplied by the wafer fabricator, but user-supplied test chips are destined to be used in the procurement process.

III. TYPES OF TEST STRUCTURES

Test structures are used for a variety of purposes in the fabrication of integrated circuits. The six categories of their use are to extract device parame-

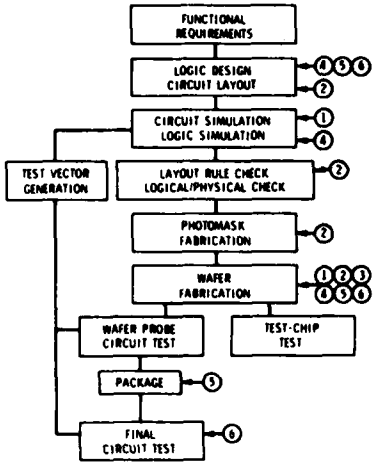


Fig. 5. Simplified integrated circuit production sequence illustrating where results from the six types of test structures are used. Test structures are used for (1) device parameter extraction, (2) layout rule checking, (3) process parameter extraction, (4) random fault analysis, (5) reliability analysis, and (6) circuit parameter extraction.

ters, check layout rules, extract wafer-fabrication parameters, analyze random faults, analyze reliability, and extract circuit parameters. Results from the test structures are used at a number of steps in the production of integrated circuits as illustrated in Fig. 5. Here a typical manufacturing sequence is shown, and it is seen that the most heavy utilization of test structures is found in wafer fabrication. Of course, one must go through the process at least once and collect results from test structures before the results can be used to influence the fabrication process.

In this chapter, test structures are discussed according to their use. The relationship between test structure uses and their description as circuit elements is shown in Table III, where it is seen that a particular circuit element has a variety of uses. For example, a discrete resistor may be used to extract process doping information, such as sheet resistance; it may be used to provide wire resistance data needed in modeling circuit propagation delay. Large-area capacitors may provide a host of process parameters, such as oxide thickness, interface density, or bulk dopant density; they may be used to evaluate random fault densities, such as shorts between conducting layers; they may be used in reliability analysis, such as time-dependent oxide breakdown studies; or they may provide circuit parameters, such as wire capacitance data.

Each of the six types of test structures is described below. At the beginning of each section a number of examples is listed, but only one is described in detail.

TABLE III

Relationship between Test Structures as Circuit Elements and Test Circuits and Their Use in Data Acquisition

	Resistor				Transistor		
Test structure use	Discrete	Array	Capacitor large area	Diode discrete	Discrete	Array	Test circuit
Device parameter extraction	X		X	X	X		
Layout rule checking		X					
Process parameter extraction	X		X	X	X		X
Random fault analysis		X	X			X	X
Reliability analysis		X	X		X		X
Circuit parameter extraction	X		X				X

A. Test Structures for Device Parameter Extraction

These structures are used to extract the parameters for circuit simulation and wafer fabrication control. Examples are

(1) test transistor for threshold voltage, conduction factor, transconductance, breakdown voltage, and leakage current measurements, and

(2) resistors and capacitors for interconnection parameter measurements such as wire resistance and capacitance.

As an example consider the extraction of MOS transistor parameters in both the linear and saturation portion of the transistor current–voltage ($I-V$) characteristics. The method follows from Ham [7] and serves to illustrate the strategy used in measuring transistors in wafer form. The method is illustrated by the $I-V$ characteristics shown in Fig. 6 where seven measurements of the drain–source current I_{DS} establish seven transistor parameters. This is a very small number of measurements to characterize a complex device such as a transistor.

Before device parameters are measured, an untested device is given a qualification test. The strategy calls for an initial stress test to ensure that the transistor has no major faults. In essence one approaches the untested device as a skeptic, requiring that the device prove that it is worthy of further detailed characterization.

The parameter extraction procedure [7] is outlined in Table IV. The initial screening tests consist of measurements that evaluate threshold voltage stability, gate-leakage current, current handling capability, and low-current lin-

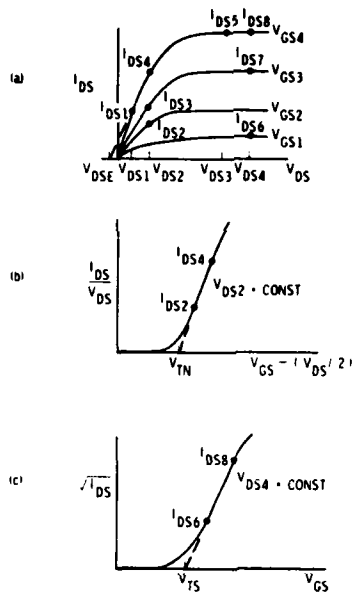


Fig. 6. NMOS transistor characteristics. (a) Common source characteristics, (b) nonsaturation region conductance curve, and (c) saturation region $\sqrt{I_{DS}}$ curve.

earity. The screening parameters must be within previously established parameter bounds or the device is rejected. This saves overall measurement time in that defective devices are not characterized further.

Referring to Table IV, it is seen that the extraction procedure requires the user to define eleven quantities: I_{DS0} is the drain–source current and V_{DS2} is the drain–source voltage; they are used to determine the threshold voltage in the linear region. V_{Tmax} is the maximum allowable threshold voltage. V_{GS0} is the gate–source leakage current, and ΔV_{Tmax} is the maximum allowable shift in the threshold voltage. V_{GS} is an estimate of the upper bound for the gate–source voltage. V_{DS4} is the maximum drain–source voltage used to characterize the device in the saturation region. I_{DSmin} is the minimum allowable drain–source current that is used to evaluate the current handling capability of the device. V_{DSmax} is the maximum allowable extrapolated drain–source voltage V_{DSE} that is used to verify the linearity of the $I-V$ curve in the linear region; see Fig. 6a. The measured quantity in most cases is a current that is measured with an electrometer. The V_{GS} measurement requires the use of an operational amplifier [7].

The parameter extraction process assumes a device model for the transistor in both the nonsaturation (linear) and saturation regions. In the

nonsaturation (linear) region where $(V_{GS} - V_{TN}) > V_{DS}$, the Sah equation [18] is used to describe the drain–source current:

$$I_{DS} = k_N[2(V_{GS} - V_{TN})V_{DS} - V_{DS}^2], \tag{1}$$

where V_{GS} is the gate–source voltage, V_{DS} is the drain–source voltage, V_{TN} is the nonsaturation (linear) threshold voltage, and k_N is the nonsaturation (linear) conduction factor. The analysis for the device parameters follows by rearranging the above equation

$$I_{DS}/V_{DS} = 2k_N[V_{GS} - (V_{DS}/2) - V_{TN}]. \tag{2}$$

A plot of this equation is shown as the linear portion of the curve in Fig. 6b, where the intercept at $I_{DS} = 0$ denotes the threshold voltage, V_{TN}. From the above equation it is seen that the conduction factor follows from

$$k_N = (1/2V_{DS})(\partial I_{DS}/\partial V_{GS})|V_{DS} = \text{const}. \tag{3}$$

The transconductance is defined by

$$g_m = (\partial I_{DS}/\partial V_{GS})|V_{DS} = \text{const}. \tag{4}$$

This relation is used to evaluate g_m in both the nonsaturation (linear) and saturation regions.

In the saturation region where $(V_{GS} - V_{TS}) \leq V_{DS}$, the drain–source current is described by

$$I_{DS} = k_S(V_{GS} - V_{TS})^2, \tag{5}$$

where k_S is the saturation conduction factor and V_{TS} is the saturation threshold voltage. A plot of the square root of this equation is represented by the linear portion of the curve shown in Fig. 6c. The threshold voltage V_{TS} is found by extrapolating the straight-line portion of the curve to $I_{DS} = 0$. The conduction factor follows from

$$k_S = [(\partial\sqrt{I_{DS}}/\partial V_{GS})|V_{DS} = \text{const}]^2. \tag{6}$$

The output conductance is determined from

$$G_{DS} = (\partial I_{DS}/\partial V_{DS})|V_{GS} = \text{const} \tag{7}$$

and provides an evaluation of the degree of flatness of the $I-V$ characteristics in saturation. In the device parameter extraction procedure presented here the body effect was ignored for the sake of simplicity.

This procedure is noteworthy in that numerous qualification tests are made to ensure the integrity of the device parameters that are determined from seven current measurements. The approach wisely requires that k and V_T be evaluated in both the nonsaturation (linear) and saturation regions, for significant differences have been observed between these parameters. The procedure for obtaining the k and V_T values utilizes a two-point method

TABLE IV

MOS Transistor Parameter Extraction Procedure

Test	Stimulus	Measured quantity	Derived parameter	Fail test	User-defined quantity[a]
Initial threshold	V_{DS2}, I_{DS0}	V_{GS01}	$V_{T1} = V_{GS01}$	$V_{T1} > V_{Tmax}$	V_{DS2}, I_{DS0}, V_{Tmax}
Gate leakage (positive stress)	$+V_{GS0}$, t	I_{GS1}		$I_{GS1} > I_{GSmax}$	V_{GS0}, t, I_{GSmax}
Threshold stability	V_{DS2}, I_{DS0}	V_{GS02}	$V_{T2} = V_{GS02}$, $\Delta V_{T1} = V_{T2} - V_{T1}$	$\lvert\Delta V_{T1}\rvert > \Delta V_{Tmax}$	ΔV_{Tmax}
Gate leakage (negative stress)	$-V_{GS0}$, t	I_{GS2}		$\lvert I_{GS2}\rvert > I_{GSmax}$	
Threshold stability	V_{DS2}, I_{DS0}	V_{GS03}	$V_{T3} = V_{GS03}$, $\Delta V_{T2} = V_{T3} - V_{T1}$	$\lvert\Delta V_{T2}\rvert > V_{Tmax}$	
			$V_{GS4} = V_{T1} + V_{GS}$		V_{GS}
			$V_{GS3} = V_{T1} + 0.9V_{GS}$		
			$V_{GS2} = V_{T1} + 0.5V_{GS}$		
			$V_{GS1} = V_{T1} + 0.25V_{GS}$		
			$V_{DS1} = 0.5V_{DS2}$		
			$V_{DS4} = 0.9V_{DS4}$		V_{DS4}
Current handling ability	V_{DS2}, V_{GS4}	I_{DS4}		$I_{DS4} < I_{DSmin}$	I_{DSmin}
Linearity check	V_{DS1}, V_{GS4}	I_{DS1}	$V_{DSE}(V_{GS4}) = \dfrac{(V_{DS1}I_{DS4} - V_{DS2}I_{DS1})}{(I_{DS4} - I_{DS1})}$	$\lvert V_{DSE}\rvert > V_{DSmax}$	V_{DSmax}
Threshold (nonsaturation)	V_{DS2}, V_{GS2}	I_{DS2}	$V_{TN}(V_{DS2}) = \dfrac{[I_{DS4}(V_{GS2} - (V_{DS2}/2)) - I_{DS2}(V_{GS4} - (V_{DS2}/2))]}{(I_{DS4} - I_{DS2})}$		
Conduction factor (nonsaturation)	V_{DS2}, V_{GS3}	I_{DS3}	$k_N(V_{DS2}) = \dfrac{(I_{DS4} - I_{DS2})}{[2V_{DS2}(V_{GS4} - V_{GS2})]}$		
Transconductance (nonsaturation)			$g_{mN}(V_{DS2}) = \dfrac{(I_{DS4} - I_{DS3})}{(V_{GS4} - V_{GS3})}$		
Threshold (saturation)	V_{DS3}, V_{GS4}	I_{DS5}	$V_{TS}(V_{DS4}) = \dfrac{(V_{GS1}\sqrt{I_{DS8}} - V_{GS4}\sqrt{I_{DS6}})}{(\sqrt{I_{DS8}} - \sqrt{I_{DS6}})}$		
Conduction factor (saturation)	V_{DS4}, V_{GS1}	I_{DS6}	$k_S(V_{DS4}) = \left[\dfrac{(\sqrt{I_{DS8}} - \sqrt{I_{DS6}})}{(V_{GS4} - V_{GS1})}\right]^2$		
Transconductance (saturation)	V_{DS4}, V_{GS3}	I_{DS7}	$g_{mS}(V_{DS4}) = \dfrac{(I_{DS8} - I_{DS7})}{(V_{GS4} - V_{GS3})}$		
Output conductance (saturation)	V_{DS4}, V_{GS4}	I_{DS8}	$G_{OS}(V_{GS4}) = \dfrac{(I_{DS8} - I_{DS5})}{(V_{DS4} - V_{DS3})}$		

