

A11106 038823

REFERENCE

NIST PUBLICATIONS

NBSIR 82-2492

# A Computer Program for Analysis of Data from Microelectronic Test Structures

---

FILE COPY  
DO NOT TAKE

U.S. DEPARTMENT OF COMMERCE  
National Bureau of Standards  
National Engineering Laboratory  
Center for Electronics and Electrical Engineering  
Semiconductor Devices and Circuits Division  
Washington, DC 20234

June 1982



U.S. DEPARTMENT OF COMMERCE  
NATIONAL BUREAU OF STANDARDS

QC  
100  
U56  
82-2492  
1982



NBSIR 82-2492

**A COMPUTER PROGRAM FOR ANALYSIS  
OF DATA FROM MICROELECTRONIC TEST  
STRUCTURES**

---

R. L. Mattis, L. J. Till, and R. C. Frisch

U.S. DEPARTMENT OF COMMERCE  
National Bureau of Standards  
National Engineering Laboratory  
Center for Electronics and Electrical Engineering  
Semiconductor Devices and Circuits Division  
Washington, DC 20234

June 1982

**U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, *Secretary***  
**NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director***



## Table of Contents

	Page
Introduction . . . . .	1
Data Array Definition . . . . .	2
Definition of Statistical Parameters . . . . .	4
Program Operation Overview . . . . .	5
Data Base Structure . . . . .	6
(1) Data Base File Format . . . . .	7
(2) Data Base Creation . . . . .	7
(3) Writing Entries in a Data Base . . . . .	10
(4) Examining a Data Base . . . . .	10
(5) Defining a Reference Data Sample . . . . .	10
(6) Searching a Data Base . . . . .	10
(7) Deleting an Entry from a Data Base . . . . .	12
(8) Warning against Editing Data Base Files . . . . .	12
Command Syntax . . . . .	12
Command Descriptions . . . . .	13
(1) Miscellaneous: COM, END, PAU, and REM . . . . .	14
(2) Data Input/Output: REA, WRT, WFM, and ASG . . . . .	14
(3) Statistics: ENN, PRS, HIS, and LIP . . . . .	16
(4) Test Site Exclusion: XOL, XPP, XNS, XGT, XLT, XIP, LXP, IIP, and RES . . . . .	17
(5) Data Value Replacement: AXP and AIP . . . . .	18
(6) Functional Fits: FPL, SPL, FQD, and SQD . . . . .	18
(7) Mapping: PLT and MAP . . . . .	19
(8) Data Base Operations: WDB, LDB, GET, SDB, and DEL . . . . .	22
Error Messages . . . . .	23
(1) Errors Associated with Command Syntax . . . . .	24
(2) Errors Associated with Precedence Violations . . . . .	25
(3) Errors Associated with Site Specification . . . . .	25
(4) Errors Associated with Noninteger Input Values . . . . .	25
(5) Errors Associated with Mapping . . . . .	26
(6) Errors Associated with Data Base Commands . . . . .	26
(7) Errors Associated with REA and ASG . . . . .	27
(8) Other Errors . . . . .	28
Addition of New Input Data Formats . . . . .	28
Addition of New Commands . . . . .	29
Program Installation . . . . .	29
Logical Unit Assignments . . . . .	29

	Page
Program Portability . . . . .	30
Acknowledgments . . . . .	31
References . . . . .	32
Appendix I: Command Summary . . . . .	I-1
Appendix II: Listing of ST2MAIN . . . . .	II-1
Appendix III: Listing of ST20V1 . . . . .	III-1
Appendix IV: Listing of ST20V2 . . . . .	IV-1
Appendix V: Listing of ST20V3 . . . . .	V-1
Appendix VI: Listing of ST20V4 . . . . .	VI-1
Appendix VII: Listing of ST20V5 . . . . .	VII-1
Appendix VIII: Listing of ST20V6 . . . . .	VIII-1
Appendix IX: Listing of ST20V7 . . . . .	IX-1
Appendix X: Listing of FETCH . . . . .	X-1
Appendix XI: Listing of CRDB . . . . .	XI-1
Appendix XII: Listing of DBDEL . . . . .	XII-1
Appendix XIII: Characterizing and Analyzing Critical Integrated Circuit Process Parameters . . . . .	XIII-1

#### List of Figures

	Page
1. Example of test site locations . . . . .	3
2. Sample label file . . . . .	8
3. Sample data file . . . . .	9
4. Sample wafer map . . . . .	20
5. Key which accompanies map of figure 4 . . . . .	21

## Disclaimer

Certain commercial equipment, instruments, or materials are identified in this report in order to adequately specify the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.





# A Computer Program for Analysis of Data from Microelectronic Test Structures

R. L. Mattis, L. J. Till, and R. C. Frisch  
Semiconductor Devices and Circuits Division  
National Bureau of Standards  
Washington, DC 20234

## Abstract

A computer program, STAT2, is described which performs the following functions: reads data as a two-dimensional array; calculates mean, sample standard deviation, and median; identifies outliers; calculates replacement values for outliers; makes a gray-tone data map on a line printer; makes a character map on the user's terminal; constructs a data base for examining correlations among various data sets; and searches the data base for correlations using several selective keys. The emphasis in this document is on program usage, and detailed descriptions of the commands are given. Program portability and data input requirements are addressed. Guidance regarding several types of program modifications is provided.

Key words: computer program; correlation coefficient; outlier; process validation wafer; statistical analysis; two-dimensional arrays; wafer map.

## Introduction

This Internal Report is intended to serve as a Program Manual for a computer program titled STAT2 [1]. As such, it contains information on program usage, installation, and internal structure. Program usage is treated in the sections titled Program Operation Overview, Command Syntax, and Command Descriptions. Program installation is treated in sections titled Program Installation, Logical Unit Assignments, and Program Portability. Program internal structure is treated in sections titled Data Array Definition, Data Base Structure, Addition of New Input Data Formats, and Addition of New Commands.

STAT2 is used to analyze test data in which each data value is associated with a test site in a two-dimensional coordinate space. The test data are stored in a two-dimensional array where the subscripts associated with each data value represent the row-column location of the test site. There are some restrictions on the data array as explained in the section titled Data Array Definition.

The program can be used to analyze data from a variety of sources. The intended application, however, is the analysis of measurements from microelectronic test structures for characterizing an integrated circuit fabrication process [2-3]. A paper which discusses this application [2] has been included as Appendix XIII. Test structures are microelectronic devices which are fabricated by the same process used to fabricate integrated circuits. They can be used to measure selected material or process parameters by means of electrical tests. Test structures are typically fabricated on a circular

silicon wafer in a pattern of test sites which is periodic in x and y over the wafer. On the wafer, there may be one or more row-column locations or test sites at which the pattern is interrupted and a different set of test structures or circuits has been inserted. Such a site is called an untested site.

In an integrated circuit process, test structures can be used to identify which parameters accurately predict or determine the degree of process control; to establish the value and range of these parameters for a given process lot; and to determine how these parameters vary across an integrated circuit die, across a wafer, from wafer to wafer, and from lot to lot. Test results must be obtained and interpreted in a timely fashion in order to be used for correcting or improving the process.

This Internal Report describes a computer program which reads data as a two-dimensional array; calculates mean, sample standard deviation, and median; identifies outliers; calculates replacement values for outliers; makes a gray-tone data map on a line printer; makes a character map on the user's terminal; constructs a data base for examining correlations among various data sets; and searches the data base for correlations using several selective keys. These techniques can provide the user with a relatively fast analysis capability for characterizing an integrated circuit process through the determination of the magnitudes of baseline parameters and their variation over the wafer for "properly" fabricated devices. It is assumed that the process being characterized is in sufficient control to produce a high percentage of "properly" fabricated test structures and that defective structures which are encountered are mainly the result of gross defects introduced by handling, by lithography voids, or by similar process irregularities.

An important aspect of the analysis of test data is the identification of test results from defective structures or defective measurements which do not accurately represent the parameter being measured. Such an incorrect data value is called an outlier. It is necessary to exclude outliers from the population of data values in order to make a more accurate statistical estimate of the parameter. A test site whose data value has been determined to be an outlier is called an excluded site. Other test sites are called included sites.

STAT2 is written primarily in FORTRAN VI to run on an Interdata 7/32 (now Perkin-Elmer) minicomputer under revision 4.3 of the OS32MT operating system. One subroutine is written in assembly language, but it is not essential to program operation. The program consists of a main segment and seven overlays and requires about 76 kilobytes of memory. The part of the program that produces the gray-tone map is written for a Printronix P300 line printer/plotter. The listing of the main segment (ST2MAIN) is given in Appendix II. The listings of the overlays (ST20V1, ST20V2, ..., ST20V7) are given in Appendices III through IX.

