

# DEUTSCHES ELEKTRONEN-SYNCHROTRON **DESY**

DESY 85-046  
May 1985



A DATA PROCESSING SYSTEM BASED ON THE 370/E EMULATOR

by

D. Notz

*Deutsches Elektronen-Synchrotron DESY, Hamburg*

ISSN 0418-9833

NOTKESTRASSE 85 · 2 HAMBURG 52

DESY behält sich alle Rechte für den Fall der Schutzrechtserteilung und für die wirtschaftliche Verwertung der in diesem Bericht enthaltenen Informationen vor.

DESY reserves all rights for commercial use of information included in this report, especially in case of filing application for or grant of patents.

To be sure that your preprints are promptly included in the  
HIGH ENERGY PHYSICS INDEX ,  
send them to the following address ( if possible by air mail ) :

DESY  
Bibliothek  
Notkestrasse 85  
2 Hamburg 52  
Germany

A Data Processing System based on the 370/E Emulator

Dieter Notz

Deutsches Elektronen-Synchrotron, DESY, Hamburg, Germany

Description

The main components of the 370/E are shown in Fig. 1. The 370/E consists of 14 boards with the dimensions 39.4cm \* 23.5cm. The whole processor therefore fits into a box of a typical crate size 45cm \* 30cm \* 40cm. The arithmetic and logic unit is divided into five parts. An integer CPU, a dedicated multiplier, two floating point boards and a control unit. The eight memory boards may contain up to 2 Mbytes of memory. If desired the backplane can be easily increased to give space for 4 Mbytes. From the address space point of view the processor can be equipped with 16 Mbytes of memory.

Connections to Host Computers

In our application the 370/E has no I/O devices. It is controlled by a host computer. At DESY an interface has been built for PADAC which allows a connection of the 370/E to a NORD 10, NORD 100, VAX, PDP 11, TMS9900 or NS32016.

The transfer rates without DMA setup time are 1.25  $\mu$ sec/byte to a NORD100/Emulator, 1.5  $\mu$ sec/byte to a VAX and 1.57  $\mu$ sec/byte to a TMS9900. Of the variety of possible connections we describe here only two alternatives, an online and an offline application.

Online Application

Fig. 2 shows a typical online application. At the beginning of a data taking run a prepared load module and the latest constants are transferred from the IBM via the online net to the online computer (NORD, VAX, PDP11) which then loads the program into the 370/E. The constants are stored on local disks. The emulator is then started and gets first the constants and afterwards the experiment's data as they are read out by the online computer. Due to the double buffering in the I/O system the processor can analyse the first event while the second event is read in. From the programmer's point of view one only has to read an event with READ(1,END=4), (EVENT(I),I=1,L) and output it with a WRITE(2)... statement. All error messages and run summaries can be transferred to the online computer via a WRITE(6,...) and printed there. At the end of a run an end-of-file is generated which will close all files and halt the processor.

Offline Application

Fig. 3 shows the offline application. The user sits at the IBM terminal and the 370/E is connected to the IBM via the online net and a TMS9900 microprocessor. To the user, the 370/E looks like an attached processor to the IBM although it is 500m away from the computer center. The TMS9900 acts as a host, checks the connection to the IBM and to the 370/E and looks every 2 minutes at the jobqueue on the IBM to determine whether or not a job has been submitted. The program in the TMS9900 runs for ever and only needs

Abstract

This paper describes the DESY implementation of a 370/E based data processing system. The 370/E was designed at the Weizmann Institute in Israel by a team led by Hanoeh Brafman and emulates an IBM 370/168 mainframe computer. This system can process large megabyte sized programs with a speed approximately 1/4 that of an IBM 3081D mainframe. Four processors are connected via PADAC interfaces to the IBM, NORD, VAX or TMS9900.

