# GCF Compatibility With Packets and Data Compression

E. C. Posner
Telecommunications and Data Acquisition Office
P. Merkey
California Institute of Technology, Graduate Student

*Some missions using packets and/or data compression may want an undetected GCF block error rate of $10^{-6}$. Here we show that the present GCF meets this requirement.*

## I. Introduction

The DSN GCF High-Speed Data Lines and Wideband Data Lines use blocks of length 4800 bits, 22 of which are parity bits generated by the NASCOM polynomial (Ref. 1):

$$G(X) = X^{22} + X^{20} + X^{14} + X^{13} + X^{12} + X^{11} + X^8 + X^7 + X^5 + X^3 + X + 1 \quad (1)$$

or, in factored form,

$$G(X) = (X+1)^2 (X^{10} + X^3 + 1)(X^{10} + X^3 + X^2 + X + 1)$$

A *codeword* or code block in this code is a block of 4800 bits of 0 or 1 which is identified with a polynomial in $X$ with 0 or 1 coefficients and modulo-2 addition. A codeword then is polynomial of degree at most 4799 divisible by $G(X)$.

If an error is detected by virtue of the fact that the received block, regarded as a polynomial, is not exactly divisible by $G(X)$, an error is detected, and a retransmission *may* be requested depending on mission requirements (Ref. 2). Measurements conducted for the TDA Engineering Office show that at most one block in 200 contains at least one bit error, so, without retransmissions, a throughput of 99.5% (or more) is obtained. The actual probability that a block contains an error will of course be slightly greater than this because of undetected errors.

The undetected error rate is the probability that a block passes the divisibility test but nonetheless contains an error. For packet telemetry systems (and all the more for packet telecommand) as well as for missions with image or other data compression, it may be necessary to keep this undetected block error probability below $10^{-6}$ (Ref. 3). Is the block error probability below $10^{-6}$? It is hard to measure directly with current instrumentation, but this article shows that the requirement is met anyway.

## II. Code Structure

Reference 4, plus a little calculation, shows that a length-4800 binary code with 22 check bits is at most single-error-correcting. Since $X + 1$ is a divisor of $G(X)$, the weights of the codewords are all even. Hence our code has minimum distance 2 or 4.

Actually, the minimum distance is only 2. This is because of the following argument. From Ref. 5, define $M^{(1)}(X) = X^{10} + X^3 + 1$, a primitive polynomial, with root, say, $a$. Then

$M^{(3)}(X)$, defined as the irreducible polynomial with root $a^3$, also has degree 10 and is the second degree -10 factor in $G(X)$ of Eq. (1). Thus:

$$G(X) = (X+1)^2 M^{(1)}(X) M^{(3)}(X)$$

Here $M^{(1)}(X)$ divides $X^{1023} + 1$ and no smaller binomial, by primitivity. The theory of equations shows that $M^{(3)}(X)$ divides $X^{341} + 1$ and no smaller (because $1023 = 3 \times 341$). So $H(X) = (X+1) M^{(1)}(X) M^{(3)}(X)$ also divides $X^{1023} + 1$ and no smaller binomial. Thus, $G(X) = (X+1) H(X)$ divides $(X^{1023} + 1)^2 = X^{2046} + 1$. But $X^{2046} + 1$ is a codeword, being a multiple of $G(X)$. Hence there are codewords of weight 2, and the code is of distance 2, not 4.

We note in passing that the code generated by $M^{(1)}(X) \cdot M^{(3)}(X)$ alone, of degree 20 (20 check bits) generates a distance-5 BCH code of length 1023 (Ref. 6). The NASCOM code however uses $(X+1)^2$ times this, of degree 22, but, more importantly, out to length 4800. How many codewords of weight 2 are there in the NASCOM code? We need to know this to estimate error probabilities.

If $C(X)$ is a codeword of weight 2, then

$$C(X) = (X^i + 1) X^j$$

for some non-negative integers with $i + j$ at most 4799. Now $G(X)$ divides $C(X)$ since $C(X)$ is a codeword, so $G(X)$ divides $X^i + 1$. All the more, $M^{(1)}(X)$ divides $X^i + 1$, so $i$ is a multiple of 1023. Since $(X+1)^2$ must divide $X^i + 1$, $i$ must be even. Since $i + j$ is at most 4800, $i = 2046$ or 4092.

If $i = 2046$, $j$ can range from 0 to 4799 - 2046 = 2753. There are 2754 codewords $X^{i+j} + X^j$ with $i = 2046$. If $i = 4092$, $j$ can range from 0 to 4799 - 4092 = 707, so there are 708 more codewords of weight 2. Altogether, the code has 2754 + 708 = 3462 codewords of weight 2.