[a] User-defined quantities are noted on the first test that requires them. The parameter t is stress time.

based on predetermined V_{GS} and V_{DS} values in the nonsaturation (linear) and saturation regions. Other more detailed approaches have been suggested. For example, k and V_T values can be determined at the maximum slope of the curves shown in Figs. 6b and c. Tradeoffs must be made between detailed measurements that take considerable time and simple measurements that can lead to erroneous results.

B. Test Structures for Layout Rule Checking

These structures are used to evaluate those geometrical circuit layout features that form the layout rules. Examples are

1. cross-bridge sheet resistor for linewidth measurements [19] and
2. alignment resistor [20, 21] or a comb resistor [22] to evaluate feature-to-feature spacing.

As an example of the use of this type of test structure, consider establishing the optimum contact window opening from the yield of good contacts in a serpentine array of contacts. The metal mask for such a structure, shown in Fig. 7 [23], consists of eight subarrays where the number of contacts progresses from a few (200) contacts to many (19,200) contacts. A cross-section of the structure is shown in Fig. 8 where the metal strap represents one of the small elements in Fig. 7. The structure is repeated many times (about 100) across a wafer. Each subarray is tested for open-circuit fault condition, and results from each string are used to construct a yield curve as seen in Fig. 9. That is, for contacts with a dimension of 3.6 μm on a side and for the subarray with 4800 contacts, the percentage of good subarrays was 33. The results shown in Fig. 9 were obtained from four different wafers each fabricated with a different contact window opening dimension as indicated in the figure.

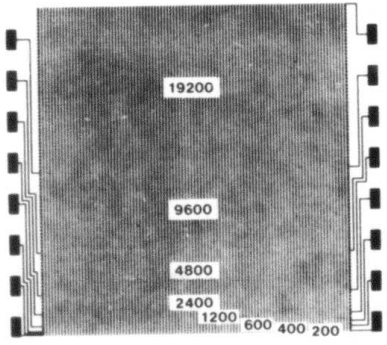


Fig. 7. Metal mask used to fabricate metal-to-silicon contact arrays with various contact window dimensions. The numerical value indicates the number of contacts in each string [23].

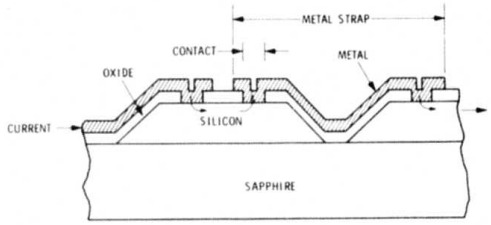


Fig. 8. Cross section of an element from the metal-to-silicon contact array shown in Fig. 7.

From the data shown in Fig. 9 optimum layout rules can be determined by trading off the higher yield of a larger contact window against fewer chips per wafer [23].

The analysis of serpentine structures requires that one assume that *the detected fault is the intended fault*. For the serpentine contact resistor structure, the intended fault is a failure of the metal to make contact with the silicon. Other faults can cause an open circuit and corrupt the data. For example, an open circuit can be caused by a step-coverage fault, a photolithographic or etching fault that omits a silicon island, or a probe fault where a probe fails to make contact with a probe pad. Also, spurious leakage currents between strings in an array can cause unintended faults.

To minimize the occurrence of unintended faults, test structures must be designed with oversize layout rules except for the feature under study, and the structure should be as simple as possible. In addition, the structures should contain double-probe pads to minimize probing errors; see Fig. 3 in [24]. Structures must be inspected visually to verify that the detected fault is the intended fault.

The expression used in the analysis of the data seen in Fig. 9 follows an exponential law. This indicates that faults are uniformly distributed across

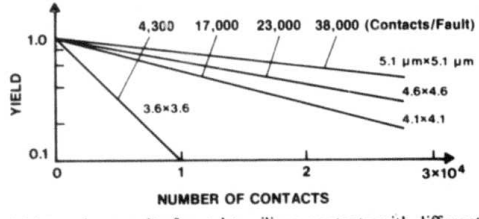


Fig. 9. The yield (number good) of metal-to-silicon contacts with different contact window dimensions [23].

the wafer and occur one at a time in any subarray. But data from other types of arrays have been measured where the yield curves are nonexponential [25]. In such cases multiple faults may be present [22], faults may be clustered [26], or faults may have a radial dependence due to increased misalignment errors toward the periphery of wafers.

Results from this kind of test structure need to be treated cautiously. Their use should be limited to evaluating layout rules and to serving as a guide in wafer processing [7]. Attempts to use the results from different arrays to predict the yield of complex circuits have not met with success. In fact, such attempts detract from the important use of these structures in establishing layout rules.

C. Test Structures for Process Parameter Extraction

These structures are used to evaluate the uniformity of the semiconductor doping processes, the quality of the interfaces between the various semiconductor materials, and the quality of the etching processes used to define the various features in the semiconductor materials. Examples are

(1) cross-bridge sheet resistor for sheet resistance and linewidth measurements [19];
(2) contact resistors for metal-to-silicon or metal-to-polysilicon contact resistance measurements;
(3) MOS capacitors for oxide thickness, interface-state measurements, flat-band voltage, and dopant density measurements [27];
(4) diodes for leakage current measurements;
(5) alignment resistors for evaluating the registration of photomask generated features;
(6) MOSFET dopant profiler for profiling the dopant profiles of various layers [28].

Also included in this type of test structure are nonelectrical structures such as alignment markers, photomask layer designators, surface topology structures, and critical dimension structures.

As an example of the use of this type of test structure, consider the evaluation of the linewidth of the metallization process. The cross-bridge sheet resistor [19] that was used is shown in Fig. 10. The linewidth is determined after measuring the sheet resistance R_s, which is determined from the van der Pauw relation [29]:

$$R_s = (\pi/\ln 2)(\Delta V/I), \tag{8}$$

where I is the current forced between I_1 and I_2, and ΔV is the voltage difference between V_1 and V_2. The linewidth W is determined from

$$W = R_s L I^*/\Delta V^*, \tag{9}$$

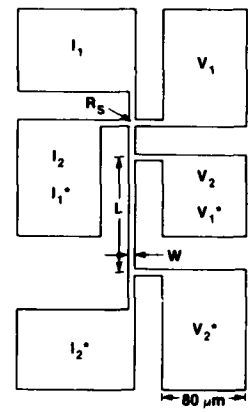


Fig. 10. Cross-bridge sheet resistor used to measure the sheet resistance and linewidth of a metal layer.

where L is the distance between the voltage tapes shown in Fig. 10, I is the current forced between I_1^* and I_2^*, and ΔV^* is the voltage difference measured between V_1^* and V_2^*.

Using the test structure shown in Fig. 10, a single photomask test chip was prepared on a 10× master reticle. The test chip was composed of a 12-by-20 array of identical test structures with a design linewidth of 6 μm. The final photomask was prepared from the master reticle by a step-and-repeat process. The photomask was used in conjunction with a contact printer and a photolithographic process to etch the test-chip pattern into an 800-nm-thick aluminum layer. The aluminum had been electron-gun evaporated and deposited on an oxide film thermally grown on a 2-in. (50.8-mm) diameter silicon wafer.

The linewidth variations shown in Fig. 11 are from a single row of test structures measured across the diameter of the wafer. The plot shown in Fig. 11 indicates that the linewidth varies periodically with the chip dimension of 250 mils (6.35 mm). This periodic or *intrachip* variation is superimposed on a nonperiodic or *interchip* linewidth variation due to those factors that affect the contact between the photomask and the photoresist-covered wafer. The periodic or intrachip linewidth variation is due to aberrations in the optics of the image repeater used to step and repeat the 10× reticle. Similar results have been reported for a 15-μm line [30]. The absolute magnitude of the variations shown in Fig. 11 is independent of the magnitude of the linewidth. Thus, the impact of such linewidth variations on device characteristics is quite dramatic especially for small devices. The linewidth variations for the lines shown in Fig. 11 is about 13%. For 1-μm lines, the variation would be 70%.

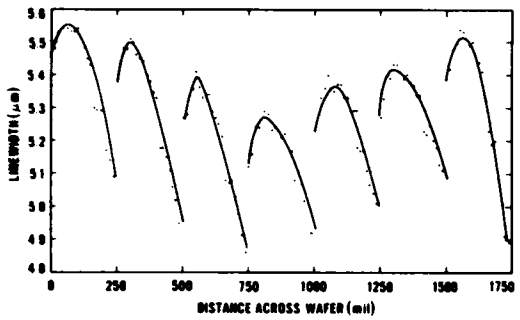


Fig. 11. The variation of the linewidth as determined from an array of identical cross-bridge resistors as shown in Fig. 10.

These results raise some important issues in developing a test-chip methodology. For instance, the circuit simulator SPICE [31] calls for one value of each parameter to be entered into the computer program. But as shown in Fig. 11, parameters can have a distribution of values that varies with position in the test chip as well as with position on the wafer. One usually does not have the luxury of detailed parametric data like that shown in Fig. 11, so guidelines must be established to indicate where test structures should be placed in the test chip and where test chips should be placed on the wafer to facilitate the acquisition of the most representative data. Such guidelines will be important in providing the circuit designer with the best data for the circuit simulators and in establishing the criteria for the positioning of test structures used in the purchase of fabricated wafers.

