Extracted Reliability Information from LA-6456-MS

LA-6456-MS is an evaluation of the CRAY-1 conducted by T.W. Keller in 1976 at Los Alamos Scientific Laboratory of the University of California (currently known as Los Alamos National Laboratory [LANL]).

This document extracts portions of that document related to reliability; other sections concerning performance have been removed. Editing was performed by Gary Grider (ggrider@lanl.gov) and John Bent (johnbent@lanl.gov), to whom any and all questions about this document may be directed.

If you have received this document in isolation, please be aware that it is one part of a larger effort to provide public access to operational data from LANL to support and enable computer science research. The homepage for this effort can be accessed at http://institute.lanl.gov/data.

LA-6456-MS
Informal Report

C.3

UC-32

Issued: December 1976

CRAY-1 Evaluation

Final Report

by

T. W. Keller

los alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87545

An Affirmative Action/Equal Opportunity Employer

UNITED STATES
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
CONTRACT W-7405-ENG. 36

Publication of this report does not imply endorsement of the products described herein.

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161
Price: Printed Copy $5.50 Microfiche $3.00

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

ACKNOWLEDGEMENTS

Much of Chapter III is drawn from reports by P. Iwanchuk and L. Rudsinski, who performed the scalar performance evaluation with the able assistance of T. Montoya and J. Miller. J. Moore performed the disk studies and much of Chapter VI is extracted from his reports. The discussion on Computation and I/O Interference in the same chapter closely follows a report by J. C. Browne (Visiting Staff Member). R. Tanner edited the manuscript with contributions from many readers and with the aid of the word processing section of group C-4. The vector performance studies (Chapter IV) were conducted by A. McKnight. Data reported in Chapter V was collected by N. Nagy and D. Buckner. T. Jordan, K. Witte, and A. Marusak performed the preliminary scalar performance test referenced in Chapter VIII. Many people reviewed the coding of the scalar performance modules (Chapter III), with special acknowledgement due D. Lindsay of FEDSIM. Discussions with B. Buzbee, J. Worlton, and F. Dorr helped define the scope of the evaluation under the time constraints involved.

(This page was not bound with the report as the result of an error on the author's part. He apologizes to all the above for this oversight.)

TABLE OF CONTENTS

SIGNATURES

Ronald S. Schwartz, Director
Office of ADP Management
Energy Research and Development Administration

Ronald Bartell

Ronald Bartell, Deputy Assistant Director for
Program Analysis and Budget
Division of Military Application
Energy Research and Development Administration

Fred W. Dorr

Fred W. Dorr, Division Leader
Computer Science and Services Division
Los Alamos Scientific Laboratory

SECTION I

EXECUTIVE SUMMARY

The performance evaluation of the CRAY-1 computer was structured to determine if the CRAY-1 meets the minimum performance standards set forth by the Los Alamos Scientific Laboratory (LASL) and the Energy Research and Development Administration (ERDA) to qualify the machine for further consideration for procurement.

The performance standards are divided into specific qualification criteria in three main areas: scalar performance, vector performance and reliability. The qualification criteria are summarized in Table I-1. The final Evaluation Plan, including precise definitions of the qualification criteria, is presented in Appendix A of this document.

It was impossible to convert large segments of the LASL computing workload to the CRAY-1 because programs to be run on the machine would require assembly language coding. Thus, for the scalar test, a sampling scheme was adopted that selected small computational kernels to be run on both the CDC 7600 and the CRAY-1. Kernels were drawn from a program by a method that weighted the probability of drawing a specific kernel by its contribution to the total execution time of the program. The sampling process was defined to a level of detail that eliminated the chances of biasing the selection toward either machine. By statistical methods it was possible to establish a test of the hypothesis that the CRAY-1 (in scalar mode) is greater than two times faster than the CDC 7600 with 90 percent confidence for any sampled program. To assure that the code kernels were representative of the potential LASL workload for the CRAY-1, the code kernels were drawn from the actual programs expected to comprise the eventual Class VI workload. Only programs consuming greater than one CDC 7600 hour per run and requiring more than one run per week were considered as potential workload candidates.

A second workload for the CRAY-1 was established from a sample that included frequently run codes not expected to be included in the immediate LASL workload for the machine. This was done as an attempt to establish a performance index of the machine based upon a more general workload, and one that might be more representative of computing throughout ERDA.

In order to eliminate the impact of compiler efficiency, kernels were coded as efficiently as was feasible in both CRAY-1 assembly language and CDC 7600 assembly language.

It was not possible to rigorously establish "representative" kernels for testing the vector performance of the CRAY-1. This was because none of the existing codes comprising the workload had been converted for vector operations, and such conversion efforts were outside the evaluation's time frame.

At the risk of oversimplifying, the vector computational speed of the CRAY-1 relative to the CDC 7600 is a function of both vector length and complexity of the vectorized arithmetic function. Relative performance of the CRAY-1 increases with vector length and complexity of operation. Performance criteria using vector lengths of 20, 100 and 500 were established, and the complexity of vector operations remained to be chosen. The simplest expressions, such as R=A+B, result in the lowest relative performance. Relative performance increases with greater numbers of different operators and operands (increasing complexity), as this allows increased overlap of functional units and chaining to occur. Chaining refers to an increase in parallelism resulting from the ability of the machine to store the result of a computation in a vector register while the result is re-entered as an operand to another vector computation in the same clock period. Thus, two or more primitive vector operations may be "chained" together. The more complex the evaluated expression, the greater the likelihood that chaining can occur. The Applications Support and Research Group of the Computer Science and Services Division at LASL was asked to furnish vector kernels that, in their judgement, represented common vector operations that user codes would perform on the CRAY-1. Five expressions of medium complexity, such as R=A*B+C, were chosen to evaluate the machine's vector performance as a function of vector length. The average of the five performance ratios was chosen to compare against the qualification criteria. Each vector kernel was coded as efficiently as was feasible in assembly language for each machine. In particular, the CDC 7600 kernels were coded as "in-stack" loops--a scheme that pushes the CDC 7600 toward its theoretical maximum performance for these algorithms.

The reliability of the CRAY-1 was evaluated by establishing a reliability test code to run on the machine for long periods of time. This "exerciser" was designed to utilize as many different hardware units of the machine as possible, at as rapid a rate as possible, for extended periods. The exerciser underwent several evolutionary stages toward this goal. The latest version of the exerciser accesses the machine's memory at a sustained rate several times greater than that plausible for a production workload. The reliability figures were determined for contiguous 20-workday periods in order to "smooth" short-term fluctuations in reliability.

In addition to the tests against the qualification criteria outlined above, the evaluation also investigated the CRAY-1 disk system performance and made other miscellaneous studies. For sake of brevity these studies will not be discussed in the Executive Summary.

Clearly, many aspects of the CRAY-1's performance were evaluated, some in considerable detail. An impartial study was accomplished despite the constraints of a primitive CRAY-1 operating system, the absence of a Fortran compiler, the relatively short evaluation period, and the necessity of agreement by LASL, ERDA, and the Federal Computer Performance Evaluation and Simulation Center (FEDSIM) upon an evaluation plan. Rigorously defensible results were obtained by adopting methods of known accuracy wherever possible. Although the constraints confined the scope of the evaluation, they did not hinder the objectivity nor accuracy of its results.