#### Data Array Definition

A typical data array to be analyzed by STAT2 is shown in figure 1. The data values are contained in an array known as DATA which is dimensioned 32 by 32. The first subscript represents the row number of the test site and the second

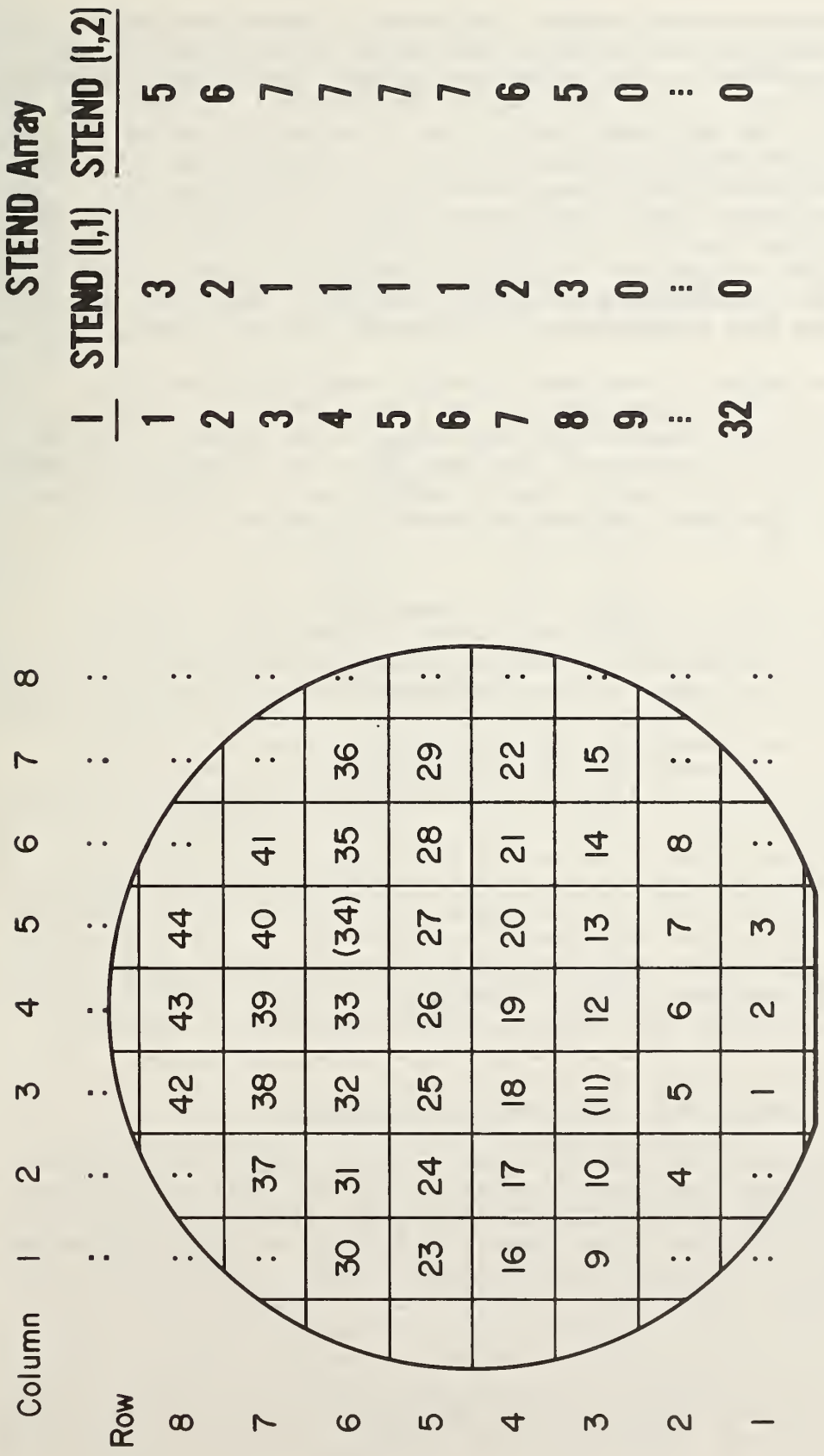


Figure 1. Example of test site locations. (The figure represents a silicon wafer with the major flat at the bottom. The square grid on the wafer represents a periodic pattern of test sites. Row and column numbers are indicated to the left of and above the wafer. Numbers within the wafer are site numbers. Parentheses around site numbers 11 and 34 indicate that these are untested sites. The STEND array for this DATA array is shown in the right part of the figure.)



subscript represents the column number. In figure 1, actual test sites exist at the locations indicated by a number. The numbers are site numbers and represent the serial order in which sites are tested. This order of testing, left to right and bottom to top, is mandatory for formats 1 through 4 but not for formats 5 and 6. No data were taken at the other points, represented by colons; these are considered nonexistent sites. In this example, the sites numbered 11 and 34 are untested sites, and these site numbers in figure 1 are surrounded by parentheses. Even though these sites are not tested, they must still be assigned a site number. The requirements for such a DATA array to be processed by STAT2 are (1) row 1 must not be empty, (2) column 1 must not be empty, (3) no row or column may have fewer than three test sites, and (4) no row or column may have a nonexistent site between two test sites.

The location of test sites is available to STAT2 through the STEND (for SStart-END) array. STEND is dimensioned 32 by 2 where STEND(I,1) is the column number of the first (leftmost) test site in row I and STEND(I,2) is the column number of the last (rightmost) test site in row I. Elements of STEND representing rows where there are no test sites must be 0. The STEND array which describes the test site locations in figure 1 is shown in the right portion of the figure.

For four of the data formats which STAT2 is capable of reading, the subscripts of the DATA array as read do not directly correspond to the row-column locations of the test sites. Rather, one of the subscripts is the site number referred to above. In these instances (formats 1 through 4 below), the STEND array and the site number are used to convert the data into a two-dimensional array based on row-column location, such that after the data have been read into STAT2, all further communication between STAT2 and the user is in terms of a two-dimensional array based on row-column location.

An important attribute of the points in the DATA array is the topological type. Each test site is classified according to whether it is an interior site, site on left boundary, site on an upper right corner, etc. Topological types are used by the outlier replacement algorithms and the mapping subroutines. The topological types are discussed in greater detail in the introductory comments to subroutine ITYPE beginning in line 686 of ST20V3 in Appendix V. Several commands give the type of data points along with other information.

#### Definition of Statistical Parameters

Several statistical terms used in this Internal Report need to be defined. MEAN represents the simple arithmetic average of all data values associated with all included sites. SIGMA represents the sample standard deviation of this same set of data values. K is a multiple of SIGMA calculated by the XOL command wherein data values which are more than K\*SIGMA from MEAN are declared to be outliers. The value of K satisfies the equation

$$\int_{-\infty}^K \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \frac{1 + (1 - p)^{1/n}}{2}, \quad (1)$$

where  $n$  is the number of included sites and  $p$  is the probability that at least one "good" test site may be excluded along with the outliers [4,5] under the assumption that the data values follow a normal distribution.

The sample correlation coefficient,  $r$ , is a measure of the similarity of the spatial variation of two sets of data. When the paired observations  $(x_1, y_1)$ ,  $(x_2, y_2)$ , ...,  $(x_n, y_n)$  are taken on two quantities, if a large value of  $x$  implies a large value of  $y$ , then the quantities are said to be positively correlated. If a large value of  $x$  implies a small value of  $y$ , then the quantities are said to be negatively correlated. If a large value of  $x$  implies nothing about  $y$ , then  $x$  and  $y$  are said to be uncorrelated. The measure of correlation is the correlation coefficient,  $\rho$ , which is estimated by the statistic  $r$ :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[ \sum_{i=1}^n (y_i - \bar{y})^2 \right]}}, \quad (2)$$

where  $\bar{x}$  and  $\bar{y}$  are the sample means of  $x$  and  $y$ , respectively, over the  $n$  points [6]. Note that  $r$  must take on values in the range  $[-1,1]$ .

#### Program Operation Overview

This section describes the capabilities of STAT2 without regard to the details of command syntax. Command names are given parenthetically so that the user can relate this description to the discussion of individual commands in a later section. The example relates to data from a microelectronic test structure.

STAT2 is first loaded with all necessary logical unit assignments made, then started. The user inputs an array of data (REA), also called a data set, for examination. If there are untested sites, the user can exclude them at the outset (XIP). The user can calculate the statistics relating to all test sites (PRS) and draw a character histogram (HIS) showing data value distribution. Data values corresponding to a short- or open-circuited device can be removed from the population by excluding sites having data values less than some lower bound (XLT) or greater than some upper bound (XGT). If the test sites on the periphery differ from the interior sites, it may be desirable to exclude them (XPP). If at some time the user wants to put a particular excluded site back in the population, he may do so (IIP). He may also want to put all sites back into the population (RES) and try a different exclusion procedure. After known outliers (such as shorts or opens) and untested sites have been excluded, the user may search the remaining data values for outliers (XOL). In some instances a user may want to specify a particular multiple of the standard deviation (ENN), see what test sites lie farther than that amount from the mean (HIS), and exclude those sites (XNS). At any time the user may list the sites which are excluded (LXP) or he may list the characteristics of any individual test site (LIP).

In some applications, the user may be interested in the functional form of the data variation over the wafer surface. The user may fit the data from the included test sites to a plane (FPL) or to a quadratic function (FQD). The user may also subtract the plane (SPL) or the quadratic function (SQD) from the DATA array and examine the residuals.

A useful output of STAT2 is a gray-tone map. The map gives a graphic representation of the variation of the measured parameter over the wafer surface. In order for the map to appear continuous, replacement values must be calculated by interpolation or extrapolation for those test sites which have been excluded (AXP and AIP). A character display can be printed on the command terminal (PLT) prior to making the map so the user can experiment with different scaling parameters. The gray-tone map (MAP) can be made in a variety of sizes and scales, and a label can be placed on it if desired.

The user may save the DATA array in its present form, that is, with replacement values for outliers or with residuals of a functional fit, on disc or tape (WRT). A filemark can be placed between such tape files (WFM).

If the present data set is one of many on which correlation studies are to be performed, the user may write a sample of the data set in a data base (WDB). He may then assign a new input file to be read (ASG) and repeat the process until all data sets have been processed.

After a data base has been established, the user may list all or a portion of its contents (LDB). He may select a particular entry or sample as a reference sample (GET) and calculate a correlation coefficient of the reference sample and a selected group of other samples (SDB). If any entries in the data base are in error or are no longer needed, they may be marked deleted (DEL).

