Project 7000

COMPANY CONFIDENTIAL

Maintenance Development Memo #9

SUBJECT: The Effect and Use of Checking Circuitry.
BY: Checking Task Force
DATE: July 29, 1957

1. Introduction

A. This is the first of several reports on the subject of checking for Project 7000. These reports are intended as a guide for designers and machine managers. This first report will define the problem, show what factors are involved, and give a criterion by which present and future solutions can be evaluated. Then the solutions which lead to the degree of system reliability and system maintainability which management feels desirable can be incorporated into the system design.

Subsequent reports will describe specific examples which show how the criterion has been applied to examples in hardware. These reports will be released periodically.

B. Scope

Checking is a very subjective term. It includes such things as error indication, fault location and error correction. These terms are more objective, therefore we will use them to avoid mis-understanding. The objective of the task force is to determine how each of these techniques can be used to the best advantage to achieve a system which has the best balance between reliability and maintainability.

C. Definitions

1. Synchronous Machine - a machine in which the time periods are of a fixed duration and are directly related to a fixed time base.

2. Asynchronous Machine - a machine in which the time periods are not related to a fixed time base but only to the characteristics of the circuitry.

3. Availability - fraction of power on time that the system is operational (sometimes expressed in percent).

4. Error - the failure to properly perform logical work in the time allotted.

5. Error, Apparent Intermittent - consistent failure under a random occurrence of specific circumstances to properly perform logical work in the time allotted.

6. Error, False - error of a checking circuit that signals an error in the circuit checked.

7. Error, Intermittent - random failures under many varied circumstances to properly perform logical work in the time allotted.

8. Error, Non-Redundant - a machine error leading to an error in the output of the machine.

9. Error, Redundant - a machine error that does not lead to an error in the output of the machine.

10. Error, Undetected- a class of errors that are not detected by the machine itself.

11. Fail Safeness - characteristic of a device which specifies the extent to which it will indicate its own errors.

12. Fault - a characteristic of any component or device which causes an error. (Note: to each type of error that has been defined there is a corresponding type of fault, which is defined as causing it.)

13. Mean Free Error Time - average time interval between errors.

14. Mean Repair Time - average time interval for all types of maintenance.

15. Mean Operation Time - average time interval between intervals of repair time.

16. Reliability - the probability of error free operation for a specified time interval.

II Factors Influencing the Quantity of Checking

A. Cost

The cost of hardware and engineering for checking is difficult to justify. The reason for this is the difficulty in placing a monitary value on the products which checking generates. There are basically two direct products of checking - ease of maintenance, and error indication to the user at the time of occurrence.

1. Experience from 700 series computers has shown that the problem of error diagnosis in electronic computers consumes in the order of 90% of the time required to clear the fault. In a system containing in the order of 100,000 transistors, the problem of fault location, utilizing previously employed schemes, appears to be a formidable one. In order to bring the system into a region with which we can cope, it is clear that we must have error detection and fault location.

   In addition to the cost of lengthy service techniques, which must be employed in the absence of checking, there is another more subtle cost. It is extremely expensive in engineering time to write diagnostic routines for a complex system. Besides this cost, many hours of machine time are spent testing machines or areas of machines which have no faults. If the precise nature and weakness of checking circuitry is understood in a system then only the checking circuits and the unchecked areas of the machine must be tested with diagnostic routines.

2. Our users object to any mal-function in our equipment but most violently object to errors which are undetected.

B. Reliability

As was pointed out above, error detection and fault location through additional logic is of vital importance to the ease of maintenance. For each component we add however, we also add another device which must be maintained. It is evident that if this process were carried to a full duplication point that as much time would be spent maintaining the checking equipment as would be spent on the system proper. There are two considerations which alleviate this problem:

1. Utilize highly efficient checking schemes. This means schemes which do the job required with a minimum of additional components.

2. Make the checking circuitry redundant in order that the system will not be out of service while a checking circuitry fault is being cleared. (assuming the user would allow temporarily unchecked operation)

A special consideration, in the category of reliability, is the use of checking with single bit error correction. The probability is highest that single faults will arise in a system rather than multiple faults. It is possible to detect and correct a single error caused by a single fault through the use of a Hamming type code. This will have considerable effect on the overall reliability of the system.