Results

A brief summary of the evaluation results is presented in Table I-1.

Scalar. Timing results for the scalar kernels are summarized in Table III-6 of Section III. On the basis of these results the hypothesis that the CRAY-1 is at least two times faster than the CDC 7600 for scalar kernels was satisfied in all tests.

Vector. Results of the vector kernel timings are summarized in Table I-1. The machine met the vector performance qualification criteria for the three vector lengths.

Reliability. The machine met the reliability criteria for many reported 20-day periods. The Mean-Time-To-Failure (MTTF) fluctuated from a low of approximately 2.5 hours to a high of approximately 7.5 hours during the six-month period. No trend was observed.

Approximately 89 percent of all machine failures during the evaluation were memory parity errors. If one assumes that all memory errors were correctable single-bit errors, then installation of single-bit memory error correction would have resulted in an increase in MTTF by a factor of nine. Such an increase would result in extremely good reliability for a machine of this complexity.

The conclusion of the evaluation is that the CRAY-1 satisfies the threshold performance criteria in all categories. FEDSIM, in a separate report to be issued, concurs in this conclusion.

Extracted Reliability Information from LA-6456-MS

This page is a place-holder indicating that pages
have been removed from the original document here.

SECTION V

RELIABILITY

Approach

It was considered crucial that any Class VI computer considered for purchase be very reliable. Extended periods of downtime would be intolerable since the programmatic functions served by the computer could not be absorbed by other machines at the Laboratory. Thus rigorous reliability standards were formulated. In addition to "system availability," the single measure that is commonly used to define reliability, two additional reliability criteria were specified. Threshold reliability criteria of at least 80 percent system availability, at least four hours Mean-Time-to-Failure (MTTF) and at most one hour Mean-Time-to-Repair (MTTR) were established by LASL and ERDA as defining an acceptable level of reliability. The reader is referred to the Evaluation Plan (Appendix A) for a precise definition of these measures.

The reliability criteria would have to be met for a contiguous twenty-workday period for the machine to be considered for further procurement. The twenty-day period was established in order to smooth daily fluctuations expected for the measures; the period was considered long enough to prevent a machine meeting the criteria during a "fluke" period of good behavior. In order not to unfairly penalize newly constructed machines, the measures from the best twenty-workday period would be applied against the criteria.

Commonly adopted logging procedures for determining machine reliability were deemed inappropriate due to the special evaluation environment of the CRAY-1. This conclusion is the result of the two observations that:

1. The burden of logging machine reliability falls upon a large number of operators and programmers, and is vulnerable to human error; and

2. The complex environment and primitive operating system of the CRAY-1 make it difficult to isolate CRAY-1 hardware failures from a) operator errors, b) ECLIPSE hardware/software failures, and c) CRAY-1 benchmark operating system software errors (unless a hardware error

interrupt occurs and is handled correctly by the system).

The evaluation environment of the CRAY-1 resulted in the CPU being idle the greatest fraction of time. The mode of operation for running on the machine typically consists of a programmer performing nearly all tasks on the ECLIPSE and running a program on the CRAY-1 for only brief intervals. The CRAY-1 benchmark operating system presently does not have implemented the capability of running a "background" job to keep the machine busy.

The EXERCISER Program

The EXERCISER program was written in an attempt to overcome these limitations. The objective of the program is to approximate a production environment on the CRAY-1 by utilizing as many hardware features of the machine as possible for substantial periods of time. The EXERCISER program is self-verifying so that all machine failures, detected at the hardware level or not, will be noted. The program keeps a printed log of the time of each failure. Reliability statistics may be gathered from the printed log, thus minimizing the possibility of human error.

Hardware components specifically being tested are the vector and scalar functional units of the CPU, the memory, and input/output (I/O) components. Unfortunately, I/O was not available on the machine during the first two months of the evaluation.

EXERCISER was adapted from one of the first programs written for the CRAY-1 at LASL. Coded in CAL with PASCAL drivers, it solves for the vector x the matrix equation Ax=y by LU decomposition.

For each NXN matrix, N+1 systems of equations are solved. The A and y are parameterized such that every element of x in the N+1th system is near unity. Before N is incremented, every element of x is tested against unity. If a "near" unity test fails for any component of x, a message is dispatched to the operator and the event is logged by the program to a print file. The program then restarts the cycle until it is terminated by the operator. Most failures, such as parity errors, cause termination of the program and a message is dispatched to the operator by the operating system.

After June 15 substantial changes were made in the program in order for it to more fully meet its objective. First, N was fixed at 705, so as to "exercise" almost all of available memory. The

EXERCISER code and data require approximately 501 000 words.

The second substantial change was the addition of I/O to a DD-19 disk. The control flow of the EXERCISER program remains unchanged with the following exception. At the completion of solving the Nth system, the LU matrix is written to disk and then read from disk. The N+1th system is then solved and the test against unity is made. Any disk errors will result in this test failing. As before, upon detection of a failure, the program logs the failure and transmits a message to the operator.

EXERCISER was run from five to eight hours per day, five days a week throughout the entire six-month period. A sample of the daily log is shown in Figure V-1.

Results and Conclusions

Results, by twenty-workday interval, calculated weekly, are displayed in Table V-1. The MTTF measures are also displayed graphically in Figure V-2. Failures are categorized in Table V-2. One sees that the CRAY-1 meets the threshold criteria for numerous periods.

An analysis of failures occuring during EXERCISER runs revealed that "intermittent" memory parity errors dominate. Hardware detection of a memory parity error prompts an interrupt which idles the machine and saves the program counter. EXERCISER is so constructed that from the program counter value and a memory dump, the exact memory bit causing the failure can be determined. According to Cray Research, Inc. (CRI), roughly 60 percent of the intermittent failures were diagnosed to this level of detail, of which all were caused by failure of a single bit. In all but 12 of the memory parity error failures the memory module incorporating that bit could not be made to fail again, neither during EXERCISER nor during various CRI memory diagnostic routines. In the 12 cases of reproducible failure, the memory module was judged defective and replaced. In the remaining cases, tests of the modules by CRI could determine no differences in operating characteristics between "once failing" and "never failing" modules. Once reinstalled in the machine, no module failed again.

CRI made various hardware changes in an attempt to eliminate the failure mode or cause it to replicate at a more rapid rate. None of the hardware changes appeared to affect the failure rate. An obstacle to diagnosing this problem was the relatively long average interval between failures (4 hours), which implied very long running times during a change to determine if the failure rate had been changed.

Since the "once failing" module was seldom replaced, the MTTR measure should be more accurately labeled "mean-time-to-return," since hardware repairs were infrequent. In order to obtain a more accurate estimate of repair times, the mean-time-to-repair for all EXERCISER failures during which a repair was effected was calculated. For these 20 failures the total down-time was 8.37 hours, resulting in a MTTR of 0.41 hours. From this measure we conclude that only a minor portion of repair time was consumed by the actual hardware change, the major portion being consumed by alerting the engineer and running diagnostic routines.