At any time, the user may insert a comment (COM) which appears in the printed output. If the user repeatedly executes the same command sequence, he may create a file of these commands and run STAT2 in the remote mode (REM). The user may pause STAT2 (PAU), such as to change logical unit assignments, and may terminate STAT2 (END) when done.

#### Data Base Structure

A group of commands in STAT2 provide for constructing and using data bases. A data base is a pair of files which contains a sample of data from several or many data sets. The user can search for correlations among data sets by performing calculations of a sample correlation coefficient [6] for one or more samples against other samples. The greater the number of data values included in the sample, the smaller the uncertainty in the calculated sample correlation coefficient. As the sample size increases, however, disc storage requirements and computation time go up. It has been found empirically that for a data set containing 95 points, a sample of 13 points produces sample correlation coefficients of acceptable uncertainty. The nature and use of a data base are presented in this description.



## (1) Data Base File Format

A data base may be constructed for any collection of data sets which may be expected to be related and for which the samples of wafer data stored in the data base are all selected in the same way. Examples of collections of data sets which may be expected to be related are (1) all runs of a particular test pattern, (2) all data from a given run of wafers, and (3) all data from a particular set of test structures. A data base consists of a label file containing 80-byte records and a data file containing 24-byte records. The two files are linked together by bidirectional pointers. A sample label file and data file are shown in figures 2 and 3. A data base entry consists of one record in the label file and (NDAT+1) records in the data file where NDAT is the number of data values in the sample.

The label record contains data in eleven fields. Each data base entry is identified by a sequence number in the SEQ field. The sequence number is assigned when the entry is written (by a WDB command) to be 1 greater than the sequence number of the (previously) last entry in the data base. The sequence number is the identifier by which the entry is known, and it is used by the GET, DEL, and LDB commands. The DLT field is normally an ASCII space, but an asterisk is placed there when an entry is marked deleted (by a DEL command). The PTR field contains the record number of the header record in the associated data file. The NDAT field contains the number of data values in the data file per entry. The next five fields -- PAT, LOT, WAF, DEV and PCODE -- give the pattern number, lot number, wafer number, device number and parameter code relating to the data. The remaining two fields give the date and time at which the entry was written.

The data themselves are in the data file preceded by a header record. The header contains 'H' in the first character position. It also contains the sequence number and a pointer back to the label record number. The data records, their number given by NDAT in the label record, follow the header. Each data record contains a 'D' in the first character position, the sequence number, the data value, and a one-byte EXC field. If that particular data value came from an excluded test site, an asterisk is written in the EXC field; otherwise that field contains an ASCII space.

The first label record and first two data records in a data base are title records which contain a user-assigned title and the sampling plan code (defined below).

The actual format of the data base as depicted in figures 2 and 3 need not be known by the user. STAT2 contains routines to display the information which is needed. The user does need to decide how many data points are to be included in the sample and from what x-y locations on the wafer they are to be taken. This is discussed in the following paragraphs.

## (2) Data Base Creation

To create a data base, the user must decide on a sampling plan, make the necessary software changes if it is a new sampling plan, and write the first record of the label file and the first two records of the data file. Four sampling plans are presently available, with test site locations as indicated





```

T NBS-12 CROSS-BRIDGE
T DATA                                2
H   1   1
D   1 0. 419399E-01
D   1 0. 414922E-01* ← Indicates an
D   1 0. 411353E-01   Excluded Site
D   1 0. 405800E-01
D   1 0. 410900E-01
D   1 0. 409767E-01
D   1 0. 407613E-01
D   1 0. 398434E-01
D   1 0. 401721E-01
D   1 0. 400248E-01
D   1 0. 400021E-01
H   2   2 ← Header Record
D   2 0. 278208E+01
D   2 0. 253562E+01
D   2 0. 280397E+01
D   2 0. 277452E+01
D   2 0. 286600E+01
D   2 0. 260859E+01
D   2 0. 278904E+01
D   2 0. 274292E+01
D   2 0. 267681E+01
D   2 0. 279821E+01
D   2 0. 275982E+01
H   3   3
D   3 0. 410106E-01
D   3 0. 402287E-01
D   3 0. 398548E-01
D   3 0. 401721E-01
D   3 0. 400134E-01
D   3 0. 396848E-01
D   3 0. 393902E-01
D   3 0. 392088E-01
D   3 0. 389935E-01
D   3 0. 390275E-01
D   3 0. 387102E-01

```

Figure 3. Sample data file. (The first two records contain the data base title and sampling plan code. The header records contain the sequence number and the label record number.)

in table 1. Other sampling plans to include up to 256 sites can be added, and the required software changes are not difficult. Refer to comments in subroutine DBØ2 beginning in line 1539 of ST20V7 in Appendix IX.

Allocate a label file having record length of 80 bytes and a data file having record length of 24 bytes. Compile, establish, and run program CRDB, listed in Appendix XI, with logical units 5 and 6 assigned to a user terminal, logical unit 7 assigned to the label file, and logical unit 8 assigned to the data file. Enter a data base description or title, up to 36 characters, as requested, and enter the sampling plan code.

Note in figures 2 and 3 that the title appears in the first label record along with the sampling plan code. The same information appears in the first two data records, but because of the smaller record length of the data file, the title is divided between the two data records. After this program is executed, entries may be written in the data base.

### (3) Writing Entries in a Data Base

To write an entry in the data base (1) read a data set using REA, (2) perform data point exclusions, (3) calculate replacement values for excluded points using AXP, and (4) type the WDB command as directed in the section titled Command Descriptions. After the data base entry has been written, it is read and displayed in readable form on the user's terminal and logged to the line printer.

### (4) Examining a Data Base

The data base can be examined by using the LDB command as explained in the section titled Command Descriptions. The output contains all the information in the label record plus the data values themselves if requested.

### (5) Defining a Reference Data Sample

When searching for correlations among data base entries, it is necessary to define an entry, identified by its sequence number, with which other entries are to be compared. This is done by the GET command as explained in the section titled Command Descriptions.

### (6) Searching a Data Base

A search of a data base for entries which correlate with the reference data sample is initiated by the SDB command. In general, the user does not want to seek correlations between the reference data sample and all entries in the data base.

If any of the data values in either of the two data samples came from an excluded site (as indicated in the EXC field of the data record), those values are not used in sample correlation coefficient calculations.

The sample correlation coefficients thus calculated are based on the data samples and are intended to indicate possible correlations (depending on the magnitude of the sample correlation coefficient). The user should examine

Table 1. Site Locations for the Four Available Sampling Plans.

Sample Number	Site Coordinates for Sampling Plan			
	Code 0	Code 1	Code 2	Code 3
1	1,6	2,4	3,5	2,6
2	2,4	3,2	5,3	3,4
3	2,8	3,6	5,5	3,8
4	4,4	4,4	5,7	5,2
5	4,8	5,1	7,3	5,4
6	5,2	5,3	7,5	5,8
7	5,6	5,5	7,7	5,10
8	5,10	5,7	9,3	7,2
9	6,4	6,4	9,5	7,4
10	6,8	7,2	9,7	7,8
11	8,4	7,6	11,5	7,10
12	8,8	8,4	-	9,4
13	9,6	-	-	9,8
14	-	-	-	10,6
15	-	-	-	-

gray-tone maps of the correlating data sets to evaluate the correlation more fully.

#### (7) Deleting an Entry from a Data Base

An entry is deleted from a data base in a two-step process. A user may mark an entry deleted by using the DEL command. Such an entry is still physically in the data base and is recognized by LDB commands but is ignored by SDB commands and produces an error when referenced by GET and DEL commands.

Actual removal of the deleted entries is accomplished by a stand-alone program DBDEL, listed in Appendix XII. To use DBDEL, load the program and assign logical unit 7 to the old label file, logical unit 8 to the old data file, logical unit 9 to the new label file (record length 80 bytes), and logical unit 10 to the new data file (record length 24 bytes). Messages are printed on logical units 4 and 6. All undeleted records are copied from the old to the new files. Sequence numbers are preserved. If any error occurs in reading a label, header, or data record, a message is printed and DBDEL aborts. At the conclusion of a run, the number of deleted entries which were removed is logged.

#### (8) Warning against Editing Data Base Files

When an error is made in writing an entry in the data base using WDB, the user may try to correct such an error with a text editor. The user is advised that the insertion or deletion of any records from either the label file or data file causes all entries following that insertion or deletion to be inaccessible (the pointers are no longer correct). The safest way to make a correction to the data base is to mark the bad entry deleted using a DEL command, then create a new entry using a WDB command.

### Command Syntax

A command consists of a three-letter mnemonic followed in some cases by a parameter list. Parameters when present are from 1 to 8 in number. A comma must be used as a delimiter between mnemonic and parameter and between parameters. Alphabetic characters (except E), semicolons, terminal commas, and successive commas are illegal and result in a request for repeated command entry. The entire command must not occupy more than 72 characters; excess characters are ignored. Considerable freedom is available in the format of the parameters themselves. They must be no more than 20 characters in length and may be expressed in I, F or E format as may be convenient. A minus sign is legal as the first character in a parameter or as the first character after an 'E'. An optional plus sign is also legal in these two locations. An exponent must be one or two digits plus sign if present.

A space signals the end of a command, and processing of a command line ends as soon as a space is encountered. Character positions to the right of a command by at least one space may be used to place a comment if desired. An exception to this is the ASG command which is parsed in a different manner. No comment may be placed on a line with an ASG command.



There are two commands (LDB and SDB) for which certain parameters can be represented legally by a minus sign only. These parameters are called privileged parameters and are explained in the discussion of LDB and SDB.