D. Test Structures for Random Fault Analysis

These structures are used to evaluate the physical faults in the semiconductor material system before it is subjected to significant stressing. A knowledge of the faults is necessary for logic design, logic simulation, and test-vector generation. To detect these faults, test structure arrays are constructed out of series, parallel, or addressable arrays of elements. Examples are

1. serpentine resistor for metal step coverage analysis,
2. comb resistor for measuring the quality of the etching process in separating conducting lines,
3. MOS capacitor for oxide integrity (pinhole) measurements, and
4. addressable MOSFET for identifying the exact nature of a defect and for accumulating fault statistics.

The serpentine and comb-type structures are designed to evaluate the occurrence of certain physical failures. For example, the metal step-coverage resistor is used to evaluate breaks in the metallization at oxide steps. These failures are known as "intended failures" for the test structure. But other unintended failures can masquerade as intended failures. For the example of the metal step-coverage resistor, an open circuit can occur due to the failure of a probe to touch a probe pad or the failure of the photomasking process to properly define the metal lines. In order to identify unintended failures, the test structure must be examined visually. This is a time-consuming process when looking for a failure in a large number of elements. In order to reduce the search time, an addressable MOSFET array was developed to pinpoint the location of physical failures so that the site can be examined visually.

The MOSFET array shown in Fig. 12 is composed of 100 MOSFETs where the gate is connected to the drain. This structure appears on test chip NBS-16 [32], which includes two *p*-channel and two *n*-channel MOSFET arrays. On a 76.2-mm-diameter wafer, 380 arrays containing 38,000 MOSFETs were tested. The results shown in Table V are from 26,760 MOSFETs located in the interior portion of the wafer. Both the fault location and the relative density of different fault types, both clustered and nonclustered, can be determined from the electrical data. A fault is considered to be clustered when two or more adjacent MOSFETs containing the same fault type are detected in an array. As seen in the table, the most frequent fault was number 8, a combination of excessive leakage current and low breakdown voltage.

The advantage of the MOSFET array test structure is that it pinpoints the location of faults so that one can visually examine the site to determine the exact nature of the failure. The disadvantage of this test structure compared with a serpentine or comb structure is twofold. First, the time to test the structure is much longer. Second, the number of elements tested is much smaller. The number of elements tested could be increased by using a memorylike structure with addressable registers. But in this kind of structure one would have to sort out faults in the address registers from faults in

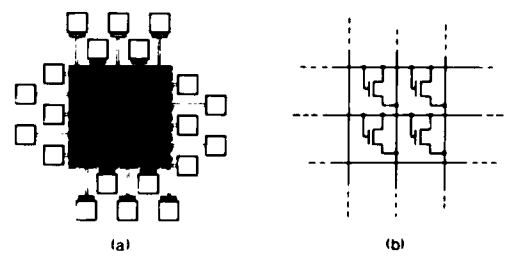


Fig. 12. (a) MOSFET array test structure consisting of 100 individually addressable transistors. (b) A schematic diagram. (Developed by L. W. Linholm.)

TABLE V

MOSFET Array Test Results

Number	Fault			Number of faults	
				Nonclustered	Clustered
1	Polycrystalline void/break			4	0
2	Epitaxial void			0	1
3	Metal void/break			0	0
4	Metal bridge			1	0
5	Gate short			4	0
	V_T[a]	I_L	V_B		
6	—	—	L	5	0
7	—	H	—	25	0
8	—	H	L	62	0
9	L	—	—	1	1
10	L	H	—	0	1
11	L	H	L	0	1
12	H	—	—	2	2
13	H	H	L	0	1

[a] V_T = threshold voltage; I_L = leakage current; V_B = breakdown voltage. H = parameter too high: L = parameter too low.

the array. An ideal structure is one where a large number of elements can be tested quickly and where failure sites can be pinpointed.

E. Test Structures for Reliability Analysis

These structures are used to identify failure mechanisms that result from atomic motion or changes in ionic charge states. The approach is to measure the change in a parameter after the semiconductor material system is given an environmental stress such as overvoltage, temperature, humidity, and radiation. Both the reliability of circuit chips and the package are addressed by these structures.

Examples of the kinds of structures used to evaluate the reliability of silicon devices are listed in Table VI. As seen in the table, the structures vary from single devices to arrays of devices. This table is similar to the one presented by Peck [33] and includes in a similar manner the stress conditions for each failure mechanism.

Test structures for measuring package-induced stress are discussed first. A strain gauge test structure [34] was fabricated by diffusing or implanting a serpentine resistor. With this structure Spencer and his co-workers showed

TABLE VI

Test Structure for Evaluating Some Time-Dependent Failure Mechanisms in Silicon Devices[a,b]

Device association	Failure mechanism	Test structure	Stress condition[b]	Reference
Package	Chip stress	Strain gauge resistor	P, T	[34], [35], [36]
Oxide	Dielectric breakdown	Large-area MOS capacitor	T, J	[37], [38]
Oxide	Charge injection	MOSFET	E, T	[39]
Oxide	Surface charge spreading	MOSFET	T	[40]
Oxide	Ion migration	MOSFET	E, T	[41]
Metallization	Electromigration	Serpentine resistor	T, J	[42]
Metallization	Corrosion	Comb resistor	H, V, T	[43], [44]

[a] See Peck [33].

[b] E = electric field, H = humidity, J = current density, P = pressure, T = temperature, V = voltage.

that package-induced stresses can be quite large, approaching 60% of the breaking strength of silicon. In addition, they indicate that these stresses can produce significant shifts in the k' of MOSFETs. Other workers [35] observed that the package-induced stresses caused significant shifts in resistors used in monolithic D/A converters. Their strain gauge, which consisted of three diffused silicon resistors arranged at 60° with respect to each other, was used to map the strain in chips with different crystal orientations. They discovered the best crystal orientation for the chip and the correct crystal direction along which to fabricate their resistors so as to minimize stress-induced resistance changes. Still other workers [36] have used commercially available Wheatstone-bridge-type strain sensors in their studies.

Test structures for measuring oxide-related failure mechanisms [37–41] are listed in Table VI. These structures are usually large-area devices or special transistor structures. Schroen and his co-workers [40] show a combination of test structures for measuring local stress, ionic accumulation, leakage current, and metal corrosion.

The test structures for measuring metallization-related failure mechanisms [42–44] are the serpentine and comb resistors. The aluminum electromigration studies [42] used a structure that had a serpentine resistor 3-cm long. This is a rather long structure when compared to most electromigration test structures, but, as the authors pointed out, the total length of the wires on present-day VLSI chips is about 100 cm. Structures with long wires are complicated by the increasing probability that a fatal structural defect will occur. This means that the structures must be prescreened for photolithographic defects before electromigration studies can begin or that outliers from the main distribution be discarded. Fortunately in this study outliers due to photolithographic defects were not present. The structures were fabricated with

linewidths and spacings that varied from 1 to 8 μm. Accelerated life testing was done at temperatures ranging from 150 to 250°C and at current densities of 1×10^5 to 2×10^6 A/cm^2. The failure rates were observed to gradually increase as the linewidth decreased, although for the smallest width of 1 μm the failure rates were lowest in certain large-grain aluminum films.

A summary of comb and serpentine test structures is shown in Figure 13. The comb resistor shown in Fig. 13a is used to measure the corrosion currents flowing between the metal fingers [44] and to determine layout rules by measuring the frequency of bridging faults [23]. The comb MOSFET shown in Fig. 13b is used to measure surface leakage currents and appears in the test chip shown in Fig. 2 (structures 31 and 32). By measuring the connectivity of the serpentine resistor shown in Fig. 13c, layout rules or the effect of metal corrosion can be evaluated by analyzing the frequency of opens. The step coverage resistor shown in Fig. 13d appears in the test chip shown in Fig. 1 (structures 1–3). The combined serpentine-comb resistor shown in Fig. 13e was used by Bartlett and Schoenberg [22] to evaluate linewidth and spacing layout rules for printed circuit boards. Finally, the combined serpentine-comb resistor shown in Fig. 13f was used by Vaidya *et al.* [42] to study electromigration and by Skar and Kozakiewicz [43] to study metal corrosion.

The combined serpentine-comb resistors are the recommended structures because both electrical connectivity and isolation can be measured, which allows an efficient use of test structure area. As illustrated above, these structures are used not only for reliability analysis but also for layout rule checking and device parameter extraction (leakage current).

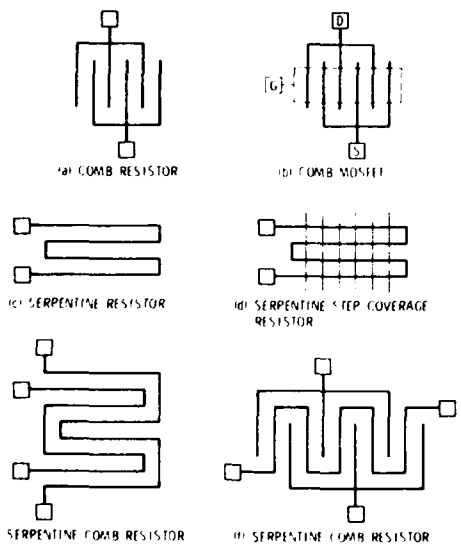


Fig. 13. Schematic diagrams of various comb and serpentine resistors used to evaluate electromigration, metal corrosion, layout rules, and leakage currents.

F. Test Structures for Circuit Parameter Extraction

These structures are used to extract circuit parameters that characterize ac and dc circuit performance and to verify that wafer fabrication process can produce functional circuits. Examples are

1. inverters for measuring the inverter threshold, gain, and noise immunity and
2. ring oscillators for measuring the oscillation frequency and stage delay.

The full characterization of circuit performance involves evaluating such parameters as power dissipation, maximum clock frequency, fan-out capability, and the propagation delay.

The ring oscillator is a popular test structure for characterizing the capabilities of a wafer fabrication process. The oscillation frequency and power are measured, and the stage delay and gate power dissipation are calculated. Relating these results to processing details is difficult. As expressed by Ham [7] the frequency of oscillation is of secondary interest, since it is very difficult to relate the frequency to a specific parameter or processing step. Operation in a reasonable frequency range guarantees that the process is capable of producing working circuits.