Since memory parity errors so clearly dominate the statistics, an analysis of EXERCISER was made in an effort to relate its memory access rate to that of a production environment. Memory utilization by EXERCISER is divided into two phases. In the I/O phase, 497 025 memory locations are accessed twice during a 3.10-second period. This results in a memory access rate of approximately 0.0062 times the theoretical maximum of 80 million accesses per second (80 MAPS). In the computation phase of the program, a 705 by 705 matrix is initialized and decomposed and a system of 705 linear equations is solved. This phase lasts 14.11 seconds. The use of vector instructions in the inner computational loop results in 64 memory accesses every 78 clock cycles. The inner loop consumes roughly 90% of the time for this phase. Thus memory is driven at approximately 0.9 x 64/78 = .74 of its maximum of 80 MAPS. Ignoring the memory accesses during the I/O phase, EXERCISER drives the CRAY-1 memory at .74 of its capacity for 14.11/(14.11 + 3.10) = .82 of the program's execution, for an overall memory access rate that is 0.61 of its maximum.

Predicting a typical memory access rate for a production environment with the above accuracy is not possible. However, it is plausible that the EXERCISER bandwidth is at least several times higher than that expected for a production workload. Thus we would expect the MTTF for a production workload on the CRAY-1 to be significantly higher than that observed under EXERCISER.

Figure V-1. Daily EXERCISER Log.

EXERCISER LOG

EXERCISER VERSION:

SYSTEM VERSION:

COMMENTS:

DATE:

EXERCISER BEGIN TIME:

EXERCISER END TIME:

OPERATOR'S INITIALS:

CE'S INITIALS:

EXERCISER INITIATED	EXERCISER DETECTED ERROR	SYSTEM DETECTED ERROR	EXERCISER TURNED OFF	SYSTEM CRASHED	CE REPAIR INITIATED	CE RETURNED MACHINE	CE REPLACED HARDWARE MODULE				

SUMMARY

Time between EXERCISER start and failure/between failures (minutes).

1.______ 2.______ 3.______ 4.______ 5.______

6.______ 7.______ 8.______ 9.______ 10.______

Time CE has machine for repair (minutes)

1.______ 2.______ 3.______ 4.______

5.______ 6.______ 7.______ 8.______

Time from last failure to shut off

1.______

No time intervals

1.______

Figure V-2. MTTF urapn.

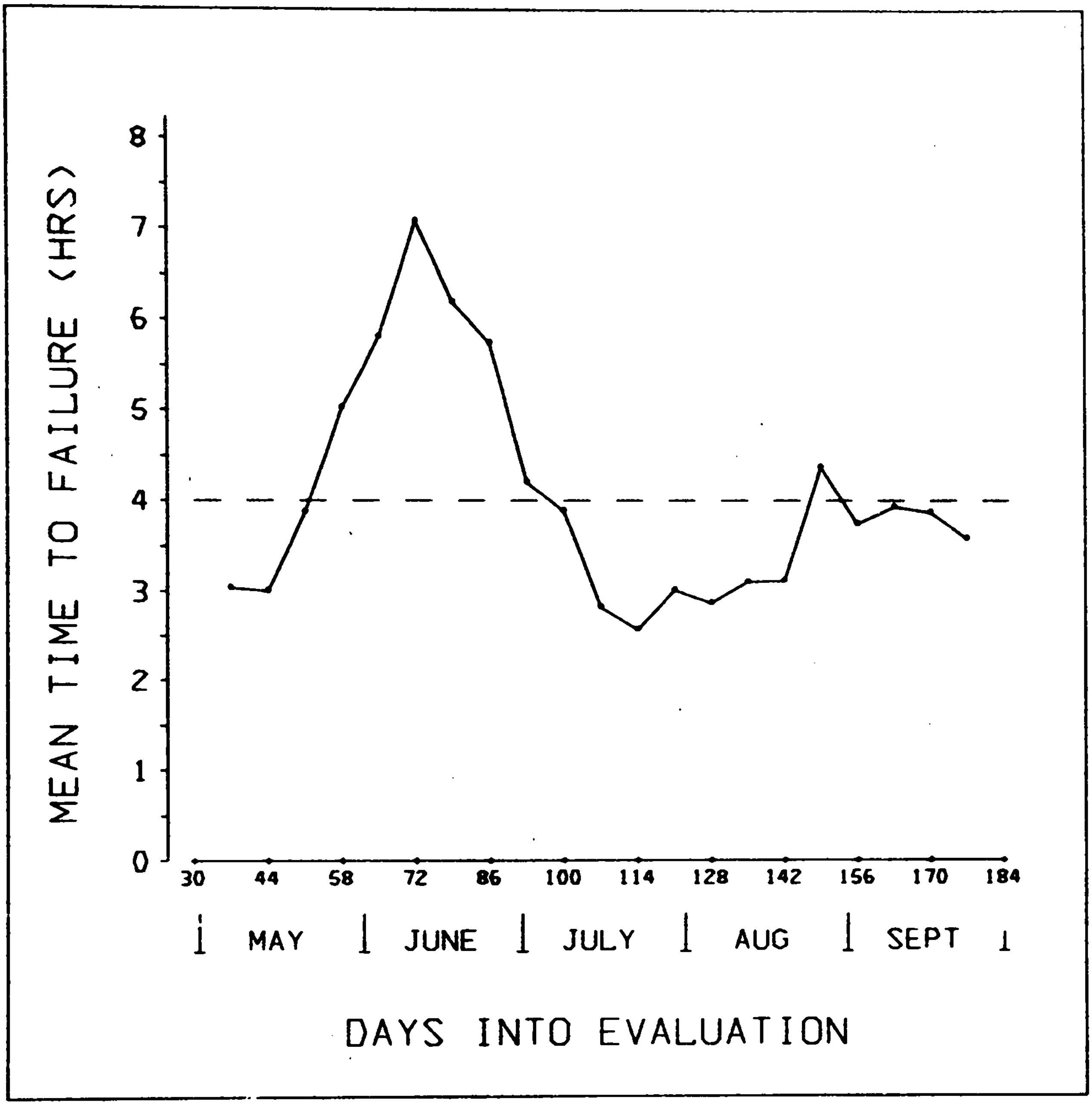


Table V-1.

Reliability Statistics Summarized by Week.

Twenty-Workday Period	Operational Use time (Hrs.)	Remedial Maintenance Time (Hrs.)	No. of Failures	MTTF (Hrs.)	MTTR (Hrs.)	SA
4/5 -5/7	115.15	17.72	40	2.88	.44	.87
4/14-5/14	113.50	16.35	38	2.99	.43	.87
4/23-5/21	112.05	11.83	29	3.86	.41	.90
4/30-5/28	115.37	9.25	23	5.02	.40	.93
5/6 -6/4	110.13	7.08	19	5.80	.37	.94
5/14-6/11	106.25	5.65	15	7.08	.38	.95
5/21-6/18	104.98	6.33	17	6.18	.37	.94
5/28-6/25	109.02	7.07	19	5.74	.37	.94
6/8 -7/2	100.78	10.40	24	4.20	.43	.91
6/15-7/9	97.12	9.55	25	3.88	.38	.91
6/22-7/16	87.42	16.67	31	2.82	.54	.84
6/29-7/23	88.57	18.12	35	2.53	.52	.83
7/6 -7/30	99.45	15.67	33	3.01	.47	.86
7/13-8/6	97.55	17.68	34	2.87	.52	.85
7/19-8/13	108.45	13.25	35	3.10	.38	.89
7/27-8/20	102.77	12.47	33	3.11	.38	.89
8/4 -8/27	104.60	8.77	24	4.36	.37	.92
8/10-9/3	104.73	10.85	28	3.74	.39	.91
8/12-9/10	105.85	9.82	27	3.92	.36	.92
8/17-9/17	108.45	10.02	28	3.87	.36	.92
8/23-9/24	114.25	10.43	31	3.69	.34	.92

Table V-2.