### Command Descriptions

The STAT2 command set consists presently of 34 commands. These commands are now discussed individually. The symbols P1, P2, ... represent parameters which would be entered as numerical values as explained above. A summary of the commands is given below and is repeated in Appendix I.

AIP - Alter an individual point.  
ASG - Assign input data file.  
AXP - Alter excluded points.  
COM - Comment.  
DEL - Delete data base entry.  
END - Terminate STAT2 execution.  
ENN - Set N to a specified value.  
FPL - Fit DATA array to a plane.  
FQD - Fit DATA array to a quadratic function.  
GET - Define data base entry as reference sample.  
HIS - Draw character histogram.  
IIP - Include an individual point.  
LDB - List data base entries.  
LIP - List an individual point.  
LXP - List excluded points.  
MAP - Draw gray-tone map of DATA array.  
PAU - Pause STAT2 execution.  
PLT - Draw character display of DATA array.  
PRS - Print statistics.  
REA - Read input data file.  
REM - Set or reset remote mode.  
RES - Restore all points to included status.  
SDB - Search data base for correlations.  
SPL - Subtract fitted plane from DATA array.  
SQD - Subtract fitted quadratic function from DATA array.  
WDB - Write data base entry.  
WFM - Write filemark to logical unit 2.  
WRT - Write STEND and DATA arrays to logical unit 2.  
XGT - Exclude points greater than a value.  
XIP - Exclude an individual point.  
XLT - Exclude points less than a value.  
XNS - Exclude points beyond  $N \cdot \text{SIGMA}$  from mean.  
XOL - Exclude outliers.  
XPP - Exclude peripheral points.

In the command descriptions, frequent references are made to error conditions. An error condition occurs when a command as entered cannot be executed. The error condition may arise due to a command syntax error, an illegal value of a parameter, a failure to execute another command which must be issued prior to the present command, or some other cause. When remote mode is enabled (see the REM command description), such as when STAT2 commands are being read from a command file, an error condition causes STAT2 to pause.

This is done because subsequent commands would, in many cases, be rendered meaningless or misleading until the error condition is corrected. A message indicating an error condition is preceded by three asterisks. On the other hand, warning messages are intended to inform the user of possibly unintentional consequences of a command without interrupting the execution of the command. An explanation of all error messages is given in the section titled Error Messages.

(1) Miscellaneous: COM, END, PAU, and REM

COM - Comment. COM may be followed by any text. The command line is echoed to logical units 4 and 6, but no other action is taken.

END - End. Terminate STAT2 execution.

PAU - Pause. Pause STAT2, such as to change logical unit assignments or perform a system function.

REM,P1 - Remote Enable/Disable. Enable Remote Mode when P1 is nonzero. Disable Remote mode when P1 is zero. Remote Mode is initially disabled when the program is started. When Remote Mode is enabled, any error condition causes STAT2 to pause. It is recommended that Remote Mode be enabled whenever STAT2 is directed from a command file. REM,1 should be the first line in such a file.

(2) Data Input/Output: REA, WRT, WFM, and ASG

REA,P1,P2 - Read Data. Read data from a disc or tape file. The data formats which can be read are those used by NBS for internal data storage. Five of them are unformatted, applicable to disc only, and generally do not allow portability of data files between computers. The sixth format does allow for a portable data format and allows reading from disc or tape. Addition of the capability to read other formats is not difficult and is treated in the introductory comments to subroutine REA, beginning in line 53 of ST20V3 in Appendix V.

When an REA command is issued, two arrays are read into the computer memory, the STEND array and the DATA array. The organization of these two arrays is discussed in the section titled Data Array Definition. The six formats are defined below in terms of the FORTRAN WRITE statement and surrounding code which produced the data file being read. The array dimensions are given also, and these must be exact for formats 1 through 5. For format 6, the arrays could actually be dimensioned larger, but only a 16-by-16 subarray would be read.

```
Format #1:  
INTEGER*2 STEND(50,2)  
DIMENSION DATA(6,100)  
WRITE (LU) STEND, DATA
```

```
Format #2:  
INTEGER*2 STEND(20,2)
```

```
DIMENSION DATA(6,200)
WRITE (LU) STEND, DATA
```

```
Format #3:
INTEGER*2 STEND(50,2)
DIMENSION DATA(240)
WRITE (LU) STEND, DATA
```

```
Format #4:
INTEGER*2 STEND(20,2)
DIMENSION DATA(200)
WRITE (LU) STEND, DATA
```

```
Format #5:
INTEGER*2 STEND(32,2)
DIMENSION DATA(32,32)
WRITE (LU) STEND, DATA
```

```
Format #6:
INTEGER*2 STEND(16,2)
DIMENSION DATA(16,16)
```

```
C
C WRITE THE STEND ARRAY
C
WRITE (LU,10) (STEND(I,1), I=1,16)
WRITE (LU,10) (STEND(I,2), I=1,16)
10 FORMAT (1X,16(I2,1X))
C
C WRITE THE DATA ARRAY.
C
DO 110 I=1,16
DO 120 J=1,16,4
WRITE (LU,130) DATA(I,J), DATA(I,J+1), DATA(I,J+2), DATA(I,J+3)
130 FORMAT (1X,4(E13.6,2X))
120 CONTINUE
110 CONTINUE
```

As discussed in the section titled Data Array Definition, the second subscript of DATA in formats 1 and 2 and the subscript of DATA in formats 3 and 4 contain the site numbers. The row-column coordinates of each site are determined by the REA command from the site numbers and the STEND array (see exception below for format 3). Parameter P1 of the REA command is the format number, an integer in the range 1 through 6 inclusive, indicating which of the above six formats are appropriate for reading the DATA and STEND arrays. Other values of P1 produce an error. Using the wrong P1 value produces unpredictable results. P2 is the subarray of the DATA array, used with formats 1 and 2. That is, a data file read using formats 1 or 2 actually contains as many as six data sets, all having the same STEND array. P2 indicates which of these six data sets is to be examined. It is the value of the first subscript of the array and has a value from 1 to 6. Other values produce an error condition. In the case of formats 3 through 6, P2 must be present but can be any value since it is not used. In the case of format 3, the STEND array must be present but is ignored and is replaced with a STEND array for a



rectangular area 12 rows by 20 columns. REA must precede most other commands except the data base commands.

WRT - Write Data. Writes the STEND and DATA arrays to logical unit 2 in format 6 (see description of REA command). This can be a disc or tape file or a printing device. If any alterations have been made to the DATA array using AXP, AIP, SPL, or SQD, these affected data points are written as altered. If REA has not been executed prior to issuing WRT, an error condition is produced.

WFM - Write Filemark. Writes a file mark to logical unit 2. This is intended to be used only when logical unit 2 is assigned to a magnetic tape drive.

ASG,file-descriptor - Assign Input File. This command closes the input file and assigns a new input file to logical unit 1. In place of the parameters normally used, ASG is followed by a comma, then the complete file descriptor of the input file. An error condition is produced if logical unit 1 was not already assigned, if the file descriptor is not syntactically correct or if it refers to an off-line volume or nonexistent file. The ASG command is written specifically for an Interdata computer with usage restricted to Multi-Terminal Monitor (MTM) terminals. The command is quite useful, however, and other users may wish to implement their own version of it.

(3) Statistics: ENN, PRS, HIS, and LIP

ENN,P1 - Enter N. Enters the value of N, the user-specified multiple of the standard deviation, for use by XNS and HIS commands. P1 is the value of N. A negative P1 produces an error condition. If ENN is not executed, N has the default value of 1.0.

PRS - Print Statistics. Prints the following statistics of the included test sites (that is, the test sites currently included in the population): maximum, minimum, MEAN, median, SIGMA, percent standard deviation, N times SIGMA, and number of included and excluded sites. An error condition occurs if no data file has yet been read. If there are fewer than two included sites, the calculations done by PRS are not meaningful and an error condition occurs. If the median value occurs repeatedly in the DATA array, the algorithm which calculates the median (subroutine MEDCAL) cannot converge to the median; in this case the median is set to 0, and a warning message is printed indicating a lower and upper bound of the median. For real data, these two bounds are usually very close together.

HIS - Print Character Histogram. Divides the range of values of the included sites into 50 equal intervals or bins and places each site in the appropriate bin. It then prints a string of 50 characters (bounded by exclamation points), each character representing the number of sites in the corresponding bin. A minus sign signifies zero sites and an 'X' signifies ten or more sites in a bin. Beneath this character histogram, double quotes are placed at the bin containing the mean value and at the bins containing the N\*SIGMA limits. These limits often would fall outside of the display. When that happens they appear at the end position of the display; however, the bin numbers at which they would appear are printed beneath the display as IBMIN



and IBMAX. Bin numbers outside the range 1 to 50 indicate that one or both of the N\*SIGMA points is off the scale of the display. Following this, a list of all test sites outside the N\*SIGMA limits is printed. The column labeled TYPE refers to the topological position of the data point as explained in the comments to Subroutine ITYPE beginning in line 686 of file ST20V3 in Appendix X. The column labeled STATUS indicates whether a site is presently included or excluded. HIS cannot be executed unless PRS has been executed with no intervening site exclusion or reinsertions or change in N value.

LIP,P1,P2 - List Individual Point. Gives the data value, topological type, and status of the test site at row P1 and column P2. If this location is not an actual measurement site, an error condition occurs.

