M. Schatzoff
W. B. Harding

A Mathematical Model for Determining the Probabilities of Undetected Errors in Magnetic Tape Systems

Abstract: Mathematical models for evaluating the relative efficiencies of vertical and longitudinal redundancy-bit checking in magnetic tape systems are derived. Although these types of validity checking have been in use for some time, this is, to the authors' knowledge, the first quantitative statement of the probabilities associated with them.

Introduction

In electronic computing systems utilizing magnetic tape units, it is desirable to have validity checks at points where information is being exchanged between the tape units, computer, and printer, in order to insure the correct transmission of such information. It is also desirable to accomplish this as economically as possible.

A magnetic tape has seven channels running horizontally across it. A character is stored on the tape as a combination of bits in one or more of six of the channels, aligned vertically. On a binary-coded-decimal tape, a checking bit is added in the seventh channel if the other six channels contain an odd number of bits. This is the so-called redundant bit. On a binary tape, the checking bit is added if the character has an even number of bits. Although the argument presented here is derived for a binary-coded-decimal tape, the logic used to develop the argument for either case would be the same. Figure 1 shows the seven-channel code representation for all digits, letters, and special characters. In addition to the vertical checking it is possible to have a longitudinal checking bit for each channel, at the end of a record. In either case, the odd-even condition can be checked by means of switching circuitry which can be implemented much more easily than the circuitry which would be required for a count check. Choice of the former involves the risk of certain configurations of bit errors slipping through the system undetected. It is the purpose of this paper to derive the probabilities associated with this risk.

We may think of a magnetic-tape error as falling naturally into one of the following four categories:

	No Apparent Error	*Apparent Error*
No Machine Check	I	II
Machine Check	III	IV

This paper deals specifically with Type I and Type II errors. An apparent error is one which can readily be detected visually on a print-out. For example, substitution of a letter or special character for a digit would be classified as an apparent error. However, even this type of error would not be detected if transmitted from a tape unit to the computer. A word should be mentioned at this point concerning Type III and Type IV errors. Although a machine check indicates an error condition, it does not locate the error, or errors, as the case may be, nor does it indicate the number of errors. Thus, it is possible to have undetected errors even though there is a machine check. This is to some extent counter-balanced by the Type II errors which are sometimes detected despite the fact that there is no machine indication of an error condition.

A tape error may take the form of a lost or added bit in a character. For all intents and purposes it is not neces-

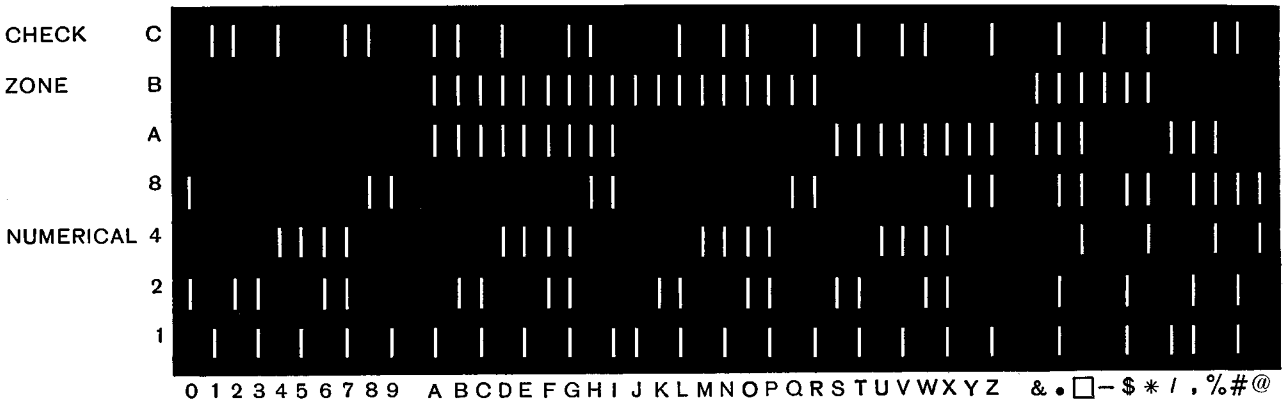

Figure 1 **Seven-channel code representation** *for all digits, letters, and special characters.*

sary to distinguish between these since the occurrence of either will have the same effect, that of changing the odd-even status. Thus, we will deal only with errors that do not cause a machine check, without regard as to whether they result from lost or added bits.