1. It will increase mean operate time. If a single fault in the system occurs during an operation period and its errors can be corrected, then the operate period can continue uninterrupted. The danger in such an operation is the occurrence of a fault in the error correction circuitry. Such a fault will cause mis-correction of valid data and may prove difficult to detect. It is most important that the probability of error in the circuitry and devices being checked and corrected is much greater than in the correction circuitry. This can be evaluated both from the component count ratio and from the relative reliability of the devices involved.

2. With any error correction scheme it is possible to define errors to a specific bit. This provides specific data to be used for fault location.

Another important consideration under reliability is the fail safeness of the checking circuitry. When checking is provided in a system the immediate reaction of maintenance people is that the section of the system being checked is working properly if no errors are being indicated. This can be shown to be false security since the checking circuits themselves can have undetected faults. The checking of this circuitry can be referred to as second level checking.

If one is wise in his choice of checking schemes then the probability of a fault in the checking circuitry should be considerably lower than in the system just due to the difference in population. If the checking circuitry is designed with maximum fail safe characteristics, the degree and frequency of second level checking for the system can be low.

An illustrative example of the statement "fail safe characteristics" can be given by discussing a comparing checking circuit. In this scheme of checking the universal connective is the "exclusive or" circuit. An analysis of the circuit shows that faults in this circuit when being used in comparing logic may go undetected. However, when the connective is used in parity counting a fault will generate an erroneous count for one of the two possible sums. Therefore one can say that for parity counting the connective is fail safe but in comparing it is not.

Unfortunately all checking schemes considered terminate in one or more comparing circuits. The need for special checking arises specifically in the case of these circuits. It is hoped that the number of these special cases remains small enough to make practical a unique testing method by which the servicing personnel can periodically make a rapid check of these circuits.

C. Availability

$$\% \text{ Avail.} = \frac{T0}{T0 + T_R}$$

$T0 = $ Mean Operation Time
$T_R = $ Mean Repair Time

The reliability and, consequently, the mean free error time of a computing system are of prime importance both to the user and designer of that system. Assuming that a practical maximum mean free error time has been realized (see paragraph on Reliability), a further step can be made to increase the time that a system is available to the user. From the above formula it is obvious that by decreasing the repair time the percent availability is increased.

Repair time is composed of two parts:

1. Emergency Repair time

   a. Diagnosis
   b. Repair

   This time commences at the time an error is detected and continues until the machine is again operational.

2. Preventive Maintenance time

   a. Detection
   b. Diagnosis
   c. Repair

   This time commences when the user releases the machine to the servicing personnel and continues until the machine is again available to the user.

It has been stated previously that approximately 90% of repair time is spent analyzing the trouble. Therefore, it is in this area that steps should be taken to reduce repair time. There are several methods of obtaining this reduction.

1. Machine analysis - by the proper use of checking circuitry to locate the fault to a pluggable module.

2. Servicing techniques

3. Use of more sophisticated test equipment

Only item (1), which is the most powerful and economical method, will be considered in this report.

Mean Free Error Time (T_{mfe})

As components are added to a computing system the probability of a fault is increased. Therefore, since the number of failures increases, the mean time between failures decreases. Assume a computer with a given number of components and mean free error time (A). If 100% added components are used then the total number of components has doubled, the failure rate has doubled and the mean free error time is halved. The mean free error time will again be halved if we again double the component count. We now have 300% added components.

Let X = % added components
Let Y = value of mean free error time
If X = 0, Y = A
X = 100, Y = A/2
X = 300, Y = A/4
X = 500, Y = A/8

(These large values of added components were used merely to point out the shape of the curve. Obviously the hyperbolic function still applies in the more realistic range of added components for checking.)