Failures Classified by Type.

Memory parity errors*	152
Disk	1
Vector modules	12
Instruction buffer module	1
Floating add module	4
Total	170

*Of the 152 memory parity errors, 12 were found to be reproducible.

NOTE: Of the 18 non-memory related failures, 8 repairs were effected at the time of failure.

If the cause of the intermittent memory failures cannot be diagnosed and eliminated, the advantages of a machine with memory error correction are obvious. If one assumes all intermittent memory failures (139) could have been avoided by a single bit correction technique, then the MTTF for such a machine over the entire evaluation period would have increased by a factor of 170/(170-139) = 5.5.

This estimate assumes that a reproducible memory parity error, of which 12 were observed, would result in the machine failing. CRI reported that all 12 memory modules replaced were single bit failures. If one assumes that error correction would allow deferring module replacement into the scheduled maintenance period, then it is possible that single bit correction would have resulted in a MTTF increase by a factor of 170/(170-152) = 9.4 over the entire evaluation period.

Extracted Reliability Information from LA-6456-MS

This page is a place-holder indicating that pages
have been removed from the original document here.

SECTION VIII

CONCLUSIONS

The purpose of this evaluation was to determine if the CRAY-1 computer meets the minimum performance standards set forth by LASL and ERDA to qualify the machine for further consideration for procurement. These standards are divided into specific qualification criteria in three main areas: scalar performance, vector performance and reliability.

Scalar Performance

The hypothesis that the CRAY-1 in scalar mode is at least two times faster than the CDC 7600 was tested on samples drawn from each of the three codes comprising the Class VI applications workload, and from a sample drawn equally from the five codes comprising the Class VI workload. The hypothesis test was structured so that the probability of a wrong result is less than 0.1. The hypothesis tested as true in all cases, with the minimum number of kernels necessary to test the hypothesis. Thus the CRAY-1 meets both the Class VI workload scalar performance criterion and the Class VI applications workload scalar performance criterion.

Kernel timings also provided estimates of the CDC 7600/CRAY-1 execution time ratios (speed ratios) for scalar computation. The speed ratios for all four workload samples ranged from about 2.5 to greater than 2.8. In addition, the speed ratio for the preliminary scalar test code was 2.5. From these results we conclude that the CRAY-1 in scalar mode has the potential for executing CPU-bound codes 2.5 times faster than the CDC 7600.

Vector Performance

Five operations were chosen to evaluate the CRAY-1's vector performance versus vector length. Each operation was coded as efficiently as was feasible for both the CDC 7600 and the CRAY-1. Each operation yielded a CDC 7600/CRAY-1 speed ratio for three vector lengths. The average of the five speed ratios was compared against the qualification criterion for each vector length. The average speed ratios were 3.39, 4.50 and 5.12 for vector lengths of 20, 100 and 500, respectively. Thus, speed ratio criteria of 3, 4 and 5 for vector lengths of 20, 100 and 500, respectively, were met.

Reliability

Reliability of the CRAY-1 was evaluated by running a reliability test code on the machine for long periods of time. The EXERCISER program was designed to access as many different hardware units of the machine as possible, at a rapid rate, for extended periods. The test program is characterized by access rates significantly higher than those of an initial production workload, and above those of longer term workloads. Mean-time-to-failure ranged from 2.53 hrs. to 7.08 hrs. for the reported periods. The criterion of at least four hours was exceeded for 7 of the 21 overlapping periods. Mean-time-to-repair ranged from 0.34 hrs. to 0.52 hrs., exceeding the criterion of no more than 1 hour for all reported periods. Likewise, system availability ranged from 0.85 to 0.95, exceeding the criterion of at least 0.80 for all reported periods. Reliability was good for a serial number one machine of this size and speed.

An analysis of machine failures revealed that memory parity errors dominated the statistics. If one assumes that single-bit memory error correction would have eliminated all memory failures, then a CRAY-1 with this feature would have resulted in a mean-time-to-failure measure approximately nine times greater than that observed over the entire six-month evaluation period, with a less dramatic increase in system availability. The indications are that the CRAY-1 with error correcting memory would be an exceptionally reliable machine.

Input/ Output Studies

Although no qualification criteria were established in the area of input/output (I/O) operations, a study of the CkAY-1 I/O subsystem was undertaken to determine if any pathologies existed. Performance tests uncovered the fact that disk revolutions were being missed during disk writes, due primarily to limitations in the interim disk controller. These limitations should not exist for the product-line controller. With this exception, expected transfer rates for the disk were observed. Head positioning times were close to expected values. An error detection test wrote and read over four billion words to disk with no errors.

The degradation to vector computation due to I/O interference by memory cycle stealing and I/O interrupts was measured to determine if this might pose a serious problem to CRAY-1 vector performance. Execution time degradations to vector computation in a worst case test with the single disk yielded degradations of 3.5 percent and 3.7 percent, which were assigned to memory cycle

Extracted Reliability Information from LA-6456-MS

This page is a place-holder indicating that pages
have been removed from the original document here.

APPENDIX A

The computer configuration being evaluated includes the following units.

- The CRAY-1 central processing unit, with 524 288 words (64 bits plus one parity bit) of memory arranged in 16 banks;
- One disk control unit (DCU);
- Two 819 disk units;
- Twelve independent channels (asynchronous, full duplex);
- One Data General Eclipse station, with the following input-output units:
    - One TEC 455 display,
    - One TEC 1440 display,
    - One Gould 5000 printer,
    - One Documation M1000 card reader,
    - One Century Data disk, and
    - One Data General nine-track tape.

THE CRAY - 1 COMPUTER

The Cray Research, Inc. CRAY-1 Computer System is a large-scale, general-purpose digital computer featuring vector as well as scalar processing, a 12.5 nanosecond clock period, and a 50 nanosecond memory cycle time. The CRAY-1 is capable of executing over 80 million floating point operations per second. Even higher rates are possible with programs that take advantage of the vector features of the computer.