Introduction

There is an increasing demand for computer power in high energy physics. In the era of the forthcoming accelerators a data production rate of 400 tapes per day is estimated. All these data have to be analysed and compared with theoretical predictions. People designing accelerators need computers to simulate the beam optics. These programs are not I/O intensive and need a lot of number crunching power. In order to support physicists with cheap and IBM compatible computer power the 370/E emulator has been developed at the Weizmann Institute by H. Brafman (Ref.1). Emulation is defined as "the desire to equal or surpass a rival". In this sense the 370/E is a computer which from the user's point of view is indistinguishable from an IBM 370. In high energy physics the term emulator has become associated with the SLAC 168/E which was designed by P. Kunz (Ref. 2). The 168/E was a successful product and many systems have been built. However, the 168/E with its limited access and separated memory for data and instructions could not run all programs without considerable user involvement. Especially formatted I/O was painful and normally not used. The advantage of the 370/E is its architectural similarity to the IBM architecture. A combined memory for data and instructions is used and the IBM instructions are emulated directly. Therefore one does not need to translate the programs before running them on the emulator. One only has to link the program together with the 370/E system and the FORTRAN I/O routines and download it. The price of the processor is 45 kDM with a 2 Mbyte memory. Approximately 23 processors are in operation in High Energy Physics (Ref. 3) so far. A new version which is 20% faster has been completed at the Weizmann Institute.

restarting in case of a power failure. The user who wants to submit a job sits in front of an IBM terminal in his known environment. We assume that a big program which has already been developed by several people should run on the 370/E. The following steps are then needed :

- a) Prepare a Load Module  
The load module is built by the IBM linkage editor. This can be performed in different ways: One can use the LKED procedure under NEWLIB which must load the 370/E system first and afterwards all user's programs
- or  
one can run a small batch job which links all routines and libraries. Fig. 4 shows an example of such a job.
- b) Allocate all Files  
All files which should be accessed by the 370/E must reside on disks. In addition one has to create a file LISTFILE which will contain the printout of the 370/E.
- c) Prepare Job Control Cards  
As in all IBM jobs one has to inform the system of the files which should be opened for each unit. Also the name of the load module and the time limit must be given. The job control file is stored in the user's library. Fig. 5 gives an example.
- d) Submit job  
In order to submit a job the user must give the submit command S370 or CALL 'TASSOL.LIBRARY(SUBM370E)' and must type the names of the file containing the job control information (i.e. TASSOL.SOURCEJCL370). The job is now placed into the jobqueue and will be executed later (Fig. 6).
- e) Check Jobstatus  
With the command J370 or CALL 'TASSOL.LIBRARY(JOBS370E)' the user gets a list of the last 16 jobs in the 370/E. He can estimate how long he has to wait before the job will be started (Fig. 7). If the job is running he may cancel the job by the CANCEL command or may look at the printout by LIST 'TASSOL.LIST370E'.

The 370/E Operating System

The operating system of the DESY 370/E is adapted to our needs and environment. Only a single user runs on the processor at one time. We do not support any multitasking. The processor is connected to an IBM and should support all I/O facilities the user normally gets on the mainframe like sequential READ, WRITE, REWIND, direct access READ, WRITE, FIND and full support in case of errors like divide check or negative SQRT. In order to get this service all programs doing input/output must be written in FORTRAN IV or FORTRAN 77 and must be compiled by the IBM compiler. The layout of the operating system is shown in fig. 8. The first locations are fixed and allocated to program status words (PSW's) and channel address

(CAW) and channel status words (CSW's) as in the IBM 370. The first word contains the PSW for initial program load (IPL) to start the program. The section for unsupported operation code contains routines to simulate some instructions which are not implemented in the hardware like move character long MVCL or CLCL. For REAL\*16 operations one can load a simulation package (TEAKPALL) from the system link library (SYS1.LINKLIB) to which control is transferred.

The supervisor call handler (SVC) supports the following IBM supervisor calls:

- SVC 3, EXIT, to terminate a task
- SVC 4, GETMAIN, to allocate dynamic memory
- SVC 5, FREEMAIN, to release dynamic memory
- SVC 8, LOAD, to load a member declared by IDENTIFY
- SVC 9, DELETE, to delete a member
- SVC 10, GETMAIN, to allocate dynamic memory
- SVC 13, ABEND, to release dynamic memory
- SVC 14, SPIE, to terminate a task abnormally
- SVC 35, WTO, to set or cancel SPIE exit
- WTOR, to write to the operator
- WTOR, to write to operator and reply
- SVC 40, EXTRACT, to provide information from task control block
- SVC 41, IDENTIFY, to add an entry point to a copy of a load module
- SVC 60, STAE, to set or cancel STAE exit.

All information concerning open files is stored in the IHOUAC table. This table indicates which unit is open for sequential or direct access I/O. Before a load module is downloaded the IBM opens all files and transfers the data control block (DCB) Parameters into this table. By this method the 370/E knows which files are accessible and which record length and blocksize should be used.

All other constants from outside like date, time, size of program and jobname are inserted into the load module on fixed locations before downloading.