This is a hyperbolic function and is expressed by the formula:

$$T_{mfe} = \frac{100A}{X + 100}$$

where
A = T_{mfe} of the unchecked system

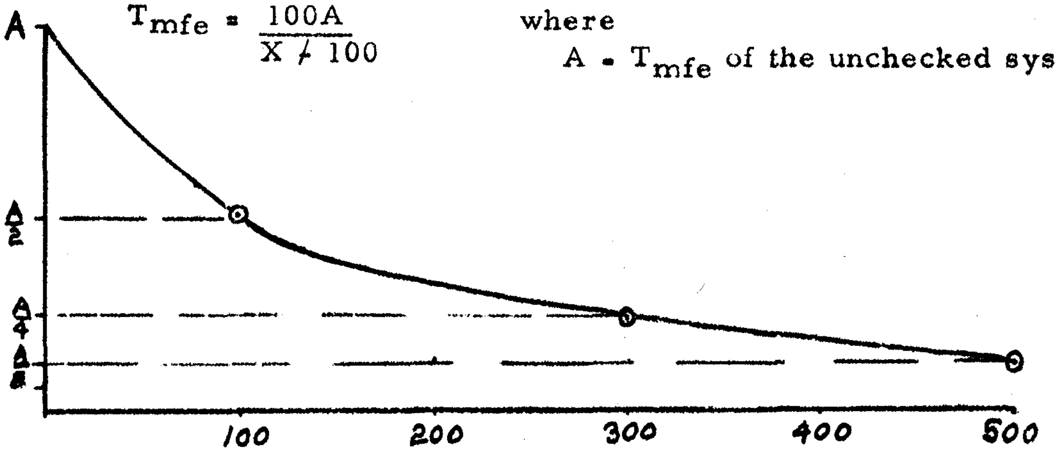


Figure C - 1

$T_{mfe} = T0 + T_R$ where $T0$ = mean operate time
T_R = mean repair time

The mean free error time has been defined as the mean time interval between errors.

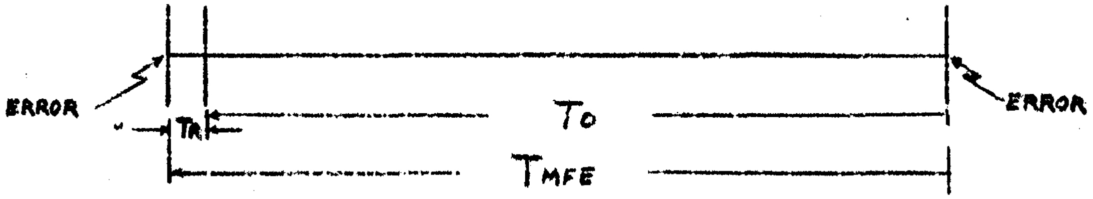


Figure C - 2

It is realized that this is not the definition generally accepted by the users of computing equipment. However, during the repair time following the detection of an error all components other than the one which caused the error, are functioning properly and are aging by a significant amount. Therefore, this measurement of mean free error time is, from the designer's and producer's viewpoint, the true indication of the ability of a system to operate error free.