A simplified expression for the pair delay of a ring oscillator was derived as an aid in their design and analysis; the derivation is presented in the appendix. The pair delay was derived for the NMOS inverter pair shown in Fig. 14 where the load transistors Q_{L1} and Q_{L2} are depletion-mode devices, the input transistors Q_{I1} and Q_{I2} are enhancement-mode devices, and C ($C = C_1 = C_2$) is the equivalent load capacitance. As depicted in the lower portion of the figure, the analysis assumes the input to stage 1 is a falling step voltage and the output of stage 2 is a rising ramp [45]. The output of stage 2 is shown in detail in Fig. 15 for both a simplified and a more complete analysis. The Mead-Conway NMOS design parameters [46] were used in the analysis. The pair delay is the time for the output of stage 2 to fall to the inverter threshold V_{TINV} for a falling step input to stage 1. The pair delay is

$$t_D = t_{on} + \tau_r \sqrt[3]{(V_{DD} - V_{TINV})/V^*}, \tag{10}$$

where

$$t_{on} = \tau_r(V_{T1} - V_L)/(V_{DD} - V_L), \tag{10a}$$

$$\tau_r = C/k_L|V_{TL}|, \tag{10b}$$

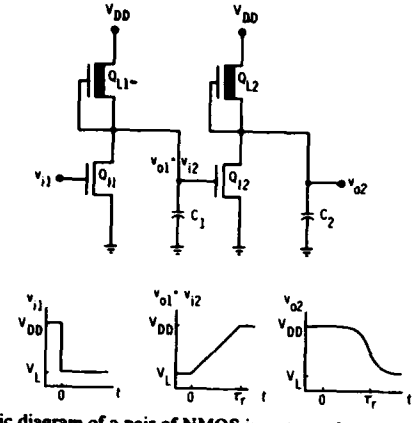


Fig. 14. Schematic diagram of a pair of NMOS inverters where the input transistor Q_I is an enhancement device and the load transistor Q_L is a depletion device.

$$V^* = (k_I/k_L)(V_{DD} - V_L)^2/(3|V_{TL}|). \tag{10c}$$

The important parameters are V_D, the power supply voltage; V_{TI}, the input transistor threshold voltage; V_{TL}, the load transistor voltage; V_L, the inverter low-state voltage; k_I, the input transistor conduction factor; k_L, the load transistor conduction factor; and C, the equivalent load capacitance. The oscillation frequency for the ring oscillator is $1/Nt_D$, where N is the number of inverters in the ring.

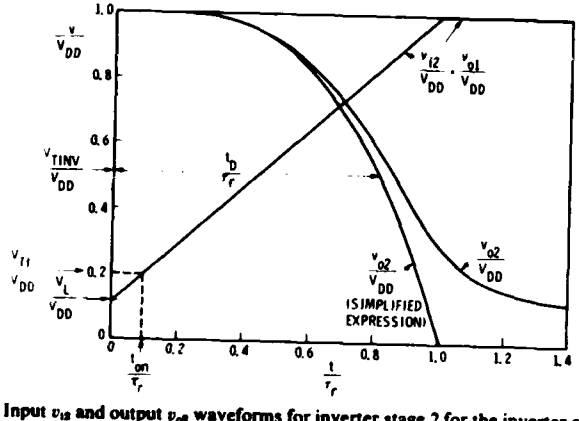


Fig. 15. Input v_{i2} and output v_{o2} waveforms for inverter stage 2 for the inverter chain shown in Fig. 14 where the input to inverter stage 1 is a falling voltage step at $t = 0$. The simplified expression is derived in the appendix and labeled Eq. (A7).

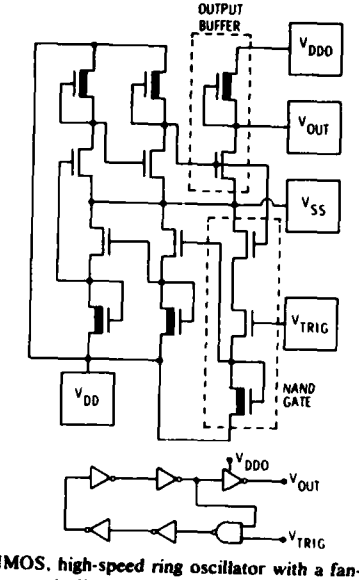


Fig. 16. Five-stage, NMOS, high-speed ring oscillator with a fan-in and fan-out of one, a NAND gate trigger, and output buffer inverter separately powered.

In designing ring oscillators, additional design factors must be considered. The number of stages in the oscillator must be an odd number (usually a prime number) that is large enough to allow each inverter to reach the high or low state after switching state. Ring oscillators with a few numbers of stages oscillate at abnormally high frequencies, because each stage does not have enough time to reach a final value. For the oscillator shown in Fig. 16, the output buffer inverter is a minimum geometry transistor that adds a minimum capacitive load to the oscillator. In addition, the output inverter is separately powered so that the power drawn by the oscillator can be accurately measured [47]. The circuit also includes a NAND gate trigger [48] that allows one to control the onset of oscillation preventing the propagation of multiple pulses.

The overall architecture of the ring oscillator can take many forms. The traditional design is the so-called high-frequency configuration where inverters are arranged as shown in Fig. 16. Here the path length and hence the capacitive load are minimized. Alternative designs involve the use of NOR gates [7] or programmable logic array-type architectures [49]. These configurations are intended to more closely simulate circuit environments by using fan-in, fan-out, and wiring loads that are typical of actual circuits. Yu *et al.* [49] show that the circuit delay is about 100 psec for the high-speed ring oscillators whereas for heavily loaded arrays the delay is about 2 nsec.

IV. TEST-CHIP ORGANIZATION AND TEST-STRUCTURE DESIGN

As mentioned in Section II, the test chip must be organized so that all test structures can be accessed by a single probe array. This restriction has two important economical considerations. First, only one common probe array need be purchased for all test chips, and second, time is saved in not having to change probe cards. Test-structure design is guided by two factors. First, the design must minimize the possibility that faults will degrade the results. Second, the design must allow for modular structures and test procedures that can be easily transferred to new test designs. This feature is particularly important in maintaining a consistent data base between various test chips.

A. Test-Chip Organization

The major issue in organizing a test chip concerns the placement of the probe pads. Over the years several philosophies have developed that are noteworthy.

Early test chips generally were designed with the probe pads located at the periphery of the chip [50]. For the test chip shown in Fig. 17, the location of the peripheral probe pads coincides with the location of the probe pads on the integrated circuit. Such a test chip, when used as a drop-in in a high-volume integrated circuit wafer as seen in Fig. 3, can be probed along with the circuits. The test chip is encoded (a short between two pads, for example) to alert the tester that the unit under test is a test chip and not a circuit. The tester then changes the test sequence from the circuit test to the test-chip test. Another example of the use of the peripheral probe approach can be seen in the work of Reynolds [5].

The peripheral probe pad approach is severely pad limited. The number of probe pads located at the periphery is limited by the dimension of the chip or the number of probes in the probe array. For example, if the test chip shown in Fig. 2 were redesigned with 100-μm square peripheral probe pads, only 88 pads could be accommodated, whereas the chip has 216 probe pads.

The peripheral probe pad approach encourages the use of common bussing where parts of several test structures are connected to the same probe pad. This practice can lead to interferences between structures and places an additional burden on the test engineer to demonstrate for all test conditions that no interference exists. If an interference can occur, then it must be shown that the interference is easily recognized and is not of a subtle nature that will lead to results that will be taken as acceptable. The transistor structure shown in Fig. 18 contains all gates and sources in common. A fault in one gate can cause all the other transistors to be inoperable, thus preventing one to isolate which device is failing. If all devices work perfectly, no difficulty is encountered. It should be noted that the test of the first

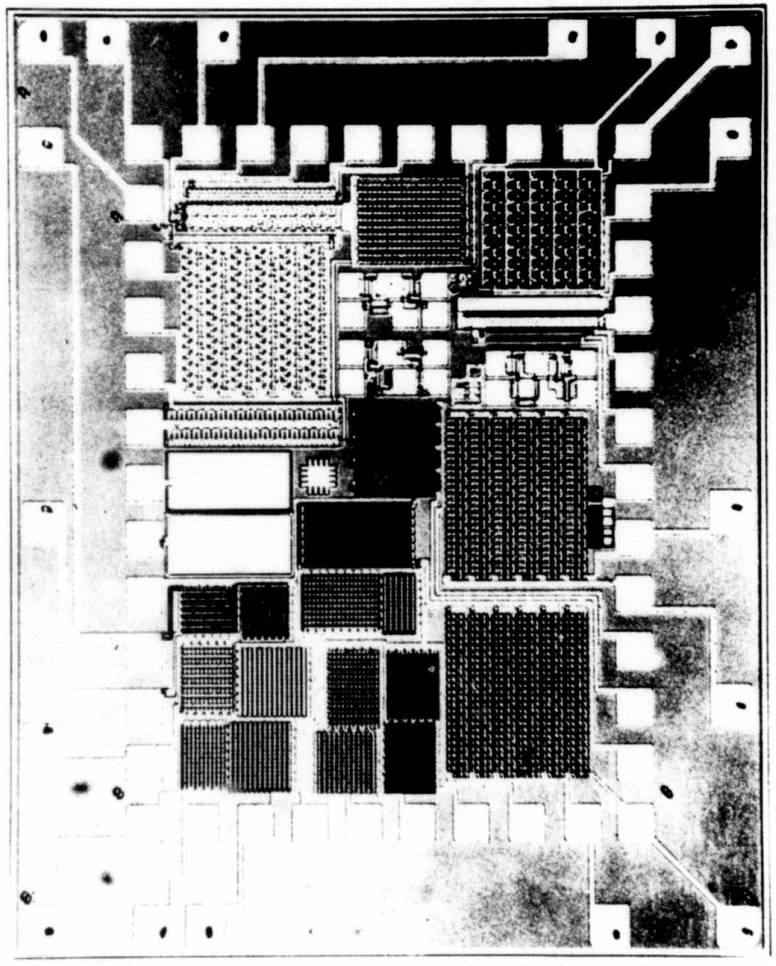


Fig. 17. Test chip (2.36 mm × 2.92 mm) designed with peripheral probe pads and internal probe pads. The outer set of probe pads is located at the same site as the probe pads on the product chips [50].

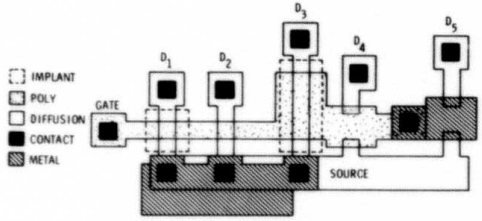


Fig. 18. Five test transistors, where all the gates are connected together as well as all the sources, illustrating the common bussing concept.

Device number	1	2	3	4	5
Transistor	Min. geom. depletion	Min. geom. enhancement	Pull-up depletion	Min. geom. field	Min. geom. enhancement
Gate	Poly	Poly	Poly	Poly	Metal
Oxide	Gate	Gate	Field	Field	Field

transistor prestresses all other transistors, which can be deleterious to the characterization of subsequent transistors, especially if oxides are unstable. In this case an additional 8 pads are required in order to isolate all transistors.