(4) Test Site Exclusion: XOL, XPP, XNS, XGT, XLT, XIP, LXP, IIP, and RES

XOL,P1 - Exclude Outliers. Uses an iterative method to identify and exclude outliers [2]. All sites having data values beyond a multiple K of SIGMA are excluded, where the multiple K is a function of the number of included sites and of P1, the probability that one or more good sites might be excluded along with the outliers. P1 must be in the range of 0.05 to 0.90 inclusive or an error condition occurs. A value of 0.20 for P1 has been found to be suitable with data from microelectronic test structures [2]. When XOL is invoked, STAT2 makes one or more passes through the data set, prints the statistics current for that pass, the K value, and the number of sites excluded on that pass. The process stops when on a given pass no sites are excluded. Note that XOL alters the value of N set by the ENN command and sets it to the final value of K.

This command and the commands which follow relating to exclusions of test sites require that REA be executed previously. If any of these commands would attempt to exclude a site which is already excluded, a warning message is logged, but this is not considered an error condition.

XPP - Exclude Peripheral Points. Excludes all sites which are located on the boundary of the test site space.

XNS - Exclude Beyond N\*SIGMA. Excludes all sites which have data values outside the range  $MEAN \pm N*SIGMA$  where N is defined by the ENN command. This command is legal only when SIGMA is current; that is, there have been no exclusions or insertions since the last time SIGMA was calculated by PRS. If SIGMA is not current, an error condition occurs.

XGT,P1 - Exclude if Greater Than. Excludes all sites having data values which are greater than the value given by P1.

XLT,P1 - Exclude if Less Than. Excludes all sites having data values which are less than the value given by P1.

XIP,P1,P2 - Exclude an Individual Point. Excludes the site at row P1 and column P2 in the data array. If there is no site at that location, an error condition occurs.

LXP - List Excluded Points. Lists all sites which have been excluded, giving their row-column location, value, and topological type.

IIP,P1,P2 - Reinsert an Individual Point. Reinserts or includes a previously excluded site located at row P1 and column P2. If there is no site at that location, an error condition occurs. If the site to be reinserted has not been excluded, a warning message is printed.

RES - Reset. Restores all test sites to included status. An REA command must be executed prior to RES.

(5) Data Value Replacement: AXP and AIP

AXP - Alter Excluded Points. Performs interpolation or extrapolation of data values from nearby sites to calculate a replacement value for each excluded site for plotting purposes. The replacement is calculated by 1 of 15 algorithms depending on the topological type and the proximity of other excluded sites. For a site having a given topological type, a "first choice" algorithm is selected to calculate the replacement value. If this algorithm would use data values from other excluded sites in the calculation, it cannot be used. A "second choice" algorithm is then selected. For sites having some topological types, a third or fourth choice algorithm is also attempted, if necessary. When the replacement is made, a message is printed giving the site coordinates, topological type, old and new values, and the algorithm used. When a replacement cannot be made, a message so stating is also printed. Note that AXP alters the DATA array. The array can be restored only by reissuing the REA command. Additional information on the mechanism of data value replacement is given in the program listing in the introductory comments to subroutines ALT and RALGA beginning in lines 99 and 394, respectively, of ST20V4 in Appendix VI.

AIP,P1,P2,P3 - Alter Individual Point. Assigns the site at row P1, column P2 the data value P3. The site must already be excluded; otherwise, an error condition occurs. This command is intended to be used judiciously when none of the replacement algorithms is able to calculate a replacement value.

(6) Functional Fits: FPL, SPL, FQD, and SQD

FPL - Fit to a Plane. Finds the equation of the plane which gives the least squares fit to the included test sites. The coefficients of the equation of the plane and the residual standard deviation are printed. REA must be executed prior to FPL.

SPL - Subtract Plane. Takes the plane, the equation of which was calculated by FPL, and subtracts it from the DATA array leaving the residuals in the DATA array. FPL must have been executed prior to SPL with no intervening exclusions or reinsertions of sites; otherwise, an error condition occurs. If a MAP or PLT of the residuals is requested with the autoscale option, PRS must be executed after SPL and prior to MAP or PLT in order that the maximum and minimum values might pertain to the residuals and not to the original DATA array. The SPL operation cannot be undone, so the original DATA array can be recovered only by reissuing the REA command.



FQD - Fit to a Quadratic Surface. Finds the equation of the quadratic surface which gives the least squares fit to the included test sites. The coefficients of the equation of the quadratic surface and the residual standard deviation are printed. REA must be executed prior to FQD.

SQD - Subtract Quadratic Surface. Takes the quadratic surface, the equation of which was calculated FQD, and subtracts it from the DATA array leaving the residuals in the DATA array. FQD must have been executed prior to SQD with no intervening exclusions or reinsertions of test sites; otherwise, an error condition occurs. If a MAP or PLT of the residuals is requested with the autoscale option, PRS must be executed after SQD and prior to MAP or PLT in order that the maximum and minimum values might pertain to the residuals and not to the original data array. The SQD operation cannot be undone, so the original DATA array can be recovered only by reissuing the REA statement.

(7) Mapping: PLT and MAP

PLT,P1,P2,P3 - Plot on Terminal. Produces a character display on logical unit 6 in which a character represents the data value at each test site. The range of values between P2 (the lower bound) and P3 (the upper bound) is divided into eight equal intervals. Data values falling in these intervals are represented by the characters 1 for the lowest interval through 8 for the highest interval. Test sites having values greater than P3 are represented by '+', and sites having values less than P2 are represented by '-'. Excluded sites are represented by ':'. Following this character display, a table is printed showing the upper and lower bounds and the number of sites which fall into each of the eight bins.

The above discussion of P2 and P3 as the lower and upper bounds of the display range applies only if P1 is 0. If P1 is positive, the range is auto-scaled to the maximum and minimum values of the included sites. If P1 is negative, the range is autoscaled to the  $MEAN \pm K * SIGMA$  limits where K is the multiple of SIGMA calculated by the XOL command on its last pass. In either case P2 and P3 must still be present, but their values are ignored. No '+' or '-' characters should appear in the display if either autoscale is used. Either PRS ( $P1 \geq 0$ ) or XOL ( $P1 < 0$ ) must be executed prior to PLT with no intervening site exclusions or reinsertions or change in N value.

In a multiuser environment, where output to the line printer is spooled, plots created by MAP are not visible until STAT2 has terminated. PLT provides a preview of what the map will look like and enables the user to adjust his scaling to produce the desired shade pattern.

MAP,P1,P2,P3,P4,P5,P6 - Draw Gray-Tone Map. Plots the DATA array as a gray-tone map with an eight-level gray scale. P1 is the width and P2 is the height of the map in inches. The width must not be greater than 13 inches nor the height greater than 8 inches. The minimum width and height in inches are given by 0.2 times the number of columns and 0.2 times the number of rows, respectively, of data in the map. This is to ensure that there are a sufficient number of picture elements (pixels) to make a meaningful plot. P4 is used to provide a label for the map. If P4 is nonzero, the user is asked to enter a label up to 72 characters long. The label is printed above the map. If P4 is zero, no label is requested or printed, but the space above

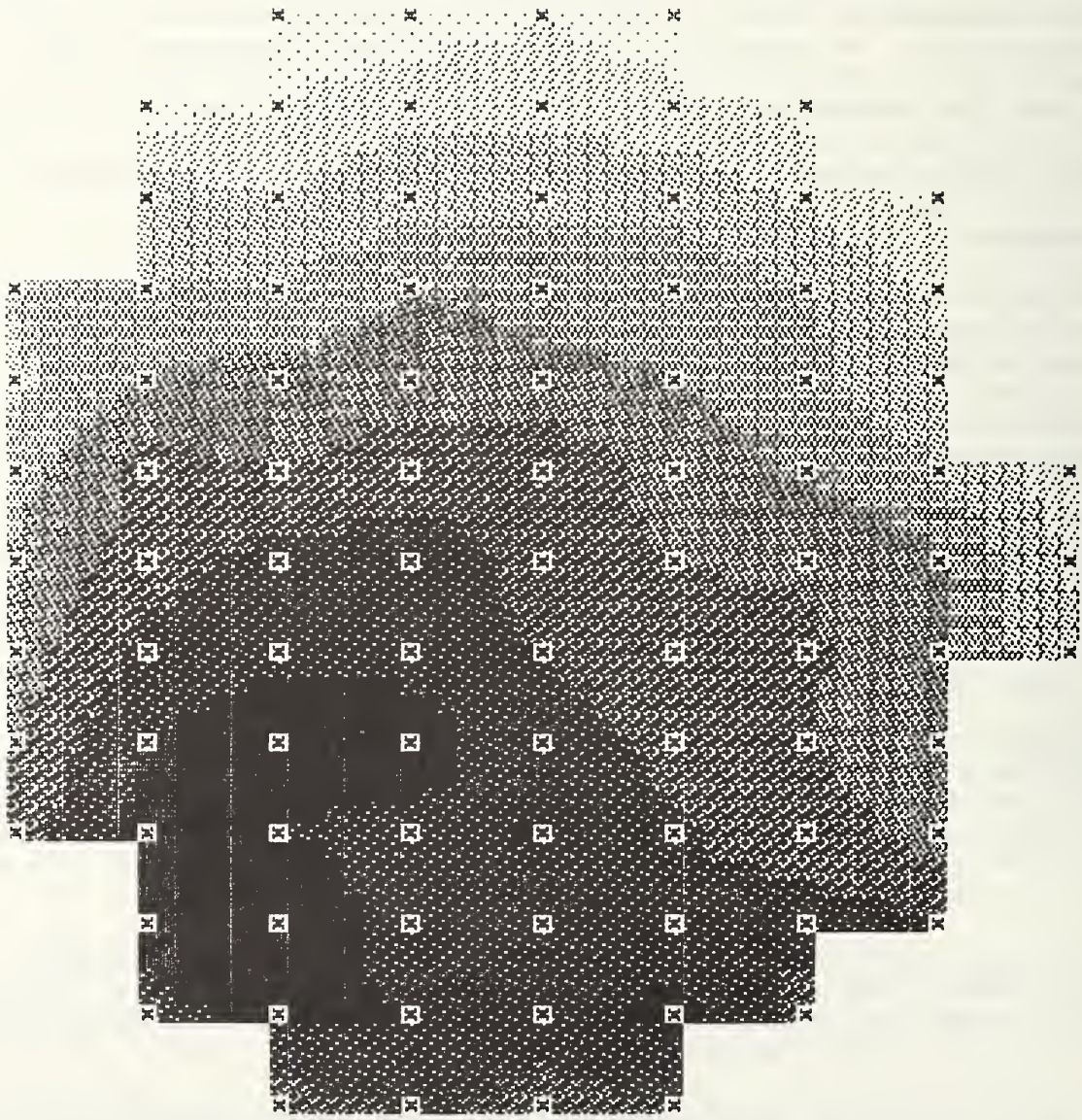


Figure 4. Sample wafer map.