Repair Time

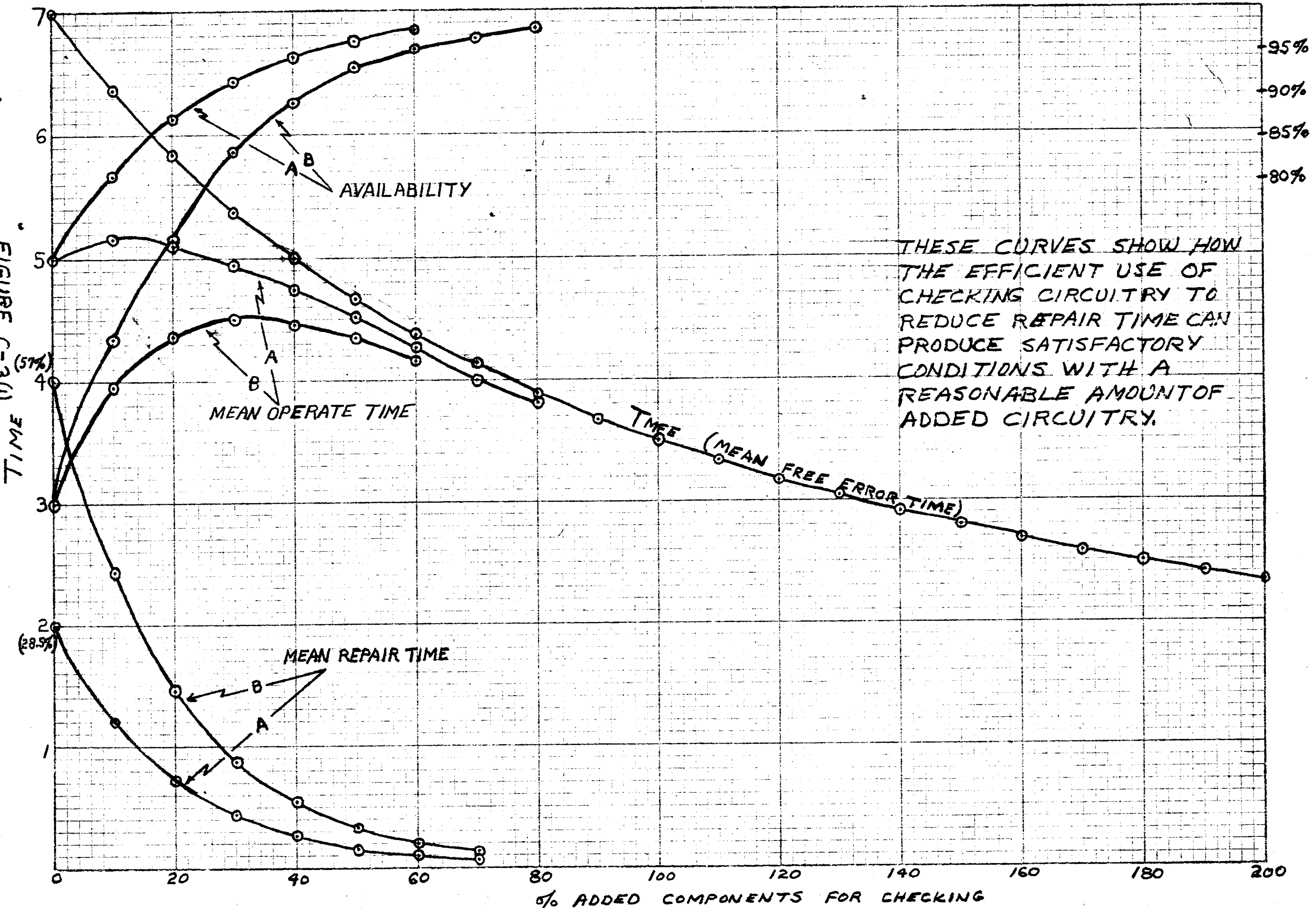
The most logical step in applying the first increment of checking circuitry to an unchecked system is to isolate the fault to a major area. Further additions of checking subdivide the major areas into intermediate and then into minor areas. Because the first increment has a greater effect on the isolation of a fault, its effect on repair time is also greater than succeeding increments. By use of this logical application of checking circuitry it would appear that the repair time is exponentially reduced by adding checking circuitry. (Refer to Figure C - 3)

Operation Time

Regardless of how refined the analysis of a fault becomes there will always remain some small amount of repair time. Since the mean free error time continually decreases there is a point where the mean free error time will equal the repair time. At this point the operation time would be zero and the machine would never be available to the user.