Mathematical derivations

There are two basic probabilities involved in determining the chances of Type I and II errors occurring. First, there is the probability of having (x) errors in a record of length (L). Let us call this p_x. Given p_x, we also have the probability (p_u) that the (x) errors will be arranged in such a manner that an error will not be detected. The conditional probability of having (x) errors, none of which are detected, is given by $(p_x|p_u)=(p_x)\,(p_u)$.

p_x is given by the $(n+1)$st term of the binomial expansion of $(q+p)^n$, where p represents the bit failure rate and n is the sample size, or in our case the number of cells. If n is large, but p is very small, then x will be small, and the general term

$$f(x)=\mathrm{C}(n,x)p^x q^{n-x}$$

of the binomial expansion can be approximated very well by the Poisson exponential function[1]:

$$f(x)=\frac{(np)^x e^{-np}}{x!}$$

which is tabulated for various values[2] of np and x.
The terms of the series

$$e^{-np}\left(1+np+\frac{(np)^2}{2!}+\frac{(np)^3}{3!}+\cdots+\frac{(np)^x}{x!}\right)$$

give the probability of exactly 0, 1, 2, $\cdots$, or x bit failures in a record of length L, where $L=n/7$. The Poisson exponential, like the binomial, has only one parameter, p, which represents the bit failure rate. Thus, knowledge of the bit failure rate will enable us to determine p_x. The value p_u, being the probability of having bit errors arranged in such a manner as to avoid detection, will differ depending on whether the system employs vertical checking only, or a combination of vertical and longitudinal checking. The probabilities for these two cases will be derived separately.

Vertical checking

If there is an odd number of bit errors in a record, the machine will always indicate an error condition. Thus, considering only the cases involving an even number of errors, we may proceed to derive the expressions for the various arrangements of bit errors which will satisfy the condition of maintaining an even bit count for each character. This probability can be expressed as the ratio of the number of such configurations to the total number of possible combinations of x bit errors in the record. We will list these for a sufficient number of cases to reduce the truncation error to a negligible amount (see Table 1). The derivation is in terms of permutations and combinations, where

P_r^n = number of permutations of n things taken r at a time

and

C_r^n = number of combinations of n things taken r at a time.

Longitudinal and vertical checking

For this type of situation it is impossible to have an arrangement of less than four bit errors, without having an error indication from one source or another. Table 2 shows the expressions for p_u in such a system.

Numerical evaluation

The expressions shown in Tables 1 and 2 can be quantified by the substitution of appropriate values of L and p. For example, assuming a record of 120 characters* and a bit failure rate of 0.001, the probabilities of non-detection for both types of checking are shown in Table 3. A failure rate of 0.001 is actually an extremely high one and probably would never be encountered unless the tape had been exposed to extremely adverse conditions of temperature and humidity. The figure is used here for computational ease as well as to illustrate the effectiveness of the checking systems under the most unfavorable type of condition. (It should be noted here that the material for this paper was derived as a result of an investigation into the reliability of a high speed

* This is a commonly used record length due to most printer limitations of 120 characters to the line.

Table 1 **Probability (p_u) of non-detection** *for vertical checking.*

Number of Bit Errors	Configuration	p_u
2		$\dfrac{C_2^7 P_1^L}{C_2^{7L}}$
4		$\dfrac{C_4^7 P_1^L}{C_4^{7L}}$
		$\dfrac{C_2^7 C_2^7 \frac{P_2^L}{2!}}{C_4^{7L}}$
6		$\dfrac{C_6^7 P_1^L}{C_6^{7L}}$
		$\dfrac{C_4^7 C_2^7 P_2^L}{C_6^{7L}}$
		$\dfrac{C_2^7 C_2^7 C_2^7 \frac{P_3^L}{3!}}{C_6^{7L}}$
8		$\dfrac{C_6^7 C_2^7 P_2^L}{C_8^{7L}}$
		$\dfrac{C_4^7 C_4^7 \frac{P_2^L}{2!}}{C_8^{7L}}$
		$\dfrac{C_4^7 C_2^7 C_2^7 \frac{P_3^L}{2!}}{C_8^{7L}}$
		$\dfrac{C_2^7 C_2^7 C_2^7 C_2^7 \frac{P_4^L}{4!}}{C_8^{7L}}$

printer which under strenuous test conditions involving over 16 million operations, experienced a bit failure rate of only 10^{-8}, which is 1/100,000 of our assumed rate.)