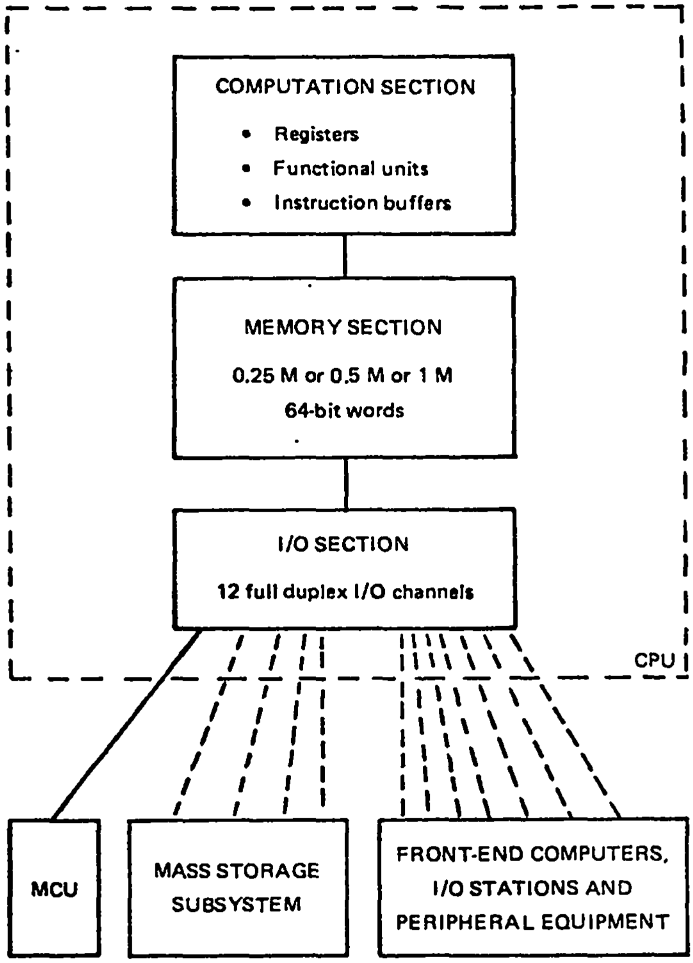
The CRAY-1 is particularly adapted to the needs of the scientific community and is especially useful in solving problems requiring the analysis and prediction of the behavior of physical phenomena through computer simulation. The fields of weather forecasting, aircraft design, nuclear research, geophysical research, and seismic analysis involve this process. For example, the movements of global air masses for weather forecasting, air flows over wing and airframe surfaces for aircraft design, and the movements of particles for nuclear research, all lend themselves to such simulations. In each scientific field, the equations are known but the solutions require extensive computations involving large quantities of data. The quality of a solution depends heavily on the number of data points that can be considered and the number of computations that can be performed. The CRAY-1 provides substantial increases with respect to both the number of data points and computations so that researchers can apply the CRAY-1 to problems not feasibly solvable in the past.

CONFIGURATION

The basic configuration of the CRAY-1 consists of the central processor unit (CPU), power and cooling equipment, one or more minicomputer consoles, and a mass storage (disk) subsystem. The CPU holds the computation, memory, and I/O sections of the computer. A minicomputer serves either as a maintenance control unit or a job entry station.

INPUT/OUTPUT

Input/output is via twenty-four I/O channels, twelve of which are input and twelve output. Any number of channels may be active at a given time. The channel transfer rate is based on the channel width (currently 8 or 16 bits). For a 16 bit channel, maximum rates of 160 million bits per second are attainable. Higher rates are possible with wider channels. In practice, this theoretical transfer rate is limited by the speed of peripheral devices and by memory reference activity of the CPU.



BASIC COMPUTER SYSTEM

MEMORY

The CRAY-1 memory is constructed of 1024-bit LSI chips. Up to 1,048,576 (generally referred to as one million) 64-bit words are arranged in 16 banks. The bank cycle time, that is, the time required to remove or insert an element of data in memory, is 50 nanoseconds. This short cycle time provides an extremely efficient random-access memory. One parity bit per word is maintained in 16 modules of the central processor. There is no inherent memory degradation for machines with less than one million words of memory.

FACTS AND FIGURES

CPU	
Instruction size	16 or 32 bits
Clock period	12.5 nsec
Instruction stack/buffers	64 words (4096 bits)
Functional units	twelve: 3 integer add 1 integer multiply 2 shift 2 logical 1 floating add 1 floating multiply 1 reciprocal approx. 1 population count
Programmable registers	8x64 64-bit 73 64-bit 72 24-bit 1 7-bit
Max. vector result rate	12.5 nsec / unit
FLOATING POINT COMPUTATION RATES (results per second)	
Addition	80×10^6 / sec
Multiplication	80×10^6 / sec
Division	25×10^6 / sec
MEMORY	
Technology	bipolar semiconductor
Word length	64 bits
Address space	4M words
Data path width (bits)	64 (1 word)
Cycle time	50 nsec.
Size	262,144 words or 524,288 words or 1,048,576 words
Organization / interleave	16 banks
Maximum band width	80×10^6 words / sec (5.1×10^9 bits / sec)
Error checking	1 parity bit / word
PHYSICAL CHARACTERISTICS / ELECTRONIC TECHNOLOGY	
Size of CPU cabinet	9 ft diameter base 4.5 ft diameter center 6 ft height
Weight of mainframe	5 tons
Cooling	Freon
Plug-in modules	1506
Module types	109
PC boards	5 layer
Circuitry (equivalent no. of transistors)	2.5M
Logic	ECL, 1 nsec.
High-density logic	SSI

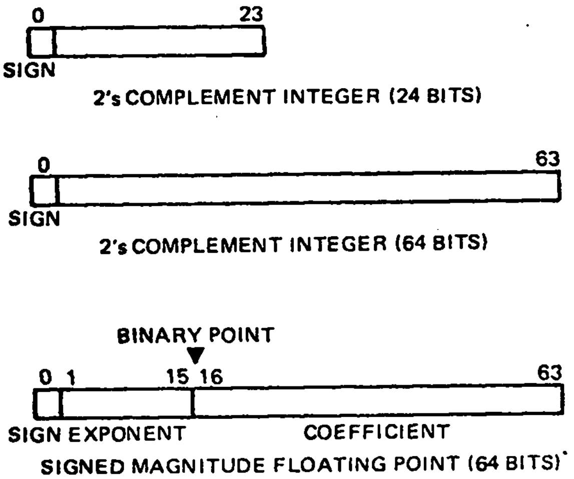
COMPUTATION SECTION

The computation section as illustrated on page 4 is composed of instruction buffers, registers, and functional units which operate together to execute sequences of instructions.

Data structure

Internal character representation in the CRAY-1 is in ASCII with each 64-bit word able to accommodate eight characters.

Numeric representation is either in two's complement form (24-bit or 64-bit) or in 64-bit floating point form using a signed magnitude binary coefficient and a biased exponent. Exponent overflow and underflow is caused if the exponent is greater than 57777_8 or less than 20000_8. For scalar operations, either of these conditions causes an interrupt except where the interrupt has been inhibited. For vector operations, these conditions do not cause an interrupt.