Another objection is that the peripheral probe pad approach does not enable the use of modular test structures that can be called from a computer cell library and placed anywhere on the test chip. With pads on the periphery, each test-chip design may be a unique design experience that does not build on previous design experience. Because of the uniqueness of a new design, it may not be possible to relate results from the new chip with previous results. In this case the data in the old data base may be useless.

Various modular probe arrays are found in the literature. Initial attempts consisted of subdividing test chips into minichips with peripheral probe pads. In the work of Jerdonek *et al.* [51], 20 probe pads were located on three sides of a 1.5-mm × 1.5-mm minichip. Tingley and Johnson [6] located 21 probe pads on three sides of a 0.71-mm × 2.03-mm minichip. The objective of both sets of workers was to provide for wafer probe and package measurements. Zucca *et al.* [52] placed 14 probe pads on all four sides of a 0.22-mm × 0.44-mm minichip. The minichip allows for a degree of modularity, but the level of modularity is above test structure level.

To achieve modularity at the test structure level where structures can be called from a cell library requires the use of the 2-by-N probe pad array [53]. The parameter N is an arbitrary positive integer limited by the number of probes allowed in the measurement system. In many test chips N is 10. The array is illustrated in Fig. 19. The layout requires a square pad 80 μm on a side with 80-μm spacing between pads. In order to conserve space, the

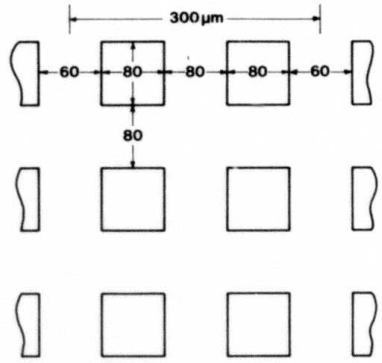


Fig. 19. The 2-by-N probe pad array [53].

pad-to-pad spacing between arrays is 60 μm. The overall width of the array is 300 μm.

The probe pad size of 80 μm was chosen after a study of the probeability of various pad sizes, including 40, 60, 80, and 100 μm [54]. A special test chip was designed, and probing failures were measured for each probe pad size. Excellent results were achieved for the 60-, 80-, and 100-μm pads. To be on the safe side, the pad size was chosen as 80 μm. It should be noted that the test chip of Zucca *et al.* [52] used 60-μm pads spaced 10 μm apart. Their test chip was fabricated in a 14-mm (0.55-in.) GaAs wafer. With a small wafer, proper alignment and run-out across the wafer is less severe than with larger 75- and 100-mm silicon wafers. Thus, a smaller probe pad size is probably acceptable for a smaller wafer.

At NBS, over 30 test chips have been designed using the 2-by-N probe pad array. As an example consider the test chip shown in Fig. 20. This chip contains 640 probe pads and 140 test structures that are electrically isolated from each other [55]. The ability to design variations of test structures for study purposes is greatly enhanced by this pad arrangement. Other investigators [56] have used this pad layout in their test chips. The flexibility of the 2-by-N probe concept was demonstrated recently by Wetterling [57]. He confined all his test structures to a 2-by-2 array. This allowed test chips to be assembled in a variety of configurations that matched the size of the circuit chips. For large chips the structures are arranged in a test strip that accompanies each circuit.

The test strip, in which test structures are accessed by a single 2-by-N pad array, is becoming an important vehicle in the purchase of wafers from silicon foundries. It may prove feasible to place a test strip on each circuit that is produced and to develop chip acceptance criteria based on test of the test strip. At this time it is not clear which structures should be included on the test strip and how they should be tested.

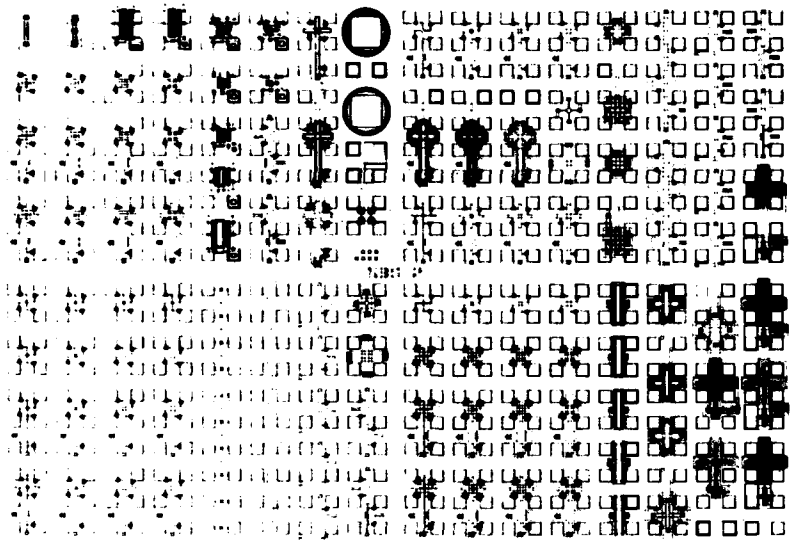


Fig. 20. Test chip NBS-12 (3.2 mm × 4.8 mm) illustrating the use of modular test structures probeable with a 2-by-10 probe array [55].

Future organizational schemes must allow for the bonding and packaging of test chips. The 2-by-N array is unsuitable for this application. A floor plan for such a test chip is shown in Fig. 21 where peripheral probe pads access selected structures that will be stressed. The central portion of the chip is organized using the 2-by-N array where structures are measured in a near-room environment. Since packaging test chips will greatly increase the cost of the measurements, alternatives should be sought. One possible alternative is to stress the test chips while they are still in wafer form. Some of this technology has been developed [58]. However, packaging test chips cannot be avoided entirely, for it will be necessary to evaluate the reliability of the package using test chips [40].

B. Test-Structure Design Rules

From the previous discussion it is clear that an appropriate choice for the probe pad arrangement is the 2-by-N probe pad array. With this as a design constraint, certain other restrictions must be followed in order to properly emulate the functional circuits and to obtain highly accurate measured values. These restrictions follow.

(1) *Use the 2-by-N probe pad array approach.* This allows (a) the test structures to be assembled in a machine-readable cell library, (b) the probe

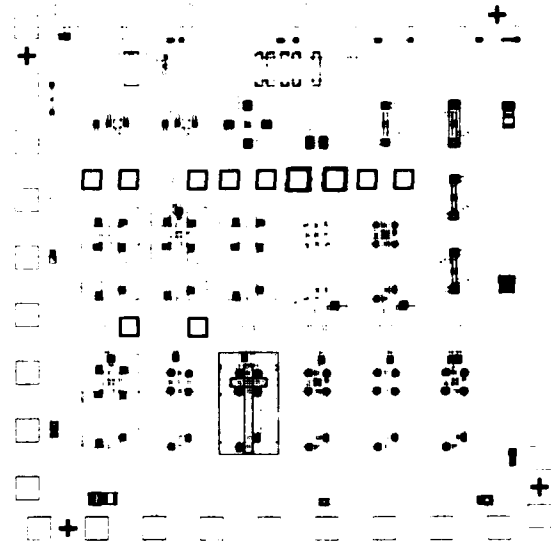


Fig. 21. Test chip organized to accommodate both wafer probe measurements and packaged measurements. The larger pads around the periphery are intended to be used for wire bonding the chip to a package.

pads to be an integral part of the test structure, (c) the electrical isolation of test structures thus minimizing structure-to-structure interference, and (d) the "standardization" of the geometries of the test structures so that many users can use the same design.

(2) *Use the same layout-rule set to design both the test structures and the functional circuits.* This is perhaps the oldest rule in test-chip design [1]. The reason for this rule is that the test structures must emulate the actual circuit geometries as much as possible. There is a notable exception to this rule. When engaging in process evaluation, it is appropriate to shrink layout rules to discover the limits of the process.

(3) *Design test structures so that only critical parts are sensitive to the layout rules.* Many parts of a structure can be designed with oversize features that will lessen the occurrence of a faulty structure. For example, contact size, metal overlap at contacts, and linewidths and spacings can be increased in noncritical areas. Design structures so that they are tolerant to photomask misalignment. Also be aware of pinholes in insulating layers as a potential source of faults; excessively large conducting features can be faulted with pinholes. Only the critical dimensions of the structure need be specified, and these can be tailored to each manufacturing facility. Recently a computer program was developed that allows one to specify certain features of a test structure while holding others constant [59]. This is illustrated

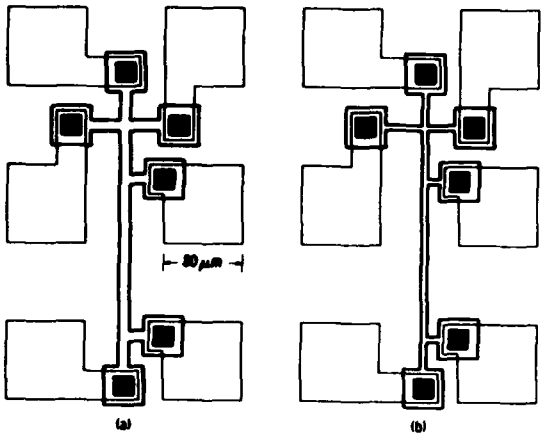


Fig. 22. Test structures drawn by a layout program where critical features are parameterized [59]. The cross-bridge test structures are shown with two linewidths for the conducting channel (outlined with a heavy line). The linewidth is 10 μm in (a) and 5 μm in (b).

in Fig. 22, which shows two cross-bridge sheet resistors with different linewidths but the same probe pad dimensions. With this approach, the critical features of a test structure can be tailored to the layout rules of a particular chip manufacturer. This saves having to redesign the entire test chip.

(4) *Avoid common bussing between test structures.* Structures that are connected together are said to share a common bus. The danger in such connections is that a fault in one structure may influence values measured in the other structure. The objective in good test structure design is to use the "separation of variables" concept to design structures that have a minimum of interferences. The 2-by-*N* array provides a pad-rich environment in which to lay out test structures where interferences can be avoided. Sometimes it is desirable to design test structures with common connections. For example, when the results of one structure are to be correlated or used with another, it is highly desirable to place the structures as close together as possible. A successful combination of structures requires a rigorous fault analysis to assure that a fault in one structure will not affect another. An example of a combined structure is the cross-bridge sheet resistor; see Figs. 10 and 22. This structure requires 6 probe pads. If the structures are separated into the cross and bridge structures, then 8 pads are required.

(5) *Use four-terminal resistors for sheet resistance and linewidth measurements.* The Kelvin contact scheme for measuring resistance uses a four-terminal resistor where current is passed between two terminals, and the voltage is measured between the other two terminals. The cross-bridge sheet resistor is an example of such a test structure. The advantage to this approach is that contact resistance encountered in injecting current into the

structure is eliminated from the measured resistance. Do not use two-terminal or "dog bone" resistors.

