REDUNDANCY REMOVAL ALGORITHMS APPLIED TO GEMINI XII DATA

H. M. JORY
Senior Engineer
McDonnell Douglas Corporation.

Summary This report presents the results of a company funded computer study to deter-mine the effectiveness of redundancy removal algorithms as applied to manned spacecraft data. The company familiarity with and access to manned space flight data provided an almost unique opportunity to study this method of data compression using data representative of that which will be required from a Manned Mars Mission. A total of 28,500 seconds of the Gemini XII flight is examined using seven algorithms and three different tolerance bands. Over eleven million samples have been examined using terminology and descriptions consistent with previously published literature to allow direct comparison of actual flight data with previous results using synthetic data.

The outputs from the computer presented the following information:

A. Compression ratios as a function of technique, channel number and type of data for each of the activity periods.
B. Buffer input rates and accumulated queue lengths every 2.4 seconds for the ZFN technique.
C. Error distribution, for each of the techniques for six different apertures.

The results indicate that the zero order - variable corridor - adjusted preceding sample transmitted (ZVA) technique can provide data compression ratios of 187:1 using a 1.2% tolerance. Nominal buffer sizes of 20K bits are adequate to handle the data activity period involved. The error distribution evaluation indicates that the error distribution is primarily a function of the technique and the aperture.

Introduction In recent years the rapid advances in spacecraft technology have produced attendant increases in system and mission complexity and in mission duration. This increase in system complexity has required that an increased number of measurements be telemetered and the lengthening of the mission duration has required that they be transmitted over greater distances as we explore farther and farther from earth. As the missions become more expensive and more “one shot” affairs, it becomes increasingly important that telemetered data be reliably received on the ground , It might appear that the gains in technology (i.e., larger boosters) would allow proportionate increases in communication systems size, weight, and power. However, these gains in technology are often used to increase the mission complexity and duration, which require even more measurements from even greater distances.

Spacecraft data acquisition systems today, have evolved as synchronous systems of either time or frequency multiplexing. In this type of system, data rates for the individual measurements are assigned as the product of the maximum frequency of interest and a constant. This constant can be determined using one of the sampling theorems in use today. Sometimes it's 2 samples per cycle; sometimes 5 samples per cycle; sometimes according to D. D. McRae's interpolation tables;[1] and sometimes by best engineering estimates. Consequently, a correctly designed, fixed format system is only efficient when the measurements are producing their maximum rate of change. Most of the time, the measurements are quiescent, and the fixed format system yields poor information efficiency.

Once the data rates are established for the various measurements, they can be summed and will determine the channel rate capability required of the communications link. In the past, channel rate was not a limiting factor, but the number of different data rates were limited so that the airborne and/or ground equipment complexity could be reduced. In order that the maximum frequency of interest was obtained, measurements frequently were sampled at higher rates than the theoretical data rate. This introduces a small amount of communications inefficiency, but this is insignificant when compared to that resulting from the necessity to fix the sampling plan.

How then, can we improve the information efficiency of the communications link? The answer, hopefully, is some form of data compression.

A. Reduced Bandwidth and Power Requirements - The savings here are fairly obvious, as shown by the range equation,[2] the received power at the ground is reduced by the square of the distance. Also, transmitters are not very efficient power converters. To achieve the 2 watt output of the Gemini transmitters for instance, required almost 20 watts of basic spacecraft power.

B. Reduced Data Storage Requirements - The flight of Gemini 12 (94.5 hour mission) produced 108 data tapes at the World Wide Tracking stations. In the Gemini mid-program Conference Report,3 it was noted that 250,000,000 data points were transmitted from Gemini V (191 hour mission) on the delayed time system alone. If this data were reduced, it would have required 1 million pages of tabulation or 750 thousand pages of plots; this in a year in which we launched six Gemini's. This overwhelming volume of data caused the NASA to adopt a policy of reducing only the launch, re-entry, and flight segments of planned activity from each flight.

Any other segments of this flight were reduced on an “as required” basis. Although data compression techniques were utilized in the processing of telemetered data record on the ground, adoption of data compression techniques, prior to transmission from the space vehicle would have reduced the storage requirement by a factor equivalent to the achieved gross compression ratio.

C. Reduced Data Handling and Analysis - The adoption of data compression techniques will decrease the amount of demultiplexing and analysis required by the user. This also decreases the time interval between data collection and dissemination of information to other users. Methods of data compression such as redundancy removal provide addressing and time correlation, thus completely deleting some of the required functions in the normal demultiplexing operation.