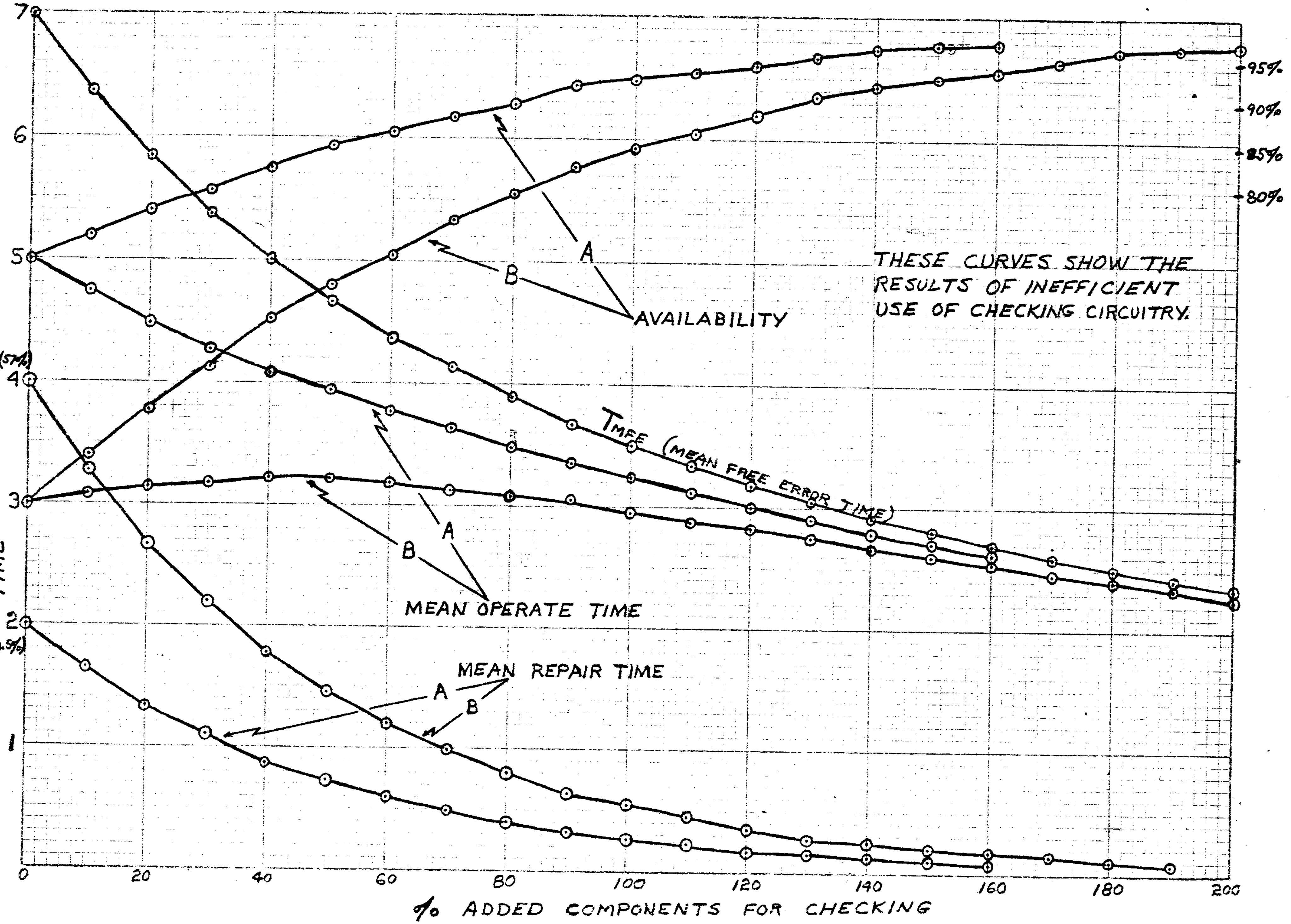
Long before this point is reached there is another condition-a crossover point - where the next increment of added checking circuitry will result in a reduction of mean free error time equal to the reduction in repair time.

Form No. 898--10 Squares to Inch
AMERICAN PAD & PAPER CO., HOLYOKE, MASS.



THESE CURVES SHOW HOW THE EFFICIENT USE OF CHECKING CIRCUITRY TO REDUCE REPAIR TIME CAN PRODUCE SATISFACTORY CONDITIONS WITH A REASONABLE AMOUNT OF ADDED CIRCUITRY.

FIGURE C-3(1)



THESE CURVES SHOW THE RESULTS OF INEFFICIENT USE OF CHECKING CIRCUITRY.

FIGURE C-3(2)

It is at this point that the operational time (T0) reaches a peak while the percent availability continues to rise and now the designer must make the decision as to whether this peak operational time or a greater amount of availability is the desirable feature.