(6) *Use channel stops wherever possible.* A channel stop serves to shut off surface currents that result from a surface inversion channel. There are several ways to eliminate these channels. The surface can be doped more heavily than the region that is inverted or a field plate can be used to accumulate the surface. Results taken on sheet resistors where surface inversion was present revealed that sheet resistance values can be five times larger than the true value [60]. In addition, the values were in the wrong direction from an intuitive explanation.

When designing test structures it is useful to keep in mind the numerous interferences that can plague the measurements. Many failure mechanisms are listed in Table VI. Additional interferences and their possible causes are listed in Table VII. These interferences were identified in a study of the measurements of junction sheet resistance [60] and can be detected by measuring the various factors listed in Table VII.

TABLE VII

Interferences Affecting the Measurement of Sheet Resistance[a]

Factor	Interference	Possible cause
Asymmetry factor large	Nonuniform sheet resistance across structure	Nonuniform lateral doping and/or junction depth variations
	Nonuniform structure boundaries	Poor photomasks and/or poor wafer fabrication
	Bulk leakage current	Low junction breakdown voltage
Zero offset factor large	Photovoltaic effect	Too much light
	Thermoelectric effect	Thermal voltages at relays and/or probe contacts
	DVM zero offset	Instrumentation problem
Linearity factor large	Joule heating	Excessive current; poor design (structure too narrow)
	Nonlinear resistivity	Grain boundaries and/or junctions
	Oxide–silicon interface currents	Oxide charge causes channels
Sheet resistance value incorrect	Part of structure missing	Poor fabrication and/or structure near edge of wafer
	High-contact resistance	Poor fabrication
	Metal-to-substrate shorts	Pinholes in oxide
	Surface currents across oxide	Incomplete metal removal
	Oxide–silicon interface currents through inversion layer	Excessive oxide charge and/or interface states
	Poor junction isolation	Low breakdown voltage or too much light

[a] See [60].

V. TEST-CHIP TESTERS AND ADVANCED TEST STRUCTURES

The commercial availability of several multifunction parametric testers [10, 61] has greatly facilitated the use of test chips in the manufacturing environment for such applications as wafer-fabrication control or equipment evaluation. These testers generally provide the capability of sequentially forcing dc currents or applying dc voltages at selected contacts of a test structure. Three quantities are measured on processed wafers: voltage, current, and capacitance. Voltage is measured in the 1-mV to 100-V range, current is measured in the 1-nA to 1-A range, and capacitance is measured in the 10 to 1000-pF range. The time to acquire one measured value can vary from 1 msec to 10 sec depending on the measurement impedances, electrical filtering requirements, and settling time. Measurements are usually made in the dark in a room ambient although a hot–cold chuck can be used for measurements at temperatures other than room temperature.

The configuration of a multifunction parametric tester is shown in Fig. 23. The essential feature of this architecture is the connection of the stimulus/measure instruments to the structures on the wafer through a mechanical switch matrix. (Capacitance measurements have not been shown because they require a special probe card in order to make precision measurements and disrupt the flow of a fully automatic test system.) The entire system is supervised by a minicomputer that controls the system through a digital data bus. The system is generally connected to a large main-frame computer that performs such higher-level tasks as data analysis and storage.

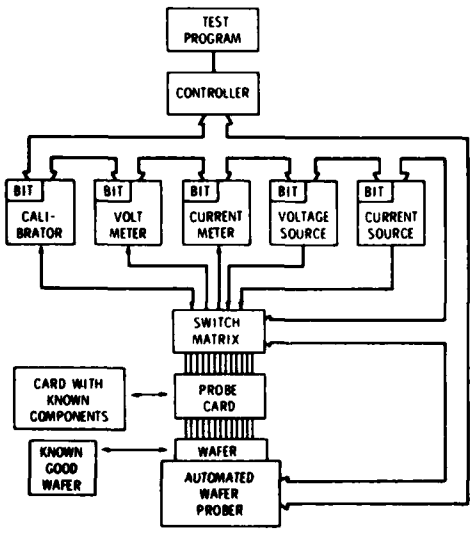


Fig. 23. Multifunction parametric tester used in characterizing test chips. The cards with known components and the known-good wafer are used in system calibration.

The first part of system validation concerns hardware verification to ensure that the system instrumentation is within prescribed precision and accuracy limits. Calibration techniques are similar to those used by the automatic test equipment community and consist of instrument and system level approaches. The instrument level calibration consists of

(a) external calibration where an instrument is removed from the system and calibrated in a calibration laboratory,

(b) self-calibration where the system is reconfigured and self-tests itself using the switch matrix and built-in calibrator, and

(c) built-in test (BIT) where the controller executes a diagnostic program that activates the BIT circuitry in each instrument.

Because the calibration of a system is always a compromise, some shortcomings of the above approaches are mentioned. The external calibration approach has the disadvantage that instruments are not calibrated in the environment in which they are used. The self-calibration approach relies on the accuracy and precision of the calibrator, which can be very good. The BIT approach relies on generally ill-defined self-test routines within each instrument so that the confidence level in system performance is not easily stated.

Once each instrument is calibrated, the system performance must be validated. System level calibration consists of testing the system after various measurement artifacts have been inserted into the front end of the system.

(a) *Known-good-component card.* Such a card is inserted in place of the probe card and allows a check of the system through the switch matrix.

(b) *Known-good wafer.* This wafer allows a system check that best simulates the actual wafer measurement environment.

The known-good wafer can take several forms. For instance, if the wafer is made of silicon it should be extra thick to reduce the chance of breakage, it should have a metal that can be probed many times, and it should be coated with an insulator that prevents ionic contamination of test structures and drift of their characteristics. Because of the difficulty in producing such a wafer, alternative approaches are being sought. At this time it is not clear what technique will be used in the future for the known-good wafer, but the need is real.

The second part of system validation concerns software validation to ensure that the test programs for each test structure are correct. This involves validating the data acquisition sequences and the data reduction routines. The data acquisition sequences must allow for the proper sequencing of the

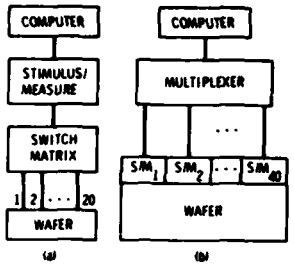


Fig. 24. System architectures of a multifunction parametric tester where (a) there is the existing, commercially available switch matrix approach and (b) there is the proposed pin-electronics approach.

forcing voltages and currents with no glitches between values, and settling times must be set correctly in order to achieve the desired measurement precision and accuracy. This task is simplified when identical systems are used where test programs can be transported between systems in a machine-readable format. The task is very difficult when systems differ in both instrumentation and test languages. A good solution to this probem seems to be the use of well-documented English-language tests that included a flow diagram [7]. Data reduction routines are more easily validated through the use of bench-mark test data and results.

A final note concerns the architecture of multifunction parametric testers. A block diagram of typical commercially available testers is shown in Fig. 24a. The disadvantage of the switch matrix-based testers is due to (a) the mechanical switch matrix, which generates thermal voltage noise and reduces the speed of the measurements (relay settling time is in the 1–10-msec range) and (b) cable noise introduced by the often long cables (10 m in some cases) that connect the stimulus/measure instruments to the probe card. An alternative is shown in Fig. 24b where the stimulus/measure (S/M) instruments are placed as close to the wafer as physically possible. This requires the replication of the S/M instruments by the number of probe pads, 40 times in Fig. 24b. With this architecture, digital signals are switched electronically (multiplexed), whereas in the current approach shown in Fig. 24a analog signals are switched mechanically. The advantages of the tester architecture shown in Fig. 24b are reduced measurement noise and improved measurement speed. For example, if the tester shown in Fig. 24b is used to test four-terminal test structures with a 40-pin probe array, the data acquisition time will be reduced by a factor of 10 since the 10 structures can be measured simultaneously. In addition, the acquisition time will improve, since the system is no longer restricted to the 10-msec matrix switch settling time.

When developing new test structures, one must decide if there is a measurement advantage to be gained by incorporating a portion of the tester into

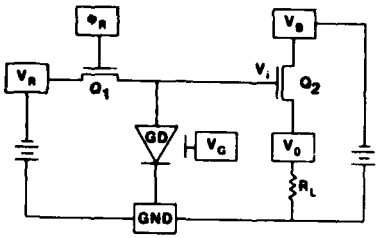


Fig. 25. Schematic diagram of an integrated gated-diode electrometer. The boxes represent probe pads. The off-chip devices are the two dc supplies V_R and V_B and the resistor R_L.

the test structure. Such an advantage has been found with the integrated gated-diode electrometer [62] shown schematically in Fig. 25. Low-level diode leakage currents can be determined by measuring the time decay of the output voltage V_o resulting from the momentary application of reverse-bias voltage V_R to the gated diode GD through the MOSFET switch Q_1. The internal gated-diode current I is determined from the expression:

$$I = (C/\beta)(-dV_o/dt), \tag{11}$$

where C is the diode capacitance and β is the incremental gain of MOSFET Q_2. The diode capacitance can be determined from $C = \epsilon A/W$ where ϵ is the permittivity of silicon, A is the area of the diode, and W is the width of the depletion region. For a one-sided step junction, $W = [2\epsilon(V_i + V_b)/(qN)]^{1/2}$ where V_i is the diode voltage, V_b is the built-in voltage, q is the electronic charge, and N is the dopant density. The dc gain of MOSFET Q_2 is determined from $\beta = \Delta V_o/\Delta V_i$, which is evaluated by closing the MOSFET switch Q_1 ($V_i = V_R$) and measuring V_o at two different values of V_R. The expression above assumes that the capacitance of the gated-diode C is large compared to the gate–source capacitance of MOSFET Q_2. Examples of this test structure are shown in the lower right-hand corner of Fig. 20 in [55, 63].

Another integrated test structure has recently been reported by Iwai and Kohyama [64]. This structure is shown schematically in Fig. 26. Here, the unknown capacitance

$$C_X = C_R[\beta(v_i/v_{oa}) - 1] \tag{12}$$

where C_R is a known reference capacitor, β is the ac gain of the output MOSFET, v_i is the rms value of the ac input signal, and v_{oa} is the output at v_o for v_i connected to point a. To measure C_X, v_i is applied to point a and the MOSFET switch Q_1 is opened by properly biasing ϕ. The ac gain β is determined from $\beta = v_{ob}/v_i$ where v_{ob} is the output at v_o for v_i connected to point b. In this measurement the MOSFET switch Q_1 is closed by properly biasing ϕ. This structure is useful in determining the value of the small capacitances typical of VLSI device geometries by connecting many capacitors into an array.

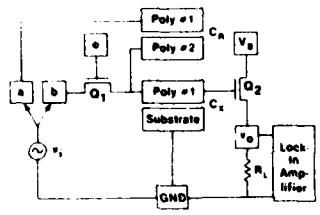


Fig. 26. Schematic diagram of an integrated precision capacitance meter. The off-chip devices are v_1, R_1, and the lock-in amplifier.