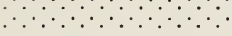
PARAMETER VALUE	# SITES
272.37695 TO 275.93018 	10
268.82373 TO 272.37695 	17
265.27051 TO 268.82373 	17
261.71729 TO 265.27051 	15
258.16406 TO 261.71729 	11
254.61110 TO 258.16406 	8
251.05791 TO 254.61110 	8
247.50471 TO 251.05791 	7
SITES INCLUDED	93
x x x x	INCLUDED SITES WITHIN SCALE
+ + + +	INCLUDED SITES ABOVE SCALE
- - - -	INCLUDED SITES BELOW SCALE
SAMPLE MEAN	263.40991
SAMPLE STD DEV	7.40973
SAMPLE MEDIAN	264.75781

Figure 5. Key which accompanies map of figure 4. (The number of significant figures is an artifact of the format used to represent the bin boundaries and does not reflect the precision of experimental data.)

the map is still reserved. A label may also be entered from remote mode by placing it immediately following the MAP command. If P3 is positive, the plot is automatically scaled between the minimum and maximum included data values. The minimum and maximum must be current, however, with no site exclusions or reinsertions or change in N value since the last PRS command. If P3 is negative, the range is autoscaled to the  $MEAN \pm K * SIGMA$  limits where K is the multiple of SIGMA calculated by the XOL command on its last pass. If this option is used, an XOL command must have been issued with no intervening site exclusions or reinsertions or change in N value. If P3 is zero, then the scaling is based on P5, the lowest value in the lightest shade, and P6, the highest value in the darkest shade. P6 must be greater than P5, and P5 and P6 must be present even if P3 is nonzero, although they are then ignored. The map is drawn on a new page. On the following page a key is drawn giving the range of values represented by each gray-tone, the number of test sites having values which lie within the range denoted by each gray tone, and the mean, sample standard deviation, and median of the included sites. If a median could not be determined when the PRS command was invoked, a value of 0 is printed for the median. On the plot, the locations of included test sites are represented by an 'X' or by a '+' or '-' in the event that autoscaling is not used and the data values are greater than or less than the scaling values given by P5 and P6. Excluded sites are not represented by one of these symbols but by the appropriate gray tone. A sample map is shown in figure 4 with the accompanying key in figure 5.

(8) Data Base Operations: WDB, LDB, GET, SDB, and DEL

WDB,P1,P2,P3,P4,P5 - Write to Data Base. Writes an entry in the data base where P1 is the pattern number, P2 is the lot number, P3 is the wafer number, P4 is the device number, and P5 is the parameter code to be associated with the entry. These must be non-negative integers having a maximum value of 999 for P1, P2 and P3, or 9999 for P4 and P5. The WDB command must be preceded by REA with no intervening GET or other WDB commands. After the entry has been written, the label and data information are printed on logical units 4 and 6.

LDB,P1,P2,P3 - List Data Base. Produces a listing of data base entries beginning with sequence number P1 and ending with sequence number P2. If P3 is zero, only the label information is printed; if P3 is nonzero, the data values in the sample are also printed. P2 must not be less than P1 or an error condition occurs. If P1 is greater than the largest sequence number, an error condition occurs. If either P1 or P2 or both refer to nonexistent sequence numbers, the message RANGE 1 NOT FOUND or RANGE 2 NOT FOUND or both is printed, but this is not considered an error condition if there are sequence numbers within the range bounded by P1 and P2.

P2 is a privileged parameter; that is, it can optionally be represented by only a minus sign. When P2 is a minus sign, LDB lists all records beginning at sequence number P1 through the end of the file. In the echo of the command to logical units 4 and 6, the minus sign appears as the value 0.111E+05.

When P3 is such that data values are also displayed, the data values immediately follow the label information, six data values per line in E-format. An

asterisk immediately following a data value indicates that the particular site was excluded.

GET,P1 - Get Sequence Number. The data base entry having sequence number P1 is defined as the reference data sample. P1 must be a legal sequence number not marked deleted or an error condition occurs. If GET is successful, the characteristics of the reference data sample are displayed.

SDB,P1,P2,P3,P4,P5,P6 - Search Data Base. Correlates the reference data sample with other entries in the data base. Parameters P1 through P5 are privileged parameters, such that they can optionally be represented in the command by a minus sign. If these are all minus signs, then correlation coefficients for all entries in the data base are calculated. When P1 is not a minus sign but is a numerical value, then correlation coefficient is calculated only for those data base entries for which the pattern number is P1. Similarly, P2 set to a numerical value restricts correlation coefficient calculation to entries for which the lot number is P2. Similarly, P3, P4 and P5 can be used to specify a particular wafer number, device number, and parameter code, respectively, for which correlation coefficient is to be calculated. Considerable flexibility is thus available for selectively searching a data base for possible correlations.

The sample correlation coefficient is calculated for all entries which are selected by parameters P1 through P5. Those entries for which the correlation coefficient is greater than or equal to P6 are printed on logical units 4 and 6. P6 provides a way of screening out entries for which correlation is poor. If P6 is zero, all results of calculations are printed.

If any of the data values in the reference data sample or the samples being searched represent excluded sites, these data values are not included in the correlation coefficient calculation. The printed output of SDB includes the number of site pairs included in each calculation. There is also a summary printed of the number of data base entries searched, number of entries selected for correlation coefficient calculation (matches), and number of results printed. A GET command must be executed prior to the SDB command.

DEL,P1 - Delete. Marks the data base entry having sequence number P1 as deleted by placing an asterisk in the DEL field of the label record. Such an entry is physically in the data base and is recognized by LDB commands but is not recognized by SDB commands and causes an error condition when referenced by GET or DEL commands. If DEL executes successfully, the entry which has been marked deleted is displayed.

### Error Messages

STAT2 contains numerous error messages intended to inform the user of error conditions and to prevent an error from causing the program to terminate abnormally. In this section, the various messages are discussed in the order of their origin. Lower case letters in the messages below represent numerical values which would appear in the actual messages.



(1) Errors associated with Command Syntax:

- \*\*\*\* NO CHARACTERS SHOULD FOLLOW MNEMONIC  
The command should have no parameters associated with it, but it has been represented as other than the three-letter mnemonic.
- \*\*\* UNDEFINED MNEMONIC  
The command mnemonic does not correspond to any of the legal mnemonics.
- \*\*\* COMMA ABSENT FOLLOWING MNEMONIC  
Either a command has been entered for which required parameters are not present, or a character other than comma has been used between the mnemonic and the first parameter.
- \*\*\* COMMAND HAS SUCCESSIVE COMMAS OR IS TERMINATED WITH A COMMA  
The command has adjacent or successive commas or ends with a comma.
- \*\*\* COMMAND HAS TOO MANY PARAMETERS  
The number of parameters following the mnemonic is more than the required number.
- \*\*\* NUMBER OF PARAMETERS/COMMAS EXCEEDS 8  
More than eight commas have been entered on the command line.
- \*\*\* COMMAND HAS TOO FEW PARAMETERS  
The number of parameters following the mnemonic is less than the required number.
- \*\*\* PARAMETER i HAS MORE THAN 20 CHARACTERS  
The ith parameter of the command has been represented by more than 20 characters.
- \*\*\* ILLEGAL CHARACTER IN POSITION i OF PARAMETER j  
The command contains an illegal character such as an alphabetic character or punctuation symbol.
- \*\*\* TOO MANY CHARACTERS IN EXPONENT OF PARAMETER i  
A parameter exponent contains more than two digits.
- \*\*\* PARAMETER i HAS NO DIGITS PRECEDING E  
The base of an E-format number is missing. This message also appears if a parameter consists of only a decimal point or plus sign or if a nonprivileged parameter consists only of a minus sign.
- \*\*\* PARAMETER i HAS NO DIGITS FOLLOWING E  
The exponent of an E-format number has been omitted.
- \*\*\* EXPONENT TOO LARGE IN PARAMETER i  
The ith parameter has an exponent whose magnitude is greater than 32.
- \*\*\* NPARS = i IS ILLEGAL VALUE  
An attempt was made to process a command having more than eight parameters without properly modifying subroutine COMPRT in ST2MAIN.