Compressed PCM System Figure I is the basic block diagram for a compressed PCM system (CPCM). The sensors convert physical phenomena into an electrical signal proportional to the phenomena. Multiplexing is accomplished by sequential sampling of the data sensors at the data rates determined by the frequency of interest and the selected sampling theorem. The A/D Converter changes the analog representation of the physical phenomena into a binary digital code which is presented to the Comparator for comparison with previous data stored in the Predictor. The Predictor anticipates the value of future data samples on the basis of selected previous sample values, and/or on the basis of knowledge of the data characteristics. An aperture or tolerance band, defining the allowable prediction error, will have been assigned previously by the data user in accordance with his accuracy requirements. If the data sample being examined falls within the same aperture or corridor as the predicted value, the sample is judged to be redundant and is discarded. However, if the data sample falls outside the aperture, it is retained and sent to the buffer for subsequent transmission over the RF link, and also sent to the predictor to be used in predicting the next sample. The samples chosen by the comparator for transmission will be sent to the buffer at an irregular rate. The role of the buffer in the system is that of accepting and storing these aperiodic data samples, so that they can be sent to the transmitter at a uniform. rate.

Redundancy Removal Algorithms What are effective redundancy removal techniques? If you examine a number of different technical papers, you will find different explanations of essentially the same processes each with different conclusions and descriptions. One of the most thorough descriptions is given in a paper presented by Dr. R. Simpson at the ITC in 19664, which presented descriptions of 8 different redundancy removal processes and their associated interpolation methods. The various redundancy removal algorithms are described using three descriptors. These are: the order, the type corridor, and the particular transmitted sample. The order of an algorithms denotes the order of the derivative -used to predict subsequent samples. The corridor, which can be fixed or variable, refers to how the prediction of subsequent samples is made. Finally, the transmitted sample determines what information is transmitted when a non-redundant sample occurs. This can be the non-redundant sample, the sample preceding the non-redundant sample or an adjusted value for the preceding sample. Seven of these eight algorithms have been adopted as standards for this study. These are:

Order	Corridor	Transmitted Sample	Descriptors
Zero	Fixed	Non-Redundant	ZFN
Zero	Variable	Preceding	ZVP
Zero	Variable	Adjusted Preceding	ZVA
First	Fixed	Preceding	FFP
First	Fixed	Adjusted Preceding	FFA
First	Variable	Preceding	FVP
First	Variable	Adjusted Preceding	FVA

The eighth method, First Order, Fixed Corridor, Non-Redundant Sample Transmitted (FFN) was not used because the error using this method can approach plus or minus full scale.

Compression Ratio Before discussing the compression ratios which have been achieved, it should be emphasized that the term "data compression" differs from the tem "gross compression". "Data compression ratio" is defined as the number of samples into the predictor (the multiplexer sample rate) over the number of non-redundant samples. "Gross compression ratio" is the number of bits into the predictor, (the product of multiplexer sample rate and word length) over the number of bits into the buffer (the product of the non-redundant words and the word length the necessary bits for addressing, time correlation, synchronization, and error correcting codes.) It is assumed that the gross compression ratio will be approximately one-half the data compression ratio.
The measurands from the Gemini XII parameter list were divided into four categories; (1) 75 channels of bi-level or event data which was investigated on a "no aperture!' basis (i.e., each change of state is transmitted), (2) 31 channels of low rate data sampled at .416 samples per second (sps), (3) 71 channels sampled at 1.25 sps, and (4) 10 channels of high rate data O at 40 sps, 3 at 20 sps, and 4 at 10 sps). The low and medium rate data was examined using three techniques. These were ZFN, ZVP, and ZVA. The ten high rate channels were examined using all seven algorithms.

Since the flight of GT 12 was 94.5 hours long, it was necessary to find a method of reducing the computer time without compromising the results significantly. The flight has been examined to determine the activity periods, and a one to one and a half hour segment was selected for each of seven periods. The results were then extrapolated to cover the entire flight. Figure II shows the periods selected and the time over which the results were extrapolated. For instance, the sleep period tape was almost one hour long

(3400 second) and the results were extrapolated to cover the 20 hours of actual sleep periods.