The rest of the operating system belongs to the FORTRAN input/output package. The operating system is linked in front of the user's program by an INCLUDE TASSO(SYST370E) statement in the linkage editor. The user's main program and all other subroutines are loaded in the middle. The remaining space is used for dynamic allocation of I/O buffers or histogram routines.

Input/Output for IBM FORTRAN programs

Fig. 9 indicates the user's program on the IBM written in FORTRAN. All I/O requests to files must be transferred in such a way that the user does not know whether his program runs on the IBM or on the emulator. This is done in the following way (Ref. 4): Each FORTRAN program which was generated by the IBM compiler generates a call to IBCOM# for each READ or WRITE. A lot of parameters like addresses and FORMAT statements are exchanged between the program and the FORTRAN I/O package. IBCOM# then does the formatting and transfers buffers to FIOCS#. Here only a few parameters like the unit number, I/O request, buffer address and buffer length are exchanged.

In the case of the 370/E the FORTRAN runtime library has been split into two parts: IBCOM# runs on the 370/E processor and FIOCS# runs on the IBM or host. For direct access I/O the routines DIOCS1 and DIOCS4 are used. The corresponding program on the IBM is DIOCS# and DEFILE. IBMTRA is the actual transfer routine between the 370/E and the host. As the information which is exchanged between the 370/E and the IBM is well known the IBM can easily be replaced by a minicomputer like NORD or VAX as long as the files are delivered in IBM format.

The transfer speed between the 370/E and the IBM is 5 usec/byte. In order to avoid a slowing down of the processor caused by this transfer rate several levels of pipelining are used: The processor has two buffers for each I/O unit and the IBM also has two buffers. For sequential input and output and for direct access output the processor can continue with its calculation while the transfer takes place. For direct access input (!) the processor must wait until the record is really shipped down from the IBM disk into the 370/E memory. If several consecutive direct access records are read into the user's program the 370/E issues a FIND request to the IBM so that the next record can be transferred while the program is still operating on the previous data.

Status and Performance

Fig. 10 shows an example of a job which was executed on the 370/E. The only indication that the job did not run on an IBM but on the emulator are the addresses in the traceback of the error messages. All addresses are the same as in the linkage editor.

Four 370/Es are running at DESY. One processor has operated for a year and has executed 1116 jobs using 150 000 min CPU (370/E time). The job profile can be seen in fig. 11. Apart from short tests many jobs remain in the 370/E for several hours. The only problems which occur from time to time are breakdowns of the IBM link and the IBM online system. In addition, some jobs need a lot of data from the mass storage device. If these data cannot be delivered within a time limit of 1 minute the job will be cancelled in order to release the link. Normally data are transferred from the mass storage system to disk when the job is submitted by the user. If the waiting time for job execution is not too long data remain on disk and are available when the job is started.

The processor does not introduce any problems to the users once they have learned how to build an IBM load module. From the point of view of the computer center the 370/E looks like one of the 40 online jobs which are running in the mainframe.

Acknowledgement

The work described in this paper was only possible due to the enormous amount of work carried out by Hanoeh Brafman and his team. One of the processors at DESY was built at the Weizmann Institute and installed by David Richard Fall and Yaron Gal. The operating system was upgraded by David Botterill (RAL) and implemented at DESY by Rafi Yaari. A lot of the construction of the DESY processor was done by Kay Rehlich, Klaus Nimmer

(DESY) and Bob Hatley (RAL). The memory with 1/4 Mbyte per card was laid out by Chris Bebeck (Cornell). The routines in the TMS9900 were written by Martin Dieckvoß (Hamburg University) and the online link has been made available by Gerd Hochweller and his group.

References

- (1) H. Brafman et.al., A Fast General Purpose IBM Hardware Emulator, Weizmann Institute, Dept. of Nuclear Physics, Internal Report, January 1983
- Review on the Impact of Specialized Processors in Elementary Particle Physics, Padova, March 23-25, 1983.
- (2) P. Kunz et. al., Experience using the 168/E Microprocessor for Offline Data Analysis, SLAC-PUB-2418, October 1979
- (3) Institutions and contact persons using 370/E's (Number of processors in brackets):  
 Weizmann Institute, H. Brafman (3); Rutherford Lab., J. Barlow (2);  
 DESY, D. Notz, (4); Aachen, G. Peise (1); Bonn, M. Kokott (1);  
 Siegen, M. Rost (1); Imperial College, G. Fayers, (1); Birmingham,  
 H. Shaylor, (1); Tel Aviv, Y. Gnat, (1); Cornell, C. Bebeck, (6);  
 CERN, P. Schmid, F. Chevrier, DELPHI, OPAL (2).
- (4) D. Notz, The Input/Output Software for the 370/E Emulator, DESY, Internal Report FI-82/01, 1982 and TASSO Note No. 251, March 1983 (unpublished).