An inspection of Table 3 reveals some very interesting information. Even for an exceptionally high bit failure rate of 0.001, the chance of having more than eight bit errors in a record of 120 characters is less than one in two million. Thus, for practical purposes, truncation of $p_x|p_u$ for x greater than eight introduces an extremely small error, and $\sum_{x=2}^{2} p_x p_u$ for both types of checking yields excellent approximations to the desired probabilities. As a matter of fact, truncation at $x=2$ and $x=4$, for vertical and two-way checking respectively, would have given satisfactory results. It is interesting to note that for the case considered here, the addition of longitudinal checking reduces the risk of non-detection to a factor of less than one ten-thousandth of that for vertical checking alone. The efficiency of the redundant-bit checking system is further emphasized by considering the probabilities associated with bit failure rates of 10^{-4} and 10^{-5} for the case of two undetected errors (vertical checking only). For these cases, $p_x|p_u$ takes values of 1.0727×10^{-5} and 1.442×10^{-7} respectively.

An interesting aspect of the probability distribution function ($p_x|p_u$) is the fact that as the length of record in-

Table 2 **Probability (p_u) of non-detection** *for two-way checking.*

Number of Bit Errors	Configuration	p_u
4		$\dfrac{C_2^7 C_2^2 \frac{P_2^L}{2!}}{C_4^{7L}}$
6		$\dfrac{C_2^7 (C_1^2 C_1^5) C_2^2 \frac{P_3^L}{3!}}{C_6^{7L}}$
8		$\dfrac{C_4^7 C_4^4 \frac{P_2^L}{2!}}{C_8^{7L}}$
		$\dfrac{C_4^7 C_2^4 C_2^2 \frac{P_3^L}{2!}}{C_8^{7L}}$
		$\dfrac{C_2^7 C_2^2 C_2^7 C_2^2 \frac{P_4^L}{4!}}{C_8^{7L}}$
		$\dfrac{C_2^7 (C_1^2 C_1^5)(C_1^2 C_1^4) C_2^2 \frac{P_4^L}{4!}}{C_8^{7L}}$

Table 3 **Conditional probability ($p_x|p_u$) of undetected errors, vertical and two-way checking.**

Number of Bit Errors	p_x	p_u Vertical	p_u Two-Way	$p_x\|p_u$ Vertical	$p_x\|p_u$ Two-Way
0	0.406570	0	0	0	0
1	0.365913	0	0	0	0
2	0.164661	7.15137×10^{-3}	0	1.17755×10^{-3}	0
3	0.049398	0	0	0	0
4	0.011115	1.53079×10^{-4}	7.27978×10^{-6}	1.70147×10^{-6}	8.09148×10^{-8}
5	0.002001	0	0	0	0
6	0.000300	5.44873×10^{-6}	1.23057×10^{-7}	1.63462×10^{-9}	3.69171×10^{-11}
7	0.000039	0	0	0	0
8	0.000004	2.72690×10^{-7}	2.96023×10^{-9}	1.09076×10^{-12}	1.18409×10^{-14}
Σ	1.000000	7.31017×10^{-3}	7.40580×10^{-6}	1.17925×10^{-3}	8.09517×10^{-8}

creases, p_x also increases, while p_u decreases. Thus, it is apparent that an optimum record length may exist—i.e., one for which the probability of non-detection of errors is a minimum. Considering the simplest case of $x=2$, and vertical checking only

$$(p_x|p_u)=\frac{147p^2L^2e^{-7Lp}}{7L-1}.$$

This cannot be treated as a continuous function due to the nature of L, but is subject to evaluation by the calculus of finite differences or by numerical analysis. Minimization of $(p_x|p_u)$ for various values of x and p will indicate the distribution of optimum L, and might be a good subject for future investigation.

Summary

A general mathematical model for determining the probabilities of undetected magnetic-tape errors has been derived, and was evaluated numerically for a specific bit-failure rate and record length. The relative efficiencies for the two redundancy checking systems were demonstrated by this example, which substantiates conclusively the intuitive concept that two-way checking is vastly superior to vertical checking alone. Finally, a method for determining an optimum record length was indicated as a possible realm for future investigation.

References

1. Kenney, J. F., *Mathematics of Statistics, Part Two*, D. Van Nostrand Company, Inc., Seventh Printing, June 1947, pp. 29–31.
2. Pearson, E. S. and Hartley, H. O., *Biometrika Tables for Statisticians*, Volume 1, Cambridge University Press, 1956, pp. 194ff.

Received January 30, 1957