0
23
SIGN
2's COMPLEMENT INTEGER (24 BITS)
0
63
SIGN
2's COMPLEMENT INTEGER (64 BITS)
BINARY POINT
0 1
15 16
63
SIGN EXPONENT
COEFFICIENT
SIGNED MAGNITUDE FLOATING POINT (64 BITS)

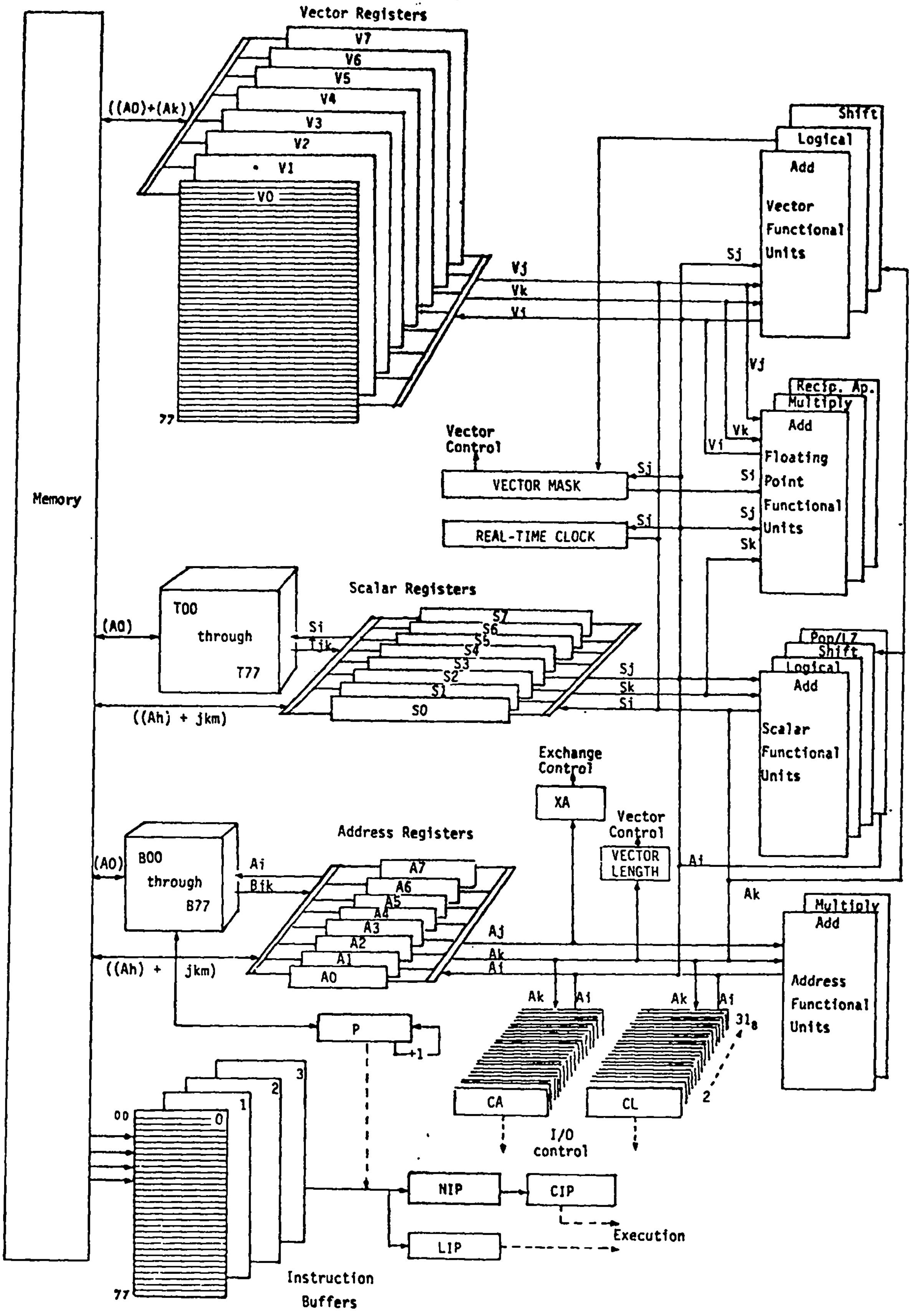
DATA FORMATS

Instruction set

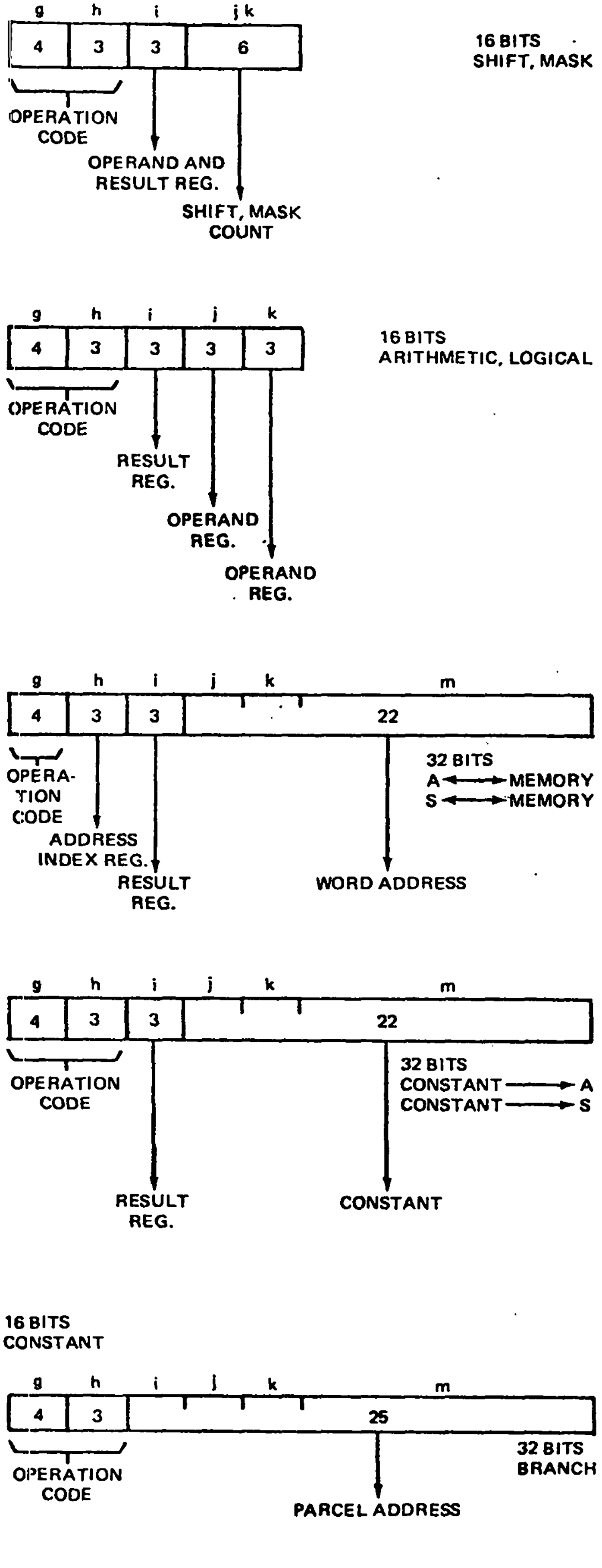
The CRAY-1 executes 128 operation codes as either, 16-bit (one parcel) or 32-bit (two-parcel) instructions. Operation codes provide for both scalar and vector processing.

In general, an instruction that references registers occupies one parcel; an instruction that references memory occupies two parcels. All of the arithmetic and logical instructions reference registers.

Floating point instructions provide for addition, subtraction, multiplication, and reciprocal approximation. The reciprocal approximation instruction allows for the computation of a floating point divide operation using a multiple instruction sequence.



COMPUTATION SECTION



INSTRUCTION FORMATS

Integer or fixed point operations are provided for as follows: integer addition, integer subtraction, and integer multiplication. An integer multiply operation produces a 24-bit result; additions and subtractions produce either 24-bit or 64-bit results. No integer divide instruction is provided. The operation can be accomplished through a software algorithm using floating point hardware.