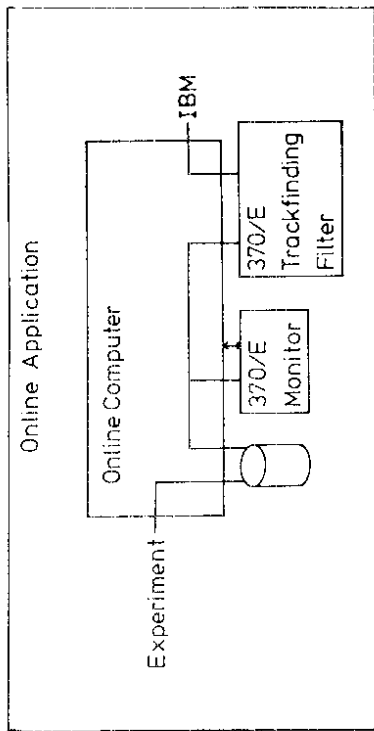


Fig. 2 Data from an experiment are written into a buffer or on a disk. Then they are analysed and filtered by the 370/E and sent to the IBM or on a tape.

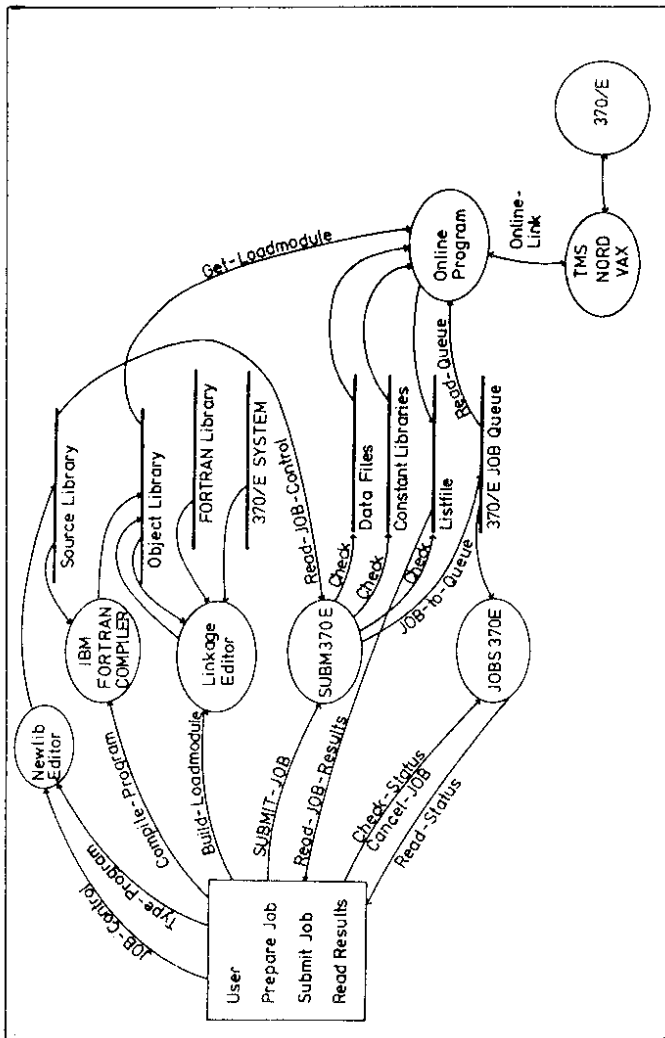


Fig. 3 Offline application. The user sits at the IBM terminal, writes a program (NEWLIB) and compiles it. Afterwards the program is linked (LINKAGE EDITOR) together with the 370/E operating system. The job is then submitted and has access to all IBM files. Results can be read from LISTFILE.