## III. Independent Errors

First suppose bit errors in a block occur independently. This is not necessarily the case. But, if it were, the input bit error probability is derivable from the fact that at most one block in 200 contains at least one error. If $p$ is the bit error probability, it will be small. So the *block* error probability $r$ is about $4800p$, and

$$4800p = \frac{1}{200}$$

$$p = 0.96 \times 10^{-6}$$

What then is the undetected block error probability of this code? We will upper-bound it as the probability of an undetected double error plus the probability of *all* quadruple (or higher) errors. We can ignore triple errors because they are of odd weight, hence detectable. The probability of a *particular* double error is

$$b = p^2 (1-p)^{4798}$$

$$b = 9.174 \times 10^{-13}$$

There are 3462 double errors which are codewords, so the probability of an undetected double error is 3462 times $b$, or $3.176 \times 10^{-9}$.

We can upper-bound the probability of quadruple or higher errors in the independent-error case by the probability of quadruple errors alone, because $p$ is so small. This can be made quantitative using the "tail estimate" for the binomial distribution (Ref. 5, App. [A.5], p. 467), but we omit it. The probability of quadruple error is $(4800 \times 4799 \times 4798 \times 4797)/24$ times $p^4 (1-p)^{4796} = 1.868 \times 10^{-11}$. Adding the previously derived probability of undetected double error, we find that an upper bound to the undetected block error probability when using the NASCOM 22-bit polynomial with independent errors is

$$3.195 \times 10^{-9}$$

This more than meets the undetected block error probability requirements.

## IV. Arbitrary Error Structure

Now suppose we know nothing of the error patterns, just that one block in 200 contains at least one detected error. This is almost the same as one block in 200 containing an error, detected or not, and it is this that we shall actually assume. Let us even assume, as the worst case, that every block with an error contains at least two errors. In fact, a little thought shows that the worst case is when *all* the errors are double errors. This is essentially because higher error patterns are so numerous that they tend to distribute themselves randomly with respect to the code. Thus, approximately $2^{-22} = 2.5 \times 10^{-7}$ of the higher error patterns can be expected to be codewords. What is the exact fraction for double errors?

There are $(4800 \times 4799)/2 = 1.152 \times 10^7$ error patterns of weight 2, but, as we have seen in Sec. II, only 3462 codewords of weight 2. If we assume, as we do, that all error patterns of

a given weight, in particular of weight 2, are equally likely to occur, the probability that a double error pattern is a codeword is

$$3462/1.152 \times 10^{7} = 3.006 \times 10^{-4}$$

This is almost 1261 times as large as the $2^{-22}$ probability we would get if double error patterns distributed themselves randomly with respect to the entire code.

The probability of undetected error in these strange circumstances can be found as follows. We start with 1/200, the probability that there is at least one (and so, by our assumptions, exactly two) errors in the block. We multiply this by the probability that the error pattern is a codeword and, hence, is undetected. Thus, the undetected block error probability is

$$\frac{1}{200} \times 3.006 \times 10^{-4} = 1.503 \times 10^{-6}$$

This slightly exceeds the $10^{-6}$ undetected block error probability requirement. However, the assumptions under which we derived this high error probability are so extreme that we can consider that we do meet the $10^{-6}$ requirement. For example, if half the blocks in error contain a single error (which is, of course, detected) and half contain a double error, the above estimate drops by a factor of 2 to $7.5 \times 10^{-7}$, and the requirement is met.

We restate here that if we monitor the links to make sure we are getting the 99.5% throughput, then we will also be confirming the $10^{-6}$ or less undetected block-error probability as well. The GCF with error detection by the NASCOM 22-bit polynomial is compatible with the extremely low undetected GCF block error probabilities that some missions may want in the packet era.

## Acknowledgments

We are indebted to the referee for sharpening the results of the last section.

## References

1. NASA Goddard Space Flight Center, *NASCOM Error Detection*, NASA Goddard Space Flight Center Document GSFC-844-71-09, Sec. 3.1.6.2, (i), NASA Goddard Space Flight Center, Greenbelt, Md. (no date given), pg. 9.

2. Nightingale, D., "High-Speed System Design, Mark IIIA," Part B (pp. 103-105) of "VIII. GCF Development," *JPL Space Program Summary 37-66*, Vol. II, The Deep Space Network, Jet Propulsion Laboratory, Pasadena, Calif., Nov. 30, 1970, pp. 99-110.

3. Posner, E. C., and R. Stevens, "TDA Assessment of Recommendations for Space Data System Standards," *TDA Progress Report 42-77* (May 15, 1984), Jet Propulsion Laboratory, Pasadena, Calif., pp. 75-85.

4. Merkey, P., and E. C. Posner, "Optimum Cyclic Redundancy Codes for Noisier Channels," *TDA Progress Report 42-76* (Feb. 15, 1984), Jet Propulsion Laboratory, Pasadena, Calif., pp. 189-195.

5. Peterson, W. Wesley, and E. J. Weldon, Jr., *Error-Correcting Codes*, Second Edition, MIT Press, Cambridge, MA (1972), App. C, pp. 472-492.

6. Mac Williams, F. J., and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North Holland, Amsterdam (1977), Chap. 7, para. 6, p. 201.