The instruction set includes Boolean operations for OR, AND, and exclusive OR and for a mask-controlled merge operation. Shift operations allow the manipulation of 64- or 128-bit operands to produce a 64-bit result. Similar 64-bit arithmetic capability is provided for both scalar and vector processing. Full indexing capability allows the programmer to index throughout memory in either scalar or vector modes of processing. This allows matrix operations in vector mode to be performed on rows, on columns, or on the diagonal.

Addressing

Instructions that reference data do so on a word basis. Instructions that alter the sequence of instructions being executed, that is, the branch instructions, reference parcels of words. In this case, the lower two bits of an address identify the location of an instruction parcel in a word.

Instruction buffers

All instructions are executed from four instruction buffers, each consisting of 64 16-bit registers. Associated with each instruction buffer is a base address register that is used to determine if the current instruction resides in a buffer. Since the four instruction buffers are large, substantial program segments can reside in them. Forward and backward branching within the buffers is possible and the program segments may be noncontiguous. When the current instruction does not reside in a buffer, one of the instruction buffers is filled from memory. Four memory words are transferred per clock period. The buffer that is filled is the one least recently filled, that is, the buffers are filled in rotation. To allow the current instruction to issue as soon as possible, the memory word containing the current instruction is among the first four transferred. A parcel counter register (P) points to the next parcel to exit from the buffers. Prior to issue, instruction parcels may be held in the next instruction parcel (NIP), lower instruction parcel (LIP) and current instruction parcel (CIP) registers.

Operating registers

The CRAY-1 has five sets of registers, three primary and two intermediate. Primary registers can be accessed directly by functional units. Intermediate registers are not accessible by functional units but act as buffers between primary registers and memory.

The figure on page 4 represents the CRAY-1 registers and functional units. The 64 address and 64 scalar intermediate registers can be filled by block transfers from memory. Their purpose is to reduce memory references made by the scalar and address registers.

The eight address registers are each 24 bits and can be used to count loops, provide shift counts, and act as index registets in addition to their main use for memory references.

The eight 64-bit scalar registers in addition to contributing operands and receiving results for scalar operations can provide one operand for vector operations.

Each of the eight vector (V) registers is actually a set of 64 64-bit registers, called elements. The number of vector operations to be performed (that is, the vector length) determines how many of the elements of a register are used to supply operands in a vector set or receive results of the vector operation. The hardware accommodates vectors with lengths up to 64; longer vectors are handled by the software dividing the vector into 64-element segments and a remainder.

Associated with the vector registers are a 7-bit vector length register and a 64-bit vector mask register. The vector length register, as its name implies, determines the number of operations performed by a vector instruction. Each bit of the vector mask register corresponds to an element of a V register. The mask is used with vector merge and test instructions to allow operations to be performed on individual vector elements.

Supporting registers

In addition to the operating registers, the CPU contains a variety of auxiliary and control registers. For example, there is a channel address (CA) register and a channel limit register (CL) for each I/O channel.

Functional units

Instructions other than simple transmits or control operations are performed by hardware organizations known as functional units. Each of the twelve units in the CRAY-1 executes an algorithm or a portion of the instruction set. Units are independent. A number of functional units can be in operation

A functional unit receives operands from registers and delivers the result to a register when the function has been performed. The units operate essentially in three-address mode with source and destination addressing limited to register designators.

All functional units perform their algorithms in a fixed amount of time. No delays are possible once the operands have been delivered to the unit. The amount of time required from delivery of the operands to the unit to the completion of the calculation is termed the "functional unit time" and is measured in 12.5 nsec clock periods.

The functional units are all fully segmented. This means that a new set of operands for unrelated computation may enter a functional unit each clock period even though the functional unit time may be more than one clock period. This segmentation is made possible by capturing and holding the information arriving at the unit or moving within the unit at the end of every clock period.

The twelve functional units can be arbitrarily assigned to four groups: address, scalar, vector, and floating point. The first three groups each acts in conjunction with one of the three primary register types, to support address, scalar, and vector modes of processing. The fourth group, floating point, can support either scalar or vector operations and will accept operands from or deliver results to scalar or vector registers accordingly.

FUNCTIONAL UNITS

Functional Unit	Unit Time (Clock Periods)	Instructions
Address integer add	2	030, 031
Address multiply	6	032
Scalar integer add	3	060, 061
Scalar logical	1	042 - 051
Scalar shift	2	052 - 055
	3	056, 057
Scalar leading zero/pop count	4	026
	3	027
Vector integer add	3	154 - 157
Vector logical	2	140 - 147, 175
Vector shift	4	150 - 153
Floating point add	6	062, 063, 170 - 173
Floating point multiply	7	060 - 067, 160 - 167
Floating point reciprocal	14	070, 174

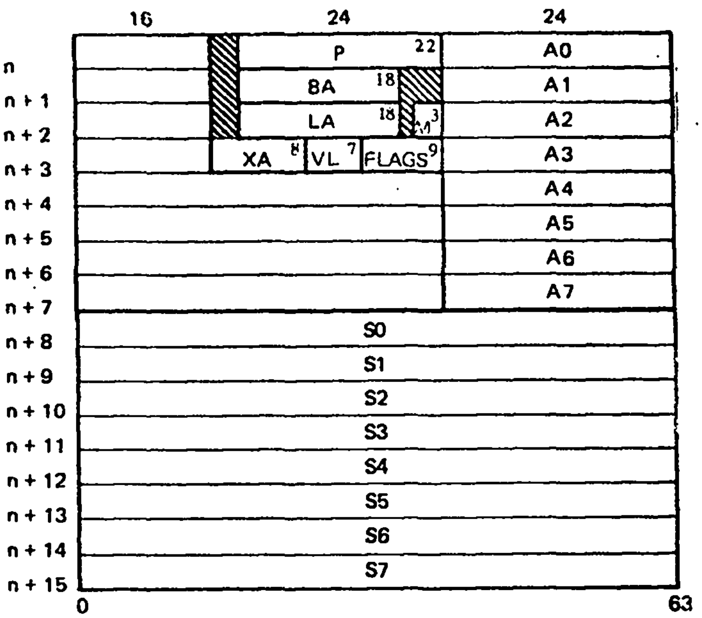
Memory field protection

Each object program has a designated field of memory. Field limits are defined by a base address register and a limit address register. Any attempt to reference instructions or data beyond these limits results in a range error.

Exchange mechanism

The technique employed in the CRAY-1 to switch execution from one program to another is termed the exchange mechanism. A 16-word block of program parameters is maintained for each program. When another program is to begin execution, an operation known as an exchange sequence is initiated. This sequence causes the program parameters for the next program to be executed to be exchanged with the information in the operating registers to be saved. The operating register contents are thus saved for the terminating program and entered with data for the new program.

Exchange sequences may be initiated automatically upon occurrence of an interrupt condition or may be voluntarily initiated by the user or by the operating system through normal and error exit instructions.