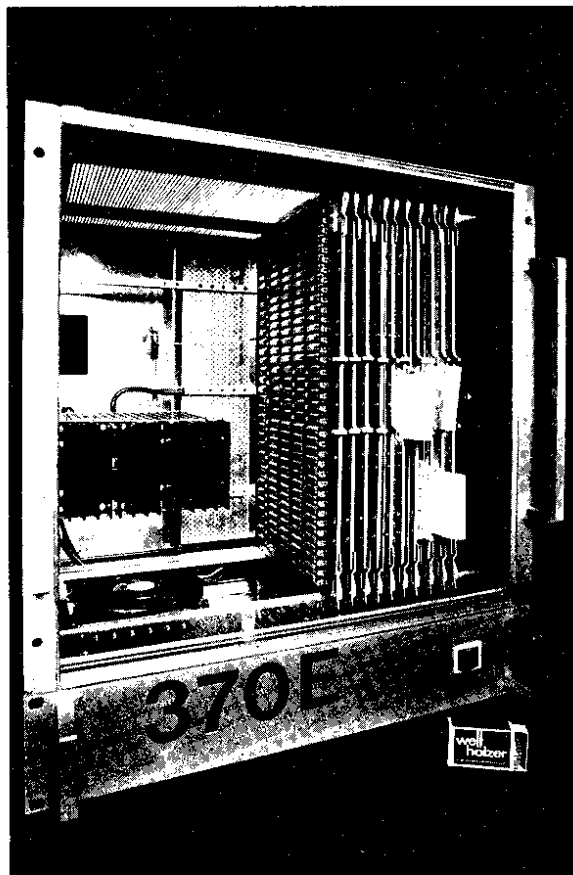
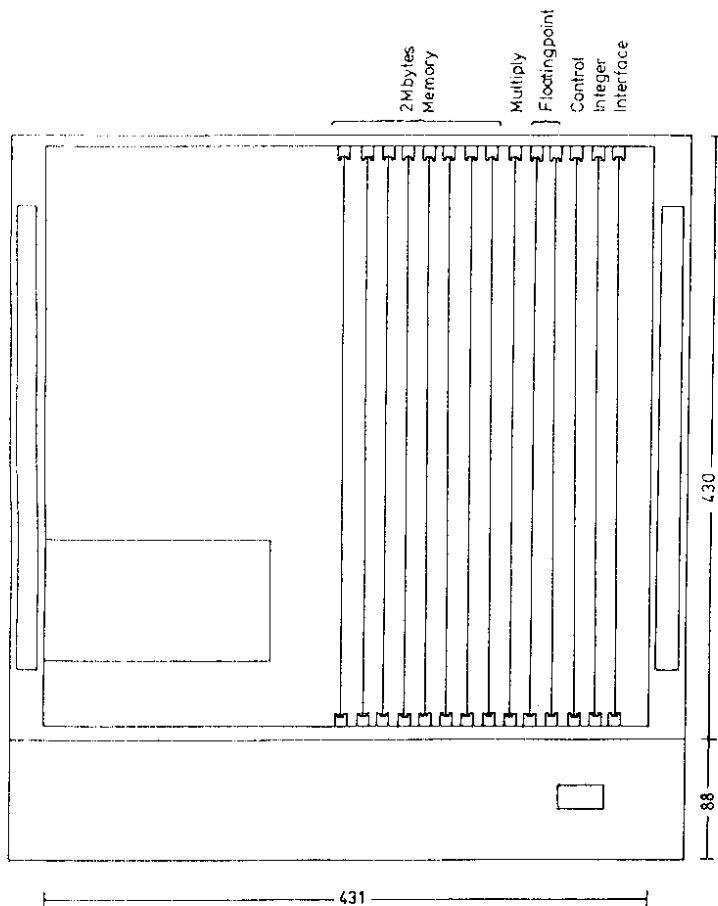


Fig. 1 The 370/E consists of 5 CPU boards, 1 interface board and 4 to 8 memory boards.

```

JOB '10601601,CEP-USER',NOTL,CLASS=L,TIME=(,5)
//MAIN ORG=EXT,KLEPRI=HIC
//EXEC PORTHCL,PARM=LKED=PAR,LIST'
CC
PORTHCL PROGRAM
STOP
END
//LKED,SYSLIN DD DDNAME=SYSJN
// DD DSN=CELESST,DISP=(OLC,DEL,L1)
// LKED,SYSLIB DD DSN=KOLBAS,GENL,DISP=SHR
// DD DSN=KOLBAS,GEPL,DISP=SHR
// LKED,SYSMOD DD DSN=FIBNOT,TSOLIBL(GEP470E),DISP=(SHR,ALF)
// LKED,TASSO DD DSN=TASSO,LIBRARY,DISP=SHR
// LKED,SYSIN DD DSN=FIBNOT,TSOLIBL,DISP=SHR
// INCLUDE TASSO(SYST370E)
//