It is obvious that the initial addition of circuitry must be such that it will drastically reduce the repair time. This is necessary in order to more than overcome the decrease in mean free error time and thus result in an overall increase in mean operation time (T0). This fact implies that we must decide initially which circuits require checking and then accomplish this checking with an absolute minimum of additional circuitry. It might be desirable in some cases to cut short the additional circuitry and leave a little more to human analysis. The human analysis could then be aided by external means. (Techniques of servicing, test equipment and error recorders.)

Analysis of an error may be accomplished in either one of two ways or a combination of both.

1. Machine analysis

   a. Error detection
   b. Fault location
   c. Error recording

2. Human analysis

   The amount required is dependent upon the amount of machine analysis.

The ideal situation appears to be where, on the occurrence of an error, the machine indicates the pluggable module containing the fault and no human analysis is required. From previous statements it is obvious that this situation may prove to be impractical or at least undesirable in some cases when the overall reliability and efficiency of the system are considered.

D. Circuit Usage

It can be easily shown that the value to the customer of checking in equally reliable circuits is greater for a circuit of high usage than one of low usage. Such is not the case for maintenance people. It is very possible that the seldomly used circuit may be much more difficult to locate faults in just due to unfamiliarity with the configuration. In summary, usage should not be employed to evaluate the need for checking.

E. Fault Location

As has been stated, the minimum degree of fault location should be to a base card. With the present size base card, an error indication shows that any one of one hundred component cards is at fault. Further location can be made in one of the three following manners:

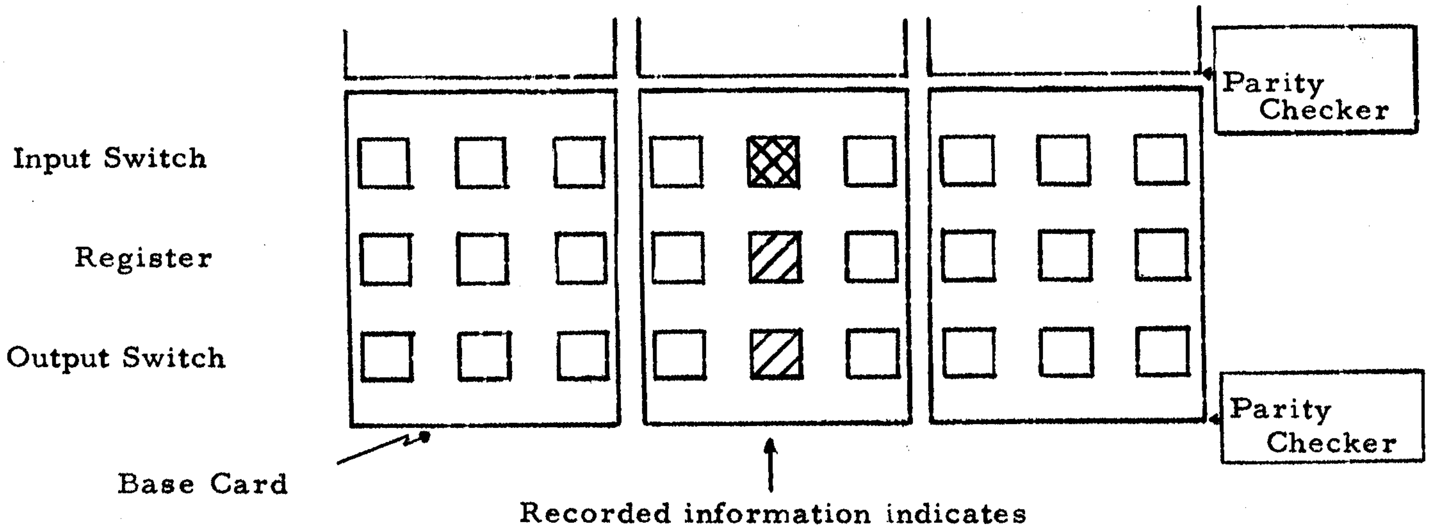
1. By observation with test equipment at the various test points.
2. By replacing component cards in part or in whole.
3. A combination of both methods.