- \*\*\* NSUBR = i IN FCALL  
A new command has been added without properly modifying subroutine FCALL in ST2MAIN.
- \*\*\* ASG CALLED IN FCALL  
The command processor has been altered so that ASG commands are not properly processed.
- (2) Errors Associated with Precedence Violations:
- \*\*\* NO DATA FILE HAS YET BEEN READ  
An input data file must be read by the REA command before the requested operation can be executed.
- \*\*\* NO FIT MADE FOR CURRENTLY INCLUDED SITES  
The SPL or SQD command must be preceded by an FPL or FQD command, respectively, with no intervening site exclusions or re-insertions.
- \*\*\* MUST EXECUTE PRS TO GET CURRENT STATISTICS  
A message produced by XNS, HIS, PLT, and MAP whenever the maximum, minimum, and standard deviation values have not yet been calculated for the present population.
- \*\*\* GET MUST PRECEDE SDB  
A reference data sample must be specified by GET before SDB can be executed.
- \*\*\* WDB ILLEGAL AFTER GET OR WDB OR BEFORE REA  
An attempt has been made to write a data base entry under improper conditions. WDB is intended to be executed no more than once after a given REA command with intervening data point exclusions.
- (3) Errors Associated with Site Specification:
- \*\*\* NO DATA POINT AT SPECIFIED ROW-COLUMN  
A row-column location at which there is no test site has been specified by a XIP, LIP, IIP, or AIP command.
- (4) Errors Associated with Noninteger Input Values:
- \*\*\* PARAMETER i MUST BE AN INTEGER  
A numerical value i has been entered for a parameter where an integer is required.
- \*\*\* PARAMETER 1 MUST BE A POSITIVE INTEGER  
The first parameter of GET or DEL as entered is not a positive (non-zero) integer.
- \*\*\* PARAMETER MUST BE A POSITIVE INTEGER  
One of the two parameters of the LDB command has been given an illegal value.

(5) Errors Associated with Mapping:

- \*\*\* PLOT WIDTH OR HEIGHT TOO BIG OR TOO SMALL  
An improper size has been entered for width or height in the MAP command. See the MAP command description for the allowed range of values.
- \*\*\* UPPER SCALE VALUE < = LOWER  
If autoscale is not enabled in a PLT or MAP command, then the upper scale value must be larger than the lower scale value.
- \*\*\* TYPE = i in MAP  
A topological type outside the range -1 to 9 has been detected in subroutine MAP in ST20V5. This error should not occur and may be caused by portability problems.
- \*\*\* XOL K VALUE IS NOT CURRENT  
The autoscale option of MAP with P3 negative or of PLT with P1 negative has been invoked, but subroutine XOL has not first been executed with no intervening site exclusions or reinsertions or change in N value.

(6) Errors Associated with Data Base Commands:

- \*\*\* SEQUENCE NUMBER i NOT FOUND  
A sequence number which does not exist has been referenced by LDB, GET, or DEL.
- \*\*\* INVALID RANGE, RANGE 1 > RANGE 2  
The range of sequence numbers of the LDB command has the first sequence number greater than the last sequence number.
- \*\*\* NO LABELS IN THIS RANGE  
The range of sequence numbers specified in the LDB command contains no entries although there are entries at larger sequence numbers.
- \*\*\* SEQUENCE NUMBER i MARKED DELETED  
An entry referenced by GET or DEL cannot be accessed because it is marked deleted.
- \*\*\* PARAMETER VALUE TOO LARGE  
In a WDB command one of the parameters exceeds the maximum allowable value. The limit is 999 for P1, P2 and P3 and 9999 for P4 and P5.
- \*\*\* ILLEGAL NEGATIVE PARAMETER  
One of the first five parameters of the SDB or WDB commands has been given a negative value.
- \*\*\* PARAMETER 6 MUST BE IN THE RANGE 0 TO 1  
The sixth parameter of the SDB command has been given an illegal value.
- \*\*\* HEADER FIRST BYTE IS "b" RATHER THAN "H".  
See below.

\*\*\* RECORD FIRST BYTE IS "b" RATHER THAN "D".

See below.

\*\*\* LABEL SEQUENCE = i, HEADER SEQUENCE = j.

See below.

\*\*\* LABEL RECORD No. = i, HEADER POINTER = j.

See below.

\*\*\* LABEL SEQUENCE = i, DATA SEQUENCE = j.

The preceding five messages indicate that the pointers which link the label file with the data file have been corrupted such as by improper removal of entries from the data base or incorrect file assignments. Refer to the section titled Data Base Structure. In the top two messages, the "b" represents the actual character which appears in the first byte position.

\*\*\* UNABLE TO FIND EOF ON L.U. i

Subroutine PEND in ST20V7, which identifies the record number of the last record in the label or data file, did not execute properly. This probably indicates a portability problem.

\*\*\* LABEL FILE HAS ILLEGAL SAMPLING PLAN CODE

The sampling plan code in the first record in the label file has an illegal value. Refer to the section titled Data Base Structure.

(7) Errors Associated with REA and ASG:

\*\*\* FORMAT TYPE MUST BE IN THE RANGE 1 to 6

The first parameter of the REA command has been given an illegal value.

\*\*\* SUBARRAY MUST BE IN THE RANGE 1 to 6

The second parameter of the REA command has been given an illegal value.

\*\*\* IFMT = i IN REA

An illegal format number has been detected in subroutine REA in ST20V3. This probably means that line labeled 14 was not properly modified when a new format was added.

\*\*\* L.U. 1 NOT ASSIGNED SO CANNOT CLOSE

An attempt has been made to use the ASG command when no input file has been assigned to logical unit 1.

\*\*\* ERROR IN ATTEMPT TO CLOSE L.U. 1

For an undetermined reason, logical unit 1 cannot be closed by the ASG command.

\*\*\* VOLUME ERROR

The volume name specified in the file descriptor of the ASG command is either off line or nonexistent.

\*\*\* FILENAME ERROR  
No file can be found having the file descriptor specified in the ASG command.

\*\*\* FILE DESCRIPTOR SYNTAX ERROR  
The file descriptor in the ASG command is not syntactically correct.

\*\*\* ERROR IN ATTEMPT TO OPEN L.J. 1  
For an undetermined reason, logical unit 1 cannot be opened by the ASG command.

(8) Other Errors:

\*\*\* PAUSED ON ERROR IN REMOTE MODE  
An error has occurred while in remote mode, causing STAT2 to pause.

\*\*\* MATRIX APPEARS SINGULAR  
An FPL or FQD command is unable to execute because of a singularity in a matrix which must be inverted. This condition should not occur when operating on real data.

\*\*\* TYPE = i AT SITE (j, k)  
A topological type outside the range -1 to 9 has been detected by subroutine AXP in ST20V4. This condition probably is caused by portability problems.

\*\*\* NEGATIVE VALUE OF N NOT ALLOWED  
An illegal value of N has been specified in the ENN command.

\*\*\* ILLEGAL VALUE FOR DELTA  
The parameter of the XOL command as specified lies outside the range 0.05 to 0.90.

\*\*\* ILLEGAL P VALUE IN NORPPF  
This message originates in a subroutine called by XOL but is not expected to occur so long as the restrictions on DELTA given in the previous error message are met.

\*\*\* DENOM IS 0.0000 IN R  
This error is produced by an SDB command when a reference sample of all zeroes is correlated with another sample of all zeroes.

\*\*\* CORRELATION COULD NOT BE DONE FOR SEQUENCE NUMBER i  
Correlation coefficient could not be calculated because too many data values in the pair of data samples represented excluded sites.

#### Addition of New Input Data Formats

Users may want to add the capability to read additional input data formats using the REA command. Instructions for doing so are given in the program listing in the introductory comments to subroutine REA, beginning in line 53 of ST20V3 in Appendix V.



### Addition of New Commands

Users may find it desirable to add one or more new commands to STAT2. Instructions for adding new commands are contained in the listing of ST2MAIN in Appendix II beginning at lines 334, 390, 805, and 905.

### Program Installation

To produce an executable module which can run in the specified environment, compile ST2MAIN, ST20V1, ST20V2, ..., ST20V7, using Interdata FORTRAN VI, then assemble the outputs of these compilations and subroutine FETCH to produce nine object files. Use the Interdata Task Establishment Task (TET) to build a main segment from ST2MAIN and FETCH and seven overlays from ST20V1 through ST20V7. FORTRAN library routines must be added as needed to the main segment and each overlay.

Files 10 and 11 are stand-alone programs which support STAT2 as explained in the section titled Data Base Structure.

### Logical Unit Assignments

The logical unit assignments used by STAT2 are given below. The logical unit number precedes the description. The logical units should be assigned after the program is loaded. As the parenthetical notes indicate, not all logical units may be needed for a given set of operations.

1. Input data file (not used if only data base operations are to be done).
2. Output data file (used only with WRT command). File can be tape or disc. Output can also be to a printing device.
4. Line printer, 132 columns.
5. Command input, normally a CRT terminal or a disc command file.
6. A CRT terminal for command echoes and other output.
7. Data base label file (used only for data base operations).
8. Data base data file (used only for data base operations).
9. Scratch file used for constructing the gray-tone map. Record length must be 132 bytes (not used if maps are not being generated).
11. Overlay 1.
12. Overlay 2.
13. Overlay 3.
14. Overlay 4.
15. Overlay 5.
16. Overlay 6.
17. Overlay 7.

The differences between logical units 4 and 6 are (1) the command line as typed is echoed to logical unit 6 as are any command syntax error messages, whereas the decoded command and parameters are written to both logical units 4 and 6; (2) gray-tone maps produced by MAP go only to logical unit 4; and (3) wafer displays produced by PLT go only to logical unit 6.

## Program Portability

As stated in the Introduction, STAT2 is written to run on an Interdata 7/32 minicomputer in conjunction with a Printronix P300 line printer/plotter. Aspects of the program which may produce portability problems are (1) use of halfword integers, (2) calls to extensions of the run-time library, (3) the SHADE subroutine in ST20V5 which creates the gray-tone maps, (4) three options associated with formatted READ statements, (5) the FETCH subroutine which is written in assembly language, (6) the manner in which overlays are loaded, and (7) the record numbering convention. In some cases the user may wish to remove the nonportable feature as not needed; in other cases he will want to write a replacement.