```

Fig. 4 Example of a job to prepare a load module (compile and link). One can connect all libraries to the linkage editor. The 370/E system is loaded by an INCLUDE TASSO(SYST370E) statement.

```

//FIBNOT00 JOB TIME=5
//STRP00 EXEC PGM=FIBNOT,TSOLIBL(GEP470E)
//LISTFILE DD DSN=FIBNOT,LIST371L
//FT4700 DD DSN=FIBNOT,GE846
//FT4800 DD DSN=FIBNOT,GE848
//

```

Fig. 5 Job control language for a 5 minutes job and 2 output files for graphic information.

```

TYPE IN NAME OF FILE CONTAINING JOB CONTROL CARDS
EXAMPLE: TASSO.SOURCE(JCL370)
//GO,FT4700 DD DSN=FIBNOT,TSOLIBL(GEP470E)
//GO,FT4800 DD DSN=FIBNOT,TSOLIBL(GEP470E)
LENGTH OF PROGRAM 211832 00033E76
//
//ALLOCATE FILE 86
//GO,FT88F001 DD DSN=FIBNOT,LIST371E
//ALLOCATE FILE 46
//GO,FT88F001 DD DSN=FIBNOT,GEF46
//ALLOCATE FILE 48
//GO,FT88F001 DD DSN=FIBNOT,GEF48
//GO,FT88F001 DD DSN=FIBNOT,GEF48
JOB NO 1871 FIBNOT00 TIME= 5 SUBMITTED 10.370/L
PROCESSOR HAS CHECKED QUEUE AT 07/02/85 09:40.45 LAST JOB: FIBNOT00

```

Fig. 6 Submit a job. The user gives the file which contains the job control information (JCL). The length of the load module is checked and all files are allocated.

2860	JOB DONE	0	JOB WAITING	PROCESSOR	18/04/85	11.31.01
2845	F35PR500	800	F35PRS-BI-L(TO1)		18/04/85	13.27
2846	F35PR500	800	F35PRS-BI-L(TO1)		18/04/85	19.30
2847	F35PR500	800	F35PRS-BI-L(TO1)		18/04/85	19.35
2848	F35PR500	1200	F35PRS-BI-L(TO1)		18/04/85	17.31
2849	F35PR500	200	F35PRS-BI-L(TO1)		19/04/85	17.04
2850	F35PR500	200	F35PRS-BI-L(TO1)		19/04/85	18.38
2851	F35PR500	1200	F35PRS-BI-L(TO1)		19/04/85	11.35
2852	TASSO100	1	TASSO1-LIBRARY(370TERP)		22/04/85	11.17
2853	F35PR500	1200	F35PRS-BI-L(TO1)		22/04/85	08.14
2854	F35PR500	1200	F35PRS-BI-L(TO1)		25/04/85	11.17
2855	F35PR500	6	F35PRS-BI-L(TO1)		25/04/85	20.03
2856	F35PR500	1200	F35PRS-BI-L(TO1)		26/04/85	14.03
2857	F35PR500	1200	F35PRS-BI-L(TO1)		26/04/85	18.49
2858	F35PR500	1200	F35PRS-BI-L(TO1)		26/04/85	19.10
2859	F35PR500	1200	F35PRS-BI-L(TO1)		26/04/85	23.10
2860	F35PR500	1200	F35PRS-BI-L(TO1)		26/04/85	23.13
TYPE: CANCEL	OK		JCLJOB	OR		
			STOP	OR		
			EXIT			

Fig. 7 The job status shows which job has finished.

```

PROGRAM STATUS WORDS
CHANNEL ADDRESS AND CHANNEL STATUS WORDS
FIXED LOCATIONS FOR PROGRAM AND PROCESSOR SIZE
INTERUPT HANDLER
SIMULATE UNSUPPORTED OPERATION CODES
OPERATION SYSTEM FOR PROCESSOR WITH OLD INTERFACE
FIXED LOCATIONS FOR INTERRUPT, DATE, TIME, SIZE
IHOUAC TABLE FOR FORTRAN UNITS AND DCB'S
INPUT/OUTPUT ROUTINES, BUFFER HANDLER
TRACE LACK ROUTINES, ERROR HANDLING
DIRECT ACCESS INPUT/OUTPUT HANDLING
LAST ADDRESS: 8710 (HEX)

```

Fig. 8 The operating system contains the supervisor call handler, the program interrupt handler and part of the FORTRAN input/output routines.