	Flags*
31	Console Interrupt
32	RTC Interrupt
33	Floating Point Error (Scalar Reference Only)
34	Operand Range
35	Program Range
36	Storage Parity
37	I/O Interrupt
38	Error Exit
39	Normal Exit

	Modes*
37	Interrupt on Floating Point
38	Interrupt on Storage Parity
39	Monitor Mode

P = Program Address
BA = Base Address
LA = Limit Address
XA = Exchange Address
VL = Vector Length

* Bit position from left of word

EXCHANGE PACKAGE

ARCHITECTURE

Construction

The CRAY-1 is modularly constructed of 1506 modules held by 24 chassis. Each module contains two 6 in. by 8 in. printed circuit boards on which are mounted a maximum of 144 integrated circuit packages per board. Emitter coupled logic (ECL) is used throughout. Four basic chip types are used: a high-speed 5/4 NAND gate, a slow-speed 5/4 NAND gate, a 16x1 register chip, and a 1024x1 memory chip.

Appearance

The esthetics of the machine have not been neglected. The CPU is attractively housed in a cylindrical cabinet. The chassis are arranged two per each of the twelve wedge-shaped columns. At the base are the twelve power supplies. The power supply cabinets, which extend outward from the base are vinyl padded to provide seating for computer personnel.

The compact mainframe occupies a mere 70 sq. ft. of floor space.

Cooling

The speed of the CPU is derived largely by keeping wire lengths extremely short in the mainframe. This, in turn, necessitates a dense concentration of components with an accompanying problem of heat dissipation. The Freon cooling system used in the CRAY-1 employs the latest in refrigeration technology to maintain a column temperature of about 68° in the unit.

MAINTENANCE CONTROL UNIT (MCU)

A 16-bit minicomputer system serves as a maintenance control unit. The MCU performs system initialization and basic recovery for the operating system. Included in the MCU system is a software package that enables the minicomputer to monitor CRAY-1 performance during production hours.

STATIONS

The CRAY-1 computer system may be equipped with one or more 16-bit minicomputer systems that provide input data to the CRAY-1 and receive output from the CRAY-1 for distribution to a variety of slow-speed peripheral equipment. A station consists of a Data General S-200 minicomputer or equivalent. Peripherals attached to the station vary depending on whether the station is a local or remote job entry station or a data concentrator used for multiplexing several remote stations.

EXTERNAL INTERFACE

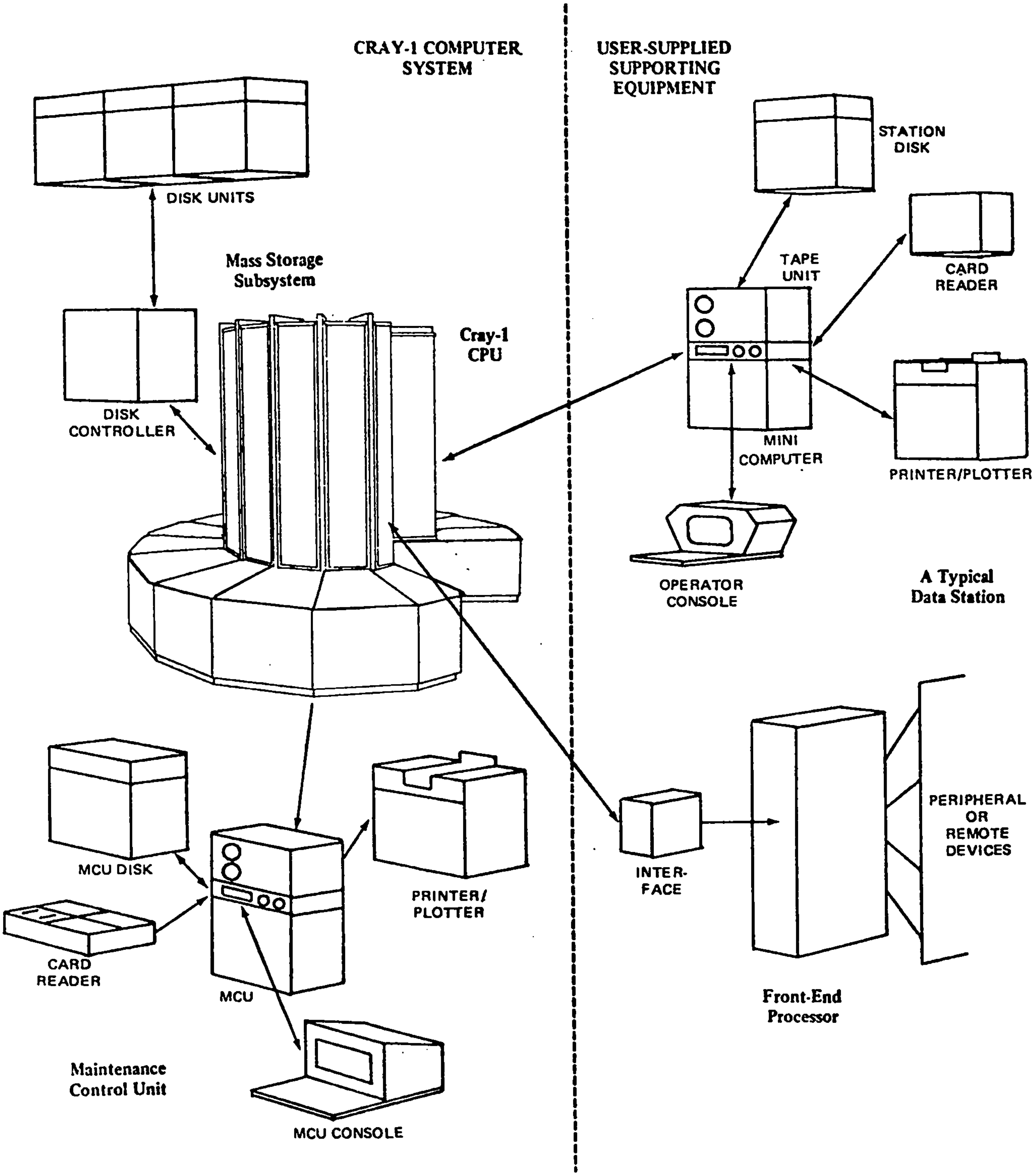
The CRAY-1 may be interfaced to front-end host systems through special controllers that compensate for differences in channel widths, machine word size, electrical logic levels, and control protocols. The interface is a Cray Research, Inc. product implemented in logic compatible with the host system.

SYSTEM MASS STORAGE

System mass storage consists of two or more Cray Research, Inc. DCU-2 Disk Controllers and multiple DD-19 Disk Storage Units. The disk controller is a Cray Research, Inc. product and is implemented in ECL logic similar to that used in the mainframe. Each controller may have four DD-19 disk storage units attached to it. Operational characteristics of the DD-19 units are summarized in the accompanying table.

CHARACTERISTICS
OF DD-19 DISK STORAGE UNIT

Bit capacity per drive	2.424×10^9
Tracks per surface	411
Sectors per track	18
Bits per sector	32,768
Number of head groups	10
Recording surfaces per drive	40
Latency	16.6 msec
Access time	15 - 80 msec
Data transfer rate (average bits per sec.)	35.4×10^6
Total bits that can be streamed to a unit (disk cylinder capacity)	5.9×10^6



DATA FLOW THROUGH SYSTEM

MAINTENANCE SERVICES

Cray Research, Inc. provides resident maintenance engineers on a contractual basis.