There are several variables which are declared INTEGER\*2, meaning a 16-bit or halfword integer, rather than INTEGER, which specifies a 32-bit integer. The halfword integer declarations are necessary for NBS to read existing data using STAT2. Users who are getting STAT2 to run on another system may want to declare these variables in some other manner.

Extensions of the run-time library are used in several places. Subroutine ASG in ST20V3 dynamically closes and reassigns logical unit 1 to a new input data file. This is not an essential feature but is useful when reading in many sets of input data. On the Interdata system it is necessary if multiple input data files are to be processed in remote or batch mode, that is, by a command file. Subroutine DATTIM contains calls to acquire the current date and time for inclusion in the data base label record. This is not essential, but there are many instances where a label record is read and the date and time information are expected to be present.

The subroutines in overlay 5 (ST20V5) contain calls to load and shift bytes as the data words are constructed by which the printer/plotter produces the gray tones. These subroutines also contain calls to SYSIO, a generalized I/O subroutine. These subroutines are written to address the Printronix P300 line printer/plotter and would have to be adapted to other plotters. It is also assumed that both printing and plotting can occur on the same device.

The formatted READ statement can contain as many as three specifications in addition to the logical unit number, format statement label, and variable name list. An example of such a READ statement is:

```
READ (lu, fmt, REC=nrec, ERR=ln1, END=ln2)X
```

The REC= option permits the user to read a particular record number where nrec is an integer or integer variable name. The ERR= option permits the user to branch to label ln1 if an error occurs. The END= option permits the user to branch to label ln2 if an attempt is made to read beyond the end of file. The ERR= option is probably not necessary, but users not having the other two options would need to set up some scheme for reading a particular record and sensing end of file.

Subroutine FETCH is written in assembly language and is listed in Appendix X. When CALL FETCH (A,B) is executed, the file descriptor corresponding to logical unit A is printed on the device which is assigned to logical unit B.

This is not an essential feature but does identify all print-out with a particular file descriptor and therefore avoids possible confusion. To remove this feature, delete all lines containing 'CALL FETCH' from STAT2.

Overlays are loaded by subroutine IFETCH, a component of Interdata's run-time library. Calls to IFETCH occur in subroutine FCALL and in the main section of the command processor, both in ST2MAIN. Each overlay is given a name which must be identified with the overlay during linking. Users of other systems will need to use an appropriate scheme to load the overlays.

In Interdata FORTRAN VI, the first record in a file is record number 0. In an environment in which the first record is record number 1, the scheme by which label records and data records in the data base are written and read would have to be altered. The code by which the data base is handled is in ST20V7.

Most of these instances of questionable portability are called to the user's attention in the program listing by a comment statement beginning with 'COMMENT' in the first seven character positions.

As part of the STAT2 documentation, SUBROUTINE statements are often preceded by comment statements which go with the subroutine. Such comments provide introductory explanations to subroutines or a family of subroutines. If a particular compiler requires that SUBROUTINE be the first statement in a subroutine, the user will have to make this change.

#### Acknowledgments

The authors are indebted to Charles P. Reeve for supplying some of the statistical techniques used in STAT2 and to Loren W. Linholm, Charles L. Wilson, and Kenneth F. Galloway for careful readings of and helpful suggestions regarding the manuscript. The manuscript was prepared with proficiency by E. Jane Walters.

### References

1. Linholm, L. W., Mattis, R. L., Till, L. J., Frisch, R. C., and Reeve, C. P., Analysis of PVW Data, Semiconductor Progress Briefs, NBSIR 81-2230, W. M. Bullis, ed., pp. 4-6, March 1981.
2. Linholm, L. W., Mattis, R. L., Frisch, R. C., and Reeve, C. P., Characterizing and Analyzing Critical Integrated Circuit Process Parameters, *Semiconductors Silicon/1981*, H. R. Huff, R. J. Kriegler, and Y. Takeishi, eds., pp. 906-920 (ECS, Pennington, New Jersey, May 1981).
3. Carver, G. P., Linholm, L. W., and Russell, T. J., The Use of Microelectronic Test Structures to Characterize IC Materials, Processes, and Processing Equipment, *Solid State Technology* 23 (9), 85-92 (September 1980).
4. Odeh, R. E., and Evans, J. O., The Percentage Points of the Normal Distribution, Algorithm AS70, Applied Statistics, *J. Royal Statistical Soc.* 23 (1), 96-97 (1974).
5. Snedecor, G. W., and Cochran, W. G., *Statistical Methods*, 6th edition, chapter 2 (Iowa State University, 1967).
6. Gross, A. M., Confidence Interval Robustness with Long-Tailed Symmetric Distributions, *J. Am. Stat. Assoc.* 71, 409-416 (1976).



## FEDERAL INFORMATION PROCESSING STANDARD SOFTWARE SUMMARY

01. Summary date Yr. Mo. Day 8 2 0 2 1 2			02. Summary prepared by (Name and Phone) Richard L. Mattis (301) 921-3621			03. Summary action New Replacement Deletion <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Previous Internal Software ID		
04. Software date Yr. Mo. Day 8 2 0 2 1 2			05. Software title A Computer Program for Analysis of Data from Microelectronic Test Structures			07. Internal Software ID STAT2		
06. Short title STAT2			08. Software type <input type="checkbox"/> Automated Data System <input checked="" type="checkbox"/> Computer Program <input type="checkbox"/> Subroutine/Module			09. Processing mode <input type="checkbox"/> Interactive <input type="checkbox"/> Batch <input checked="" type="checkbox"/> Combination		
11. Submitting organization and address Semiconductor Devices and Circuits Division National Bureau of Standards Washington, DC 20234			10. Application area <u>General</u> <input type="checkbox"/> Computer Systems Support/Utility <input checked="" type="checkbox"/> Scientific/Engineering <input type="checkbox"/> Bibliographic/Textual <u>Management/Business</u> <input type="checkbox"/> Process Control <input type="checkbox"/> Other			<u>Specific</u> Analysis of Integrated Circuit Process Validation Wafer Test Data		
12. Technical contact(s) and phone Richard L. Mattis (301) 921-3621						13. Narrative The program performs the following functions: read data as a two dimensional array in several formats; calculate mean, standard deviation, median; identify outliers; calculate replacement values for outliers; make a gray tone data map on a line printer; make a character map on the user's terminal; construct a data base for making correlations among various data sets; search said data base using several selective keys.		
14. Keywords Process validation wafer; wafer map; statistical analysis; correlation coefficient								
15. Computer manuf'r and model INTERDATA 7/32			16. Computer operating system OS32MT43			17. Programing language(s) FORTRAN		18. Number of source program statements About 7000
19. Computer memory requirements 70 KB			20. Tape drives Optional			21. Disk/Drum units Disk needed for overlays, map scratch file, data input		22. Terminals 1 video or hard copy
23. Other operational requirements Uses overlays								
24. Software availability Available <input checked="" type="checkbox"/> Limited <input type="checkbox"/> In-house only <input type="checkbox"/>				25. Documentation availability Available <input checked="" type="checkbox"/> Inadequate <input type="checkbox"/> In-house only <input type="checkbox"/> A user's manual is available as NBS Internal Report				
26. FOR SUBMITTING ORGANIZATION USE								



U.S. DEPT. OF COMM. <b>BIBLIOGRAPHIC DATA SHEET</b> (See instructions)	<b>1. PUBLICATION OR REPORT NO.</b> NBSIR 82-2492	<b>2. Performing Organ. Report No.</b>	<b>3. Publication Date</b> June 1982
<b>4. TITLE AND SUBTITLE</b>  A Computer Program for Analysis of Data from Microelectronic Test Structures			
<b>5. AUTHOR(S)</b> R. L. Mattis, L. J. Till, and R. C. Frisch			
<b>6. PERFORMING ORGANIZATION</b> (If joint or other than NBS, see instructions)  <b>NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20234</b>		<b>7. Contract/Grant No.</b>	<b>8. Type of Report &amp; Period Covered</b>
<b>9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS</b> (Street, City, State, ZIP)			
<b>10. SUPPLEMENTARY NOTES</b>  <input checked="" type="checkbox"/> Document describes a computer program; SF-185, FIPS Software Summary, is attached.			
<b>11. ABSTRACT</b> (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here) A computer program, STAT2, is described which performs the following functions: reads data as a two-dimensional array; calculates mean, sample standard deviation, and median; identifies outliers; calculates replacement values for outliers; makes a gray-tone data map on a line printer; makes a character map on the user's terminal; constructs a data base for examining correlations among various data sets; and searches the data base for correlations using several selective keys. The emphasis in this document is on program usage, and detailed descriptions of the commands are given. Program portability and data input requirements are addressed. Guidance regarding several types of program modifications is provided.			
<b>12. KEY WORDS</b> (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons) Computer program; correlation coefficient; outlier; process validation wafer; statistical analysis; two-dimensional arrays; wafer map.			
<b>13. AVAILABILITY</b> <input checked="" type="checkbox"/> Unlimited <input type="checkbox"/> For Official Distribution. Do Not Release to NTIS <input type="checkbox"/> Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. <input checked="" type="checkbox"/> Order From National Technical Information Service (NTIS), Springfield, VA. 22161		<b>14. NO. OF PRINTED PAGES</b> 38	<b>15. Price</b> \$7.50









NBSIR 82-2492

A COMPUTER PROGRAM FOR ANALYSIS OF DATA FROM MICROELECTRONIC TEST STRUCTURES

1 OF 2