It can be seen that with adequate recorded data, upon the occurrence of an error, that a considerably higher degree of fault location can be made. This assumes some logical deductions on the part of the maintenance people but should not lead to intensive systems analysis.

This requirement can be further met by populating the machine at discrete points with error detectors. The density of these devices will be determined by the logical package.

F. Packaging

If faults are to be isolated to a base card then the checking stations must be so arranged as to provide the required information. To minimize the number of checking stations required to accomplish this breakdown, it will be necessary to make the packaging compatible with the logic flow within the system. (Refer to Figures F - 1 and F - 2)



Recorded information indicates which position failed.

Figure F - 1

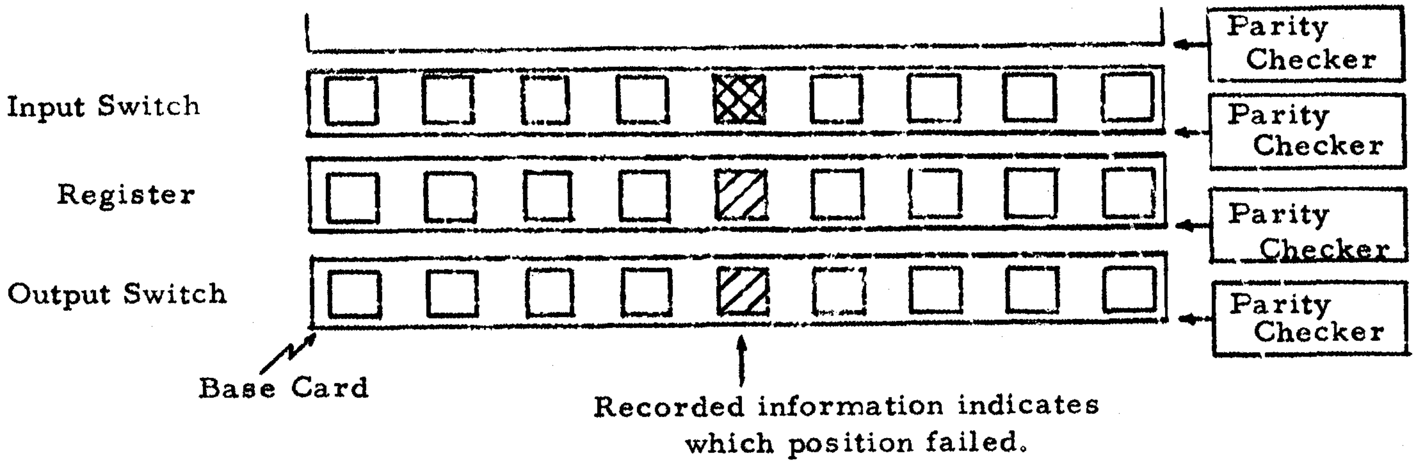


Figure F - 2

Figure F - 1

This method of packaging requires one checking station at the output switch to detect an error in any one of the three logical areas. Additional error information would indicate the position which failed. The base card for that position could then be replaced and the fault corrected external to the system.

Figure F - 2

This method of packaging would require three checking stations to accomplish the same results as stated in (1). If one checking station were used the location of the fault would be narrowed down to three base cards.

By packaging as many "series" logical elements as is physically and logically possible on a base card, the number of checking stations will be brought to an absolute minimum while maintaining a high degree of servicing "ease".

G. Effect of Asynchronous Nature of Computers

In synchronous computers the distinction between error detection and control circuits is rather sharp. For example, it is easy to distinguish the parity checker from the oscillator-controlled clock ring. By way of contrast, the functions of error detection and control within asynchronous computers are sometimes achieved within one and the same circuit.

To illustrate this fact, consider the (asynchronous) register - to - register transfer of data, a rather frequent event. To control the transfer, the following sequence of events take place:

1. A control line activates transfer gates and initiates the transfer for each bit.
2. Comparison circuits, one for each bit, recognize that each bit has completed its passage from the first register to the second.
3. An AND circuit recognizes that the outputs of all the comparison circuits are up and that the control line which initiated the transfer is still up.
4. The output of the AND circuit initiates the next action, which includes pulling down the control line.