VI. FUTURE DIRECTIONS

Test chips have been used for many years, but little effort has gone into their development compared to the massive effort that has gone into the development of other facets of the integrated circuit business. The state of affairs can be judged by the modest amount of literature that exists concerning test chips. Articles often appear in the literature that describe a test chip used in the development of a new device, but very little space is devoted to explaining the test-chip measurements. On a more positive note, the appearance of the multifunction testers has greatly enhanced the usage of test chips. They enable the chip manufacturers to concentrate on the business of producing circuits by removing the burden of having to design and build test equipment.

The major goal for test-chip metrology is to provide diagnostic information in a manufacturing environment. Thus the kind of measurements and the time allowed to acquire the diagnostic data are governed by the return on investment. The goal is met by providing the highest quality data at the lowest possible price on a timely basis. Providing high quality data means that the data are correct and that correct decisions result from the use of the data. The characteristics of test-chip metrology are as follows.

(1) Measurements must be made quickly in order for the results to be of value. This is true in both a manufacturing environment and in the purchase of fabricated wafers. An electrical parameter that takes longer than a few seconds to measure is probably not worth acquiring. For such parameters alternative measurement schemes that allow rapid data acquisition must be sought.

(2) Data must be reduced quickly to a usable form. Data that is analyzed a week or a month from the time it was taken is of little use in a manufacturing environment.

(3) Parameters must be measured at more than a few sites on a wafer in order to properly characterize the parameters. Because parameters vary significantly across wafers and within chips, parameter correlations must be done on a site-by-site basis. The use of "average values" in logic simulations must be viewed with caution.

(4) Test structures must be viewed as "virgin" untested devices. The testing strategy must include a qualification step that an untested device must pass before further tests are undertaken in order to save time and to keep from corrupting the data set.

Improvements that are needed in test chip metrology cover a wide spectrum of activities. For discussion purposes the activities are grouped into data acquisition, data reduction, and vendor transactions.

A. Data Acquisition

Some future needs have already been mentioned, such as the need for improved tester architecture using the pin electronic approach, advanced test circuitry using on-chip test circuitry, and system validation using known-good wafers. Other needs include a portable test language and stress testing for reliability evaluations.

Portable test languages would facilitate the rapid, accurate, and economical transfer of tests between groups. One impediment to the implementation of test-chip metrology is the cost of developing software. A common test language should have a high-level description that can be compiled by machine into the test languages compatible with existing multifunction testers.

Stress testing for reliability evaluation requires the use of special wafer probers equipped with hot/cold wafer chucks and controlled environmental chambers [58]. Thermal stressing in wafer form is more economical than stressing in packages. As mentioned previously, stressing in packages is essential to verify the reliability of the entire chip-package system.

B. Data Reduction

Improvements in data reduction call for developing the fundamental knowledge necessary to more fully understand the devices and the semiconductor material systems. In addition, the application of existing practices in statistical engineering will greatly enhance the data handling problems.

The data reduction algorithms used to convert the measured values to more useful values depend on physical models. For VLSI-type geometries, current transport is better modeled in two dimensions. Simplified expressions are needed that embody the essence of the complex phenomena.

The nature of test-chip metrology is characterized by rapid data acquisition and rapid data reduction using simplified algorithms. In such a situation it is essential that the test-chip measurements be constantly verified against

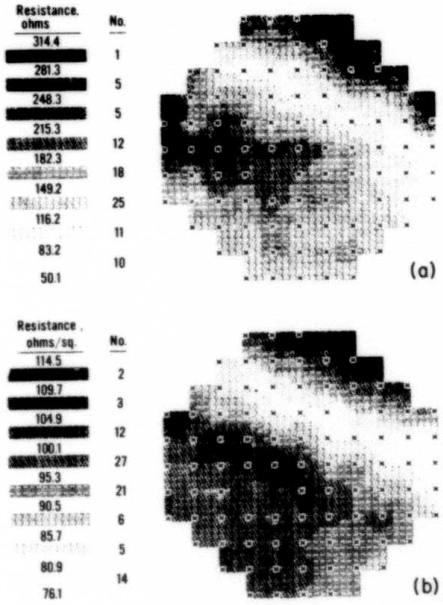


Fig. 27. Wafer maps of (a) the metal-to-n^+ contact resistance and (b) the n^+ sheet resistance. The high correlation between these two parameters leads to the conclusion that excessively high contact resistance was due to the lack of adequate control of a phosphorus implant process step [66].

measurements that allow an in-depth analysis by such analytical measurements as SIMS, AUGER, XPS, SEM, [14] spreading resistance, and ellipsometry.

Information handling is a difficult problem in test-chip metrology because parameters depend on the location on the wafer where they were measured. This causes difficulties in averaging results, in identifying outliers, and in correlating parameters. A number of data reduction and display techniques have been developed and are summarized by Cave and Smith [65]. Recently Linholm [66] presented a technique for averaging results. Also, Ham [7] has described the difficulty of the problem. An example of the nature of the problem is shown in Fig. 27. A contact resistance problem was identified and correlated to variations in the sheet resistance as shown in the figure [66]. The correlation was identified by observing a high correlation of these parameters from a host of other parameters. The maufacturer of these wafers could not have identified this problem because his production wafers contained only two test chips.

C. Vendor Transactions

Test chips are presently being used in the buying and selling of custom-integrated circuit wafers produced by the "silicon foundries." The current usage is limited to very simple parameters usually measured on various transistors. However, a more sophisticated usage can be envisioned where the reliability of the silicon material system is evaluated. Also one could expect that test-chip measurements would take the place of visual inspection performed on the final circuits. In VLSI-type circuits it is no longer economical to perform a detailed inspection for cosmetic defects. As pointed out some years ago [1], visual inspection is limited to the topmost features, for the lower layers are often hidden beneath opaque layers.

APPENDIX

The simplified expression for the pair delay of two NMOS inverters as seen in Fig. 14 was derived with the use of certain assumptions:

(1) The transistors have no storage of transit time effects [4].

(2) The circuit capacitances are combined into an equivalent load capacitance C that is voltage independent [4].

(3) The rise time of the output of inverter stage 1 is independent of the input waveform of stage 1 provided the delay time is computed from the point where the falling input crosses the inverter threshold voltage, V_{TINV} [45].

(4) The load transistor current i_L is approximated by a linear expression derived from the expression for the load transistor operating in the linear region:

$$i_L = k_L|V_{TL}|(V_{DD} - v_o), \tag{A1}$$

where k_L is the load transistor conduction factor, V_{TL} is the load transistor threshold voltage, and v_o is the output voltage.

(5) The two inverter stages are identical so $C_1 = C_2 = C$, $k_{L1} = k_{L2} = k_L$, etc.

Assumption (3) allows the rise time of the output of stage 1 to be calculated for a falling voltage step applied to the input of stage 1. This step input cuts off Q_{I1} allowing Q_{L1} to charge C_1. For these conditios, $i_L = i_{C1}$ where the capacitor current is $i_{C1} = C_1\, dv_{o1}/dt$.

The solution to the resulting differential equation for the initial and final conditions $v_{o1}(0) = V_L$ and $v_{o1}(\infty) = V_{DD}$, is

$$v_{o1} = V_{DD} - (V_{DD} - V_L)\exp(-t/\tau_r), \tag{A2}$$

where the inverter rise-time time constant is

$$\tau_r = C/k_L|V_{TL}|. \tag{A3}$$

The analysis for the simplified pair-delay expression uses the following linear expression derived from the initial slope of Eq. (A2):

$$v_{o1} = ((V_{DD} - V_L)/\tau_r)t + V_L \tag{A4}$$

for $0 \le t \le \tau_r$. For $t > \tau_r$, $v_{o1} = V_{DD}$.

The fall-time analysis for inverter stage 2 assumes that only Q_{I2} redischarges C_2 so $i_{L2} = 0$. *Initially* Q_{I2} is cut off but after v_{I2} exceeds V_{TI}, Q_{I2} operates in the saturation region with a current

$$i_{I2} = k_I(v_{I2} - V_{TI})^2, \tag{A5}$$

where k_I is the input transistor conduction factor and V_{TI} is the input transistor threshold voltage. The output of stage 2 follows from the solution to the differential equation derived from $i_{I2} = -i_{C2}$ where $i_{C2} = C_2\, dv_{o2}/dt$. The equation uses the rising ramp described by Eq. (A4) where $v_{o1} = v_{I2}$ and the initial condition that $v_{o2} = V_{DD}$ for $t \le t_{on}$. The time when Q_{I2} turns on, t_{on}, is calculated from Eq. (A4) for $v_{I2} = v_{o1} = V_{TI}$:

$$t_{on} = \tau_r(V_{TI} - V_L)/(V_{DD} - V_L). \tag{A6}$$

The output voltage for stage 2 for $t \ge t_{on}$ is:

$$v_{o2} = V_{DD} - V^*(t - t_{on})^3/\tau_r^3, \tag{A7}$$

where

$$V^* = (k_I/k_L)(V_{DD} - V_L)^2/(3|V_{TL}|). \tag{A8}$$

The pair delay t_D is defined as the time for v_{o2}, as given in Eq. (A6), to fall to the inverter threshold, V_{TINV}. This time is

$$t_D = t_{on} + \tau_r\sqrt[3]{(V_{DD} - V_{TINV})/V^*}. \tag{A9}$$

An expression for the inverter threshold V_{TINV} is derived from $i_L = i_I$ for $v_o = v_i$. For $v_o = v_i$, i_I is in saturation and described by

$$i_I = k_I(v_i - V_{TI})^2. \tag{A10}$$

The combination of this equation with the expression for i_L as given by Eq. (A1) yields the inverter threshold voltage

$$V_{TINV} = \tfrac{1}{2}[2V_{TI} - (|V_{TL}|/k_R) + \sqrt{(2V_{TI} - (|V_{TL}|/k_R))^2 + 4((|V_{TL}|V_{DD}/k_R) - V_{TI}^2)}] \tag{A11}$$

where $k_R = k_I/k_L$. The inverter output low-state voltage is calculated from $i_L = i_I$ for $v_i = V_{DD}$. For this condition i_I is in the linear range and described by

$$i_I = k_I(2(v_i - V_{TI})v_o - v_o^2). \tag{A12}$$

The combination of this equation with the expression for i_L as given by Eq. (A1) yields the inverter output low-state voltage:

$$V_L = \tfrac{1}{2}[(2(V_{DD} - V_{TI}) + (|V_{TL}|/k_R)) - \sqrt{(2(V_{DD} - V_{TI}) + (|V_{TL}|/k_R))^2 - 4(|V_{TL}|V_{DD}/V_R)}]. \tag{A13}$$

The pair delay is shown in Fig. 15, which illustrates the input and output waveforms for inverter stage 2. In the computation of these waveforms, the Mead and Conway [46] design parameters were used. That is, $V_{TI} = 0.2V_{DD}$, $V_{TL} = -0.8V_{DD}$, and $k_R = k_I/k_L = 4$. For these values, $V_L = 0.119V_{DD}$, $V_{TINV} = 0.513V_{DD}$, and $t_{on} = 0.092\tau_r$. In Fig. 15, Eq. (A7) is plotted using the above design parameters. For the sake of comparison, a more complete solution for the output voltage of stage 2 is also shown in Fig. 15. In computing the waveform, the rising ramp given by Eq. (A4) was used, but the differential equation was derived from $i_{L2} = i_{I2} + i_{C2}$. Transistor Q_{I2} was allowed to operate in all three modes starting in cutoff, then saturation, and terminating in the linear region. A comparison of the curves indicates that the pair delay as given by Eq. (A9) is a good approximation for the case considered here.