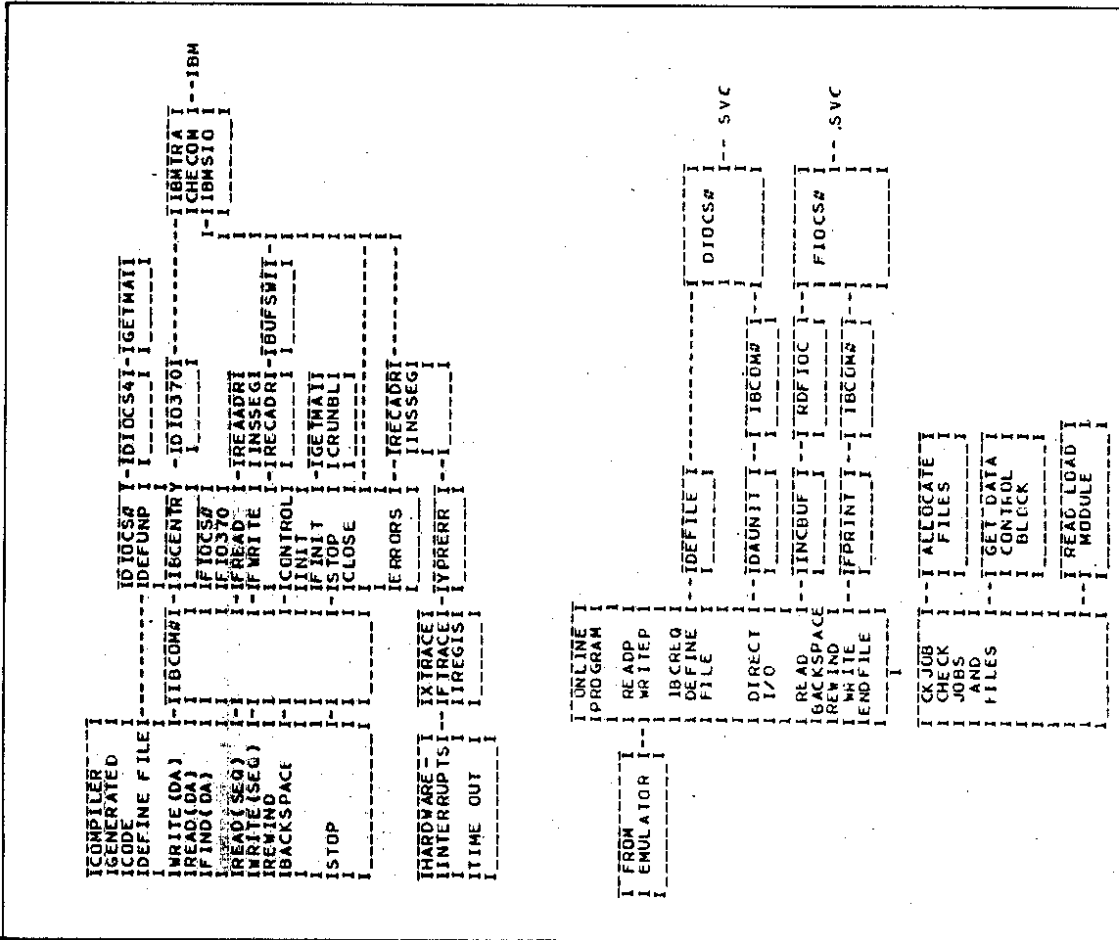


Fig. 9 Interface routines between user's program and IBM. The FORTRAN routines are written in such a way that the user does not need to change his program.

```

C 22/04/85 504291124 MEMBER NAME MAIN370 (TSOLIB)
CALL QUAD
CALL DACCES
CALL SEQUEN
CALL STA371
STOP
END
*OPTIONS IN EFFECT*NAME (MAIN) OPTIMIZE (2) LINECOUNT (60) SIZE (MAX) AUTODBL (NO)
C 12/02/85 504220948 MEMBER NAME QUAD (TSOLIB)
SUBROUTINE QUAD
EXAMPLE FOR REAL * 16
REAL*16 QA,QB,QC,QD
QA=2.
QB=4.
QC=CA*QB
QD=40%QATAN(100)
WRITE (6,2)QA,QB,QC,QD
FORMAT (17,1X,REAL * 16:*,4F20.10)
RETURN
END
*OPTIONS IN EFFECT*NAME (MAIN) OPTIMIZE (2) LINECOUNT (60) SIZE (MAX) AUTODBL (NO)
C 02/07/82 502409551 MEMBER NAME STA371 (TSOLIB)
SUBROUTINE STA371
EXAMPLES FOR SOME ERRORS
C-----
DIMENSION ADE(1),UTIME(4)
PRINT SIN AND COS
C
UNIT=6
WRITE (UNIT,1)
FORMAT (1X,5I4 COS TABLE)
PI=3.14159260,20
DO 2 I=1,20
RADE=I*PI
RAD=RAD/180.
SIN= SIN(RAD)
COS= COS(RAD)
WRITE (UNIT,4) I,SIN,COS,RAD