The function of the comparison circuits in the above sequence is two-fold, for they indicate not only <u>when</u> the particular bits have transferred but also that the bits have <u>correctly</u> transferred. The effect of an incorrect or excessively delayed transfer (an error) is a delay or suppression of the next action, which could be called an <u>error stop.</u>

Thus a basic asynchronous computer already contains a certain amount of checking hardware that is essential for asynchronous operation. A factor to be considered is that although comparison circuits may check that events actually occur when the controls require them to occur, they do not check that undesired events have occurred. For the latter errors, there are needed other checking means, such as parity checkers, parallel flow, etc.

Marginal Checking

In a system made up of components which exhibit a gradual degradation it is possible to extend the customer's mean operate time by marginal checking. Marginal checking allows the removal of components before they actually fail during the customer's operation. This eliminates some short operate time from the summation of customer's operate times and thus increases the mean. Notice that if our marginal checking scheme does not cut short the life of the component by any great amount the mean free error time is uneffected.

Few if any details are available on the drift transistor at this time. However in order to illustrate the importance of this feature let us make some assumptions. Assume that the mean end of life for a transistor is 100,000 hours (about 12 years of 24 hours a day operation) and that 160,000 transistors are used in the system. It must also be assumed that the transistor follows the exponential law of reliability. Under these conditions we would expect to get a transistor failure every 0.625 hours. This mean operate time would be too short for most applications. If we were able to predict the components that would reach end of life before the next scheduled period, then the customer's mean operate time would be extended to the time between scheduled marginal checking periods.

There will of course be a second type of failure which cannot be predicted by this method. This type of failure can be termed catastrophic and its rate will then control the mean operate time, provided that marginal checking is optimized. In a very large system, which is totally dependent upon the operativeness of each component within the system, the mean end of life for each component type is quite important. A common method of describing failures is in percent per thousand hours of operation. Assume for a moment that we populate the system with a device which fails at the rate of .05% per 1000 hours. This would give us a mean end of life of 2,000,000 hours. If the system contained 200,000 components of this reliability then we would expect a failure each ten hours of operation.

The figures quoted in the above examples are not intended to be accurate, nor to alarm the reader. They are intended to point out the importance of individual component reliability to the overall reliability of a very large system. Unfortunately no figures are available for the proposed components for the Stretch system at this time.

Error Correction

Error correction can increase the mean operate time by correcting errors which result from transient or permanent faults. The term transient fault is used in this discussion to describe either an intermittent fault or random noise which may be inherent in certain devices.

The gain in mean operate time, as a result of error correction, is non linearly dependent upon the initial mean operate time of the system before correction is added. If errors are being caused by transient faults, the frequency of errors is not a direct function of component life as it is in the case of permanent faults. Since this type of error is not permanent, error correction will allow operation to continue until the advent of a double error. The probability of a double error is dependent upon the probability for single errors but is comparatively small in most cases. In order to illustrate the effectiveness of single error correction some arbitrary examples were tested. Assume a 72 bit data flow path with a 2 micro-second data rate and 5 minutes between transient errors. Such a system will allow 150 million data words to pass between errors. Single bit error correction will extend the mean operate time between uncorrected errors from 5 minutes to 16 hours of 200 fold. Other examples were tested. Examples of mean operate times greater than the above example yielded an interesting fact. The number of times greater that the operate time will become with correction increases with an increase in the uncorrected operate time. In other words, greatest gains or multiplication in operate time between errors occurs in systems of initially low error rate.

The maximum gain to be expected in the correction of a permanent fault is twice the normal operate time. In this case error correction takes place each cycle and the advent of the next error of either type will result in a double error.

An alternative is programmed correction thru restart procedure. It works for both cases, but it costs time. The customer must have programmed correction for the double error case, so the factors to be weighed are as follows: frequency of each type of error, the loss of time for programmed correction, and the cost of error correction circuitry.

Error correction offers fault location data to the maintenance staff if error recording is available. It also allows him diagnosis time following the occurrence of an error since the system can continue to function even with a permanent fault.

One of the hazards of correction circuitry is that of mis-correcting data due to a fault in the correction circuit.