ACKNOWLEDGMENTS

This work was performed under the Product Assurance Technology Program jointly sponsored by the National Aeronautics and Space Administration and the Defense Advanced Research Projects Agency. The author wishes to acknowledge the efforts of the staff engineers and scientists at NBS who participated with the author in developing a number of test structures during the eight years the author was with NBS.

REFERENCES

1. F. J. Barone and C. F. Myers, *Electronics* **41**, 84 (July 1968).
2. G. L. Schnable and R. S. Keen, *IEEE Trans. Electron Devices* **ED-16**, 322 (1969).
3. R. J. Sahni, *Reliability Phys. Symp. Proc.* **8**, 226 (1970).
4. W. M. Penney and L. Lau (eds.), "MOS Integrated Circuits," Van Nostrand-Reinhold, New York, 1972.
5. F. H. Reynolds, R. W. Lawson, and P. J. T. Mellor, *Proc. IEEE* **62**, 223 (1974).
6. R. G. Tingley and D. W. Johnson, *Circuits Manufact.* **17**, 30 (April 1977).
7. W. E. Ham, Comprehensive Test Pattern and Approach for Characterizing SOS Technology. National Bureau of Standards Special Publ. 400-56 (January 1980).
8. M. A. Mitchell and L. W. Linholm, Test Pattern NBS-28 and NBS-28A: Random Fault Interconnect Step Coverage and Other Structures. National Bureau of Standards Special Publ. 400-65 (March 1981).
9. J. M. Mikkelson, L. A. Hall, Aron K. Malhotra, S. D. Seccombe, and M. S. Wilson, *IEEE Int. Solid-State Circuits Conf.* 106 (1981).
10. C. Chrones, Semicond. Int. **3**, 113 (October 1980).

11. H. R. Huff, R. J. Kriegler, and Y. Takeishi (eds.), "Semiconductor Silicon 1981." Chapter 12. Electrochemical Society, Princeton, New Jersey, 1981.
12. M. G. Buehler, Microelectronic Test Patterns: An Overview. National Bureau of Standards Special Publ. 400-6 (August 1974).
13. W. D. Jansen and D. G. Fairbairn, *VLSI Design* (formerly *Lambda*) **2**, 16 (First Quarter 1981).
14. R. Kossowsky, *Reliabil. Phys. Symp. Proc.* **16**, 112 (1978).
15. M. G. Buehler and W. M. Bullis (eds.), Special issue on characterization techniques for semiconductor materials, processes, and devices, *IEEE Trans. Electron Devices* **ED-27** (December 1980).
16. G. Potter, *VLSI Design* (formerly *Lambda*) **1**, 12 (Fourth Quarter 1980).
17. W. R. Scott, L. M. Hess, and F. R. Ault, *IEEE Test Conf.* 494 (1980).
18. C. T. Sah, *IEEE Trans. Electron Devices* **ED-11**, 324 (1964).
19. M. G. Buehler and W. R. Thurber, *J. Electrochem. Soc.* **125**, 650 (1978).
20. T. J. Russell and D. A. Maxwell, A Production-Compatible Microelectronic Test Pattern for Evaluating Photomask Misalignment. National Bureau of Standards Special Publ. 400-51 (April 1979).
21. T. F. Hasan, S. U. Katzman, and D. S. Perloff, *IEEE Trans. Electron Devices* **ED-27**, 2304 (1980).
22. C. J. Bartlett and L. N. Schoenberg, *Circuits Manufact.* **21**, 54 (April 1981).
23. A. C. Ipri and J. C. Sarace, *RCA Rev.* **38**, 323 (1977).
24. M. G. Buehler, *J. Electrochem. Soc.* **127**, 2284 (1980).
25. A. C. Ipri, *RCA Rev.* **41**, 537 (1980).
26. O. Paz and T. R. Lawson, *IEEE J. Solid-State Circuit* **SC-12**, 540 (1977).
27. K. H. Zaininger and F. P. Heiman, *Solid State Technology* **13**, 49 (May 1970).
28. M. G. Buehler, *J. Electrochem. Soc.* **127**, 701 (1980).
29. L. J. vander Pauw, *Philips Res. Rep.* **13**, 1 (1958); *Philips Tech. Rev.* **20**, 220 (1958).
30. L. W. Linholm and M. G. Buehler, Electrochem. Soc. Ext. Abstr. Abstract No. 191 (May 1979).
31. L. W. Nagel, SPICE2: A Computer Program to Simulate Semiconductor Circuits, Memorandum No. ERL-M510, Electronics Research Laboratory, Univ. of California, Berkeley, California (May 9, 1975).
32. L. W. Linholm, The Design, Testing, and Analysis of a Comprehensive Test Pattern for Measuring CMOS/SOS Process Performance and Control, National Bureau of Standards Special Publ. 400-66 (August 1981).
33. D. S. Peck, *Reliabil. Phys. Symp. Proc.* **13**, 253 (1975).
34. J. L. Spencer *et al.*, *Reliabil. Phys. Symp. Proc.* **19**, 74 (1981).
35. S. Komatsu, K. Suzuki, N. Iida, T. Aoki, T. Ito, and T. Sawazaki, *Int. Electron Devices Meeting Proc.* 144 (1980).
36. R. J. Usell, Jr., and S. A. Smiley, *Reliabil. Phys. Symp. Proc.* **19**, 65 (1981).
37. S. P. Li, S. Prussin, and J. Maserjian, *Reliabil. Phys. Symp. Proc.* **16**, 132 (1978).
38. D. L. Crook, *Reliabil. Phys. Symp. Proc.* **17**, 1 (1979).
39. S. Rosenberg, D. Crook, and B. Euzent, *Reliabil. Phys. Symp. Proc.* **16**, 19 (1978); A. K. Sinha, W. S. Lindenberger, W. D. Powell, and E. I. Povilonis, *J. Electrochem. Soc.* **127**, 2046 (1980).
40. W. H. Schroen, J. L. Spencer, J. A. Bryan, R. D. Cleveland, T. D. Metzgar, and D. R. Edwards, *Reliabil. Phys. Symp. Proc.* **19**, 81 (1981).
41. W. H. Schroen, *Reliabil. Phys. Symp. Proc.* **16**, 81 (1978).
42. S. Vaidya, D. B. Fraser, and A. K. Sinha, *Reliabil. Phys. Symp. Proc.* **18**, 165 (1980).
43. N. L. Sbar and R. P. Kozakiewicz, *IEEE Trans. Electron Devices* **ED-26**, 56 (1979).
44. R. B. Comizzoli, L. K. White, W. Kern, G. L. Schnable, D. A. Peters, C. E. Tracy, and R. D. Vibronek, *Reliabil. Phys. Symp. Proc.* **18**, 282 (1980).

45. I. Ohkura, K. Okazaki, and Y. Horiba, *Proceedings IEEE Int. Conf. Circuits Comput.* 953 (1980).
46. C. A. Mead and L. A. Conway, "Introduction to VLSI Systems," Chapter 1. Addison-Wesley, Reading, Massachusetts, 1980.
47. J. Shott and T. Walker, Private communication, (1981).
48. F. L. Schuermeyer, H. P. Singh, Russell L. Scherer, and D. L. Mays, *Int. Electron Devices Meeting Proc.* 441 (1980).
49. H.-N. Yu, A. Reisman, C. M. Osburn, and D. L. Critchlow, *IEEE Trans. Electron Devices* **ED-26**, 318 (1979).
50. W. Murakami, *in* ARPA/NBS Workshop III. Test Patterns for Integrated Circuits, (H. A. Schafft Ed.), National Bureau of Standards Special Publ. 400-15 (January 1976).
51. R. T. Jerdonek, H. F. Bare, Jr., and G. J. Fromen, *IEEE Trans. Electron Devices* **ED-25**, 873 (1978).
52. R. Zucca, B. M. Welch, C.-P. Lee, R. C. Eden, and S. I. Long, *IEEE Trans. on Electron Devices* **ED-27**, 2292 (1980).
53. M. G. Buehler, *Solid State Technol.* **22**, 89 (October 1979).
54. R. L. Mattis and M. R. Doggett, *Solid State Technol.* **21**, 76 (November 1978).
55. G. P. Carver, R. L. Mattis, and M. G. Buehler, Microelectronic Test Patterns NBS-12 and NBS-24, NBSIR 81-2234 (May 1981).
56. D. S. Perloff, F. E. Wahl, C. L. Mallory, and S. W. Mylroie, *Solid State Technol.* **24**, 75 (September 1981).
57. S. Wetterling, *in* "Semiconductor Silicon 1981," (H. R. Huff, R. J. Kriegler, and Y. Takeishi, eds.), p. 896. Electrochemical Society, Princeton, New Jersey, 1981.
58. R. Y. Koyama and M. G. Buehler, A Wafer Chuck for Use Between −196 and 350°C. National Bureau of Standards Special Publ. 400-55 (January 1979).
59. C. A. Pina, Private communication, (1981).
60. M. G. Buehler and W. R. Thurber, *J. Electrochem. Soc.* **125**, 645 (1978).
61. J. S. Howard and J. Nahourai, *Solid State Technol.* **21**, 48 (July 1978).
62. G. P. Carver and M. G. Buehler, *IEEE Trans. Electron Devices* **ED-27**, 2245 (1980).
63. G. P. Carver, L. W. Linholm, and T. J. Russell, *Solid State Technol.* **23**, 85 (September 1980).
64. H. Iwai and S. Kohyama, *Int. Electron Devices Meeting Proc.* 235 (1980).
65. T. Cave and D. Smith, *Comput. Design* **17**, 161 (May 1978).
66. L. W. Linholm, R. L. Mattis, and R. C. Frisch, *in* "Semiconductor Silicon 1981" (H. R. Huff, R. J. Kriegler, and Y. Takeishi, eds.), p. 906. Electrochemical Society, Princeton, New Jersey, 1981.

NASA—JPL—Coml., L.A., Calif.