FORMAT (1X,16,3F10.4)
CONTINUE
C
A=3.
B=0.
C=A/B
WRITE (UNIT,6) A,B,C
FORMAT (1X,4F10.5)
CONTINUE
C
I=5
J=0
K=I/J
WRITE (UNIT,8) I,J,K
FORMAT (1X,4F10.5)
OVERFLOW
C
EVAL=ROV*E0V
WRITE (UNIT,10) E0V,E0VL
FORMAT (1X,2E15.7)
CONTINUE
C
RUN=1.E-60
EUNV=EUN*EUN
NEGATIVE SORT
C
AA=SQRT(-1.)
WRITE (UNIT,14) AA
FORMAT (1X,4F15.5)
NEGATIVE SORT, IF 15.5)
ADDRESS EXCEPTION
C
WRITE (UNIT,16)
FORMAT (1X,16)
ADDRESS VIOLATION*)
LI=2.300.000
AB=ADR(LI)
WRITE (UNIT,20)
FORMAT (1X,FINISH*)
WRONG UNIT
C
WRITE (0,12)
WRITE (UNIT,12)
FORMAT (1X,4F10.5)
RETURN
END
*OPTIONS IN EFFECT*NAME (MAIN) OPTIMIZE (2) LINECOUNT (60) SIZE (MAX) AUTODBL (NO)
C
SUBROUTINE DEFINE
COMMON/DEFILC/IV
DEFINE FILE 12(20,25,U,IV),13 (22,100,L,IU)
RETURN
END
*OPTIONS IN EFFECT*NAME (MAIN) OPTIMIZE (2) LINECOUNT (60) SIZE (MAX) AUTODBL (NO)

```

Fig. 10 This program gives an example how input/output and errors are handled by the 370/E.







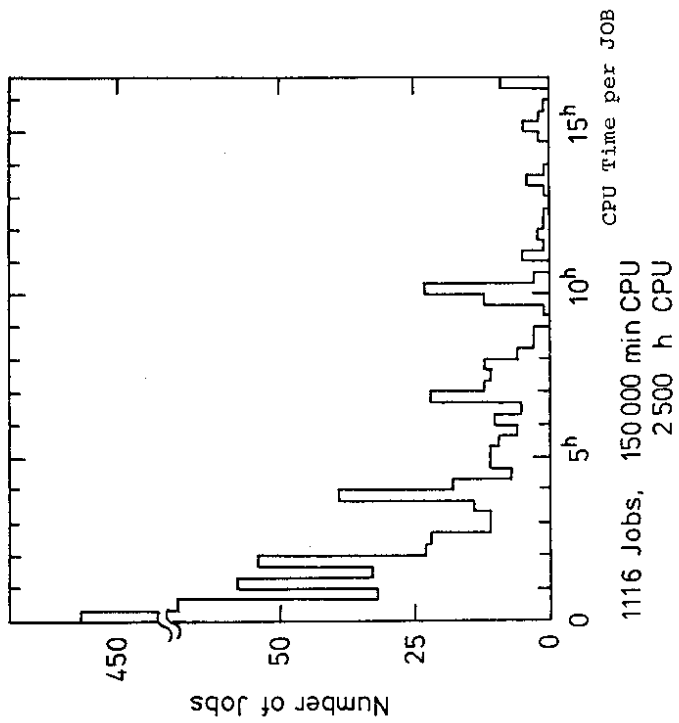


Fig. 11. Job profile of 1116 jobs executed during 1984.