The above extrapolations left almost 51% of the mission, which had not been accounted for. It was then assumed that the 51% not accounted for, had about the same average compression ratio as the 49% which was accounted for. Thus the following model was constructed:

$$C_D = \frac{R_1 T_1}{N_1 T} + \text{-----} \frac{R_8 T_8}{N_8 T}$$

Where:

C_D = overall flight, data compression ratio

R_1 = total samples examined for the period denoted by the subscript.

N_1 = non-redundant samples for the period denoted by the subscript.

T_1 = extrapolated time for the period denoted by the subscript.

T = total flight time (340,200 seconds)

Subscripts:

1 = Launch
2 = Rendezvous
3 = Standup EVA
4 = Umbilical EVA
5 = Tether
6 = Sleep
7 = Re-entry
8 = Unaccounted for periods

Table I shows the compression ratios for the various activity periods an the results of the extrapolation. The superior noise eliminating properties of the ZVA technique are apparent in the difference between the ± 1 and the ± 3 count aperture columns.

Buffer Queue Lengths In order to determine the buffer queue lengths in the Gemini XII data, the number of non-redundant words for each 2.4 second interval were printed as well as the accumulated totals for the ZFN technique. In addition to this data, time histories and histograms were printed at the end of each run.

An examination of the time histories revealed that the largest data activity peak occurred during the reentry period. From this data a set of models (Figure III) were constructed so that queue length (q) could be determined by the peak input data rate (P), the buffer readout rate (R_o) and the duration of the peak (T) or:

$$q = (P-R_o)T$$

If we assume an output word rate (R_o) of 50 words per second then for an aperture of + .4% from Figure II we have:

$$q = (75\text{-}50)612$$
$$q = 15{,}300 \text{ words}$$

And if we use 16 bit words (8 data bits plus 8 bits of correlation) a 244,800 bit uncontrolled buffer would be required to prevent loss of data due to buffer overflow. If we increase the aperture to 1.2% then:

$$q = (P\text{-}R_o)T$$
$$q = (52\text{-}50)612$$
$$q = 1224 \text{ words}$$

Assuming 16 bit words, a 19,584 bit buffer will handle this data. Figure IV shows the average input data rates for the flight periods examined. An examination of this chart shows that an output rate of 50 words/second (800 bits/second) is much too high for most of the flight. The difference between apertures of ± .4% and ± 1.2% during the sleep period is due to the fact that the spacecraft is powered down and many channels are noisy. Figure V describes the output bit rate versus queue length for the reentry model.

Error Distribution In order to determine the error distribution, ten channels of Gemini data were examined for 81 seconds during the re-entry phase of the flight. A total of 17,600 samples were examined. Each channel was examined for redundancy using the seven redundancy removal algorithms. Interpolation between the non-redundant points was accomplished as required and all points were compared with the interpolated line to determine the magnitude of the error. The total error band was divided into 40 ranges of values and the errors falling in each range were tabulated. This was done for each individual channel for apertures of 1, 2, 3, 4, 6, and 7 counts (full scale = 254 counts). The results of the ten,channels were combined and the overall error distribution was tabulated for each algorithm. The r.m.s. error was also determined and is shown in Table II. Figure VI is a sample of the error distribution histograms for ZVA, FFA, and FVA. These show the definitely bi-modal characteristics reported by Dr. R. Simpson[5] Figure VII shows the error distribution for the other four techniques.

Conclusions Although the size of th sample examined is many times that which has been previously reported,[4,6,7] the results in general, complement the earlier work. Four techniques have been adopted for future work. These are the ZFN, ZVA, FVP, and FVA. The ZVP is discarded because the ZVA is superior in noise elimination and compression ratio while the complexity is about the same. The FFA and FFP techniques are discarded because of low compression ratios which is probably due to the propensity to oscillate which has been previously reported.[4]

The examination of buffer queue lengths was probably the most revealing. Even assuming an uncontrolled buffer, the buffer size is not prohibitive, while some form of buffer control[8] would reduce the size even further. The error distributions for the data examined appear to differ from that reported in previous work,[4,5] but the use of rms error and the assumption of normal distribution appears to be practical since the difference in rms errors for the various techniques are small.

Future Work Future effort will be expended on implementation, addressing and time correlation to optimize the overall system design. Investigations are also being initiated to examine the feasibility of transmission of airplane flight test data via common telephone carrier from the remote test facilities to a centralized data collection facility. In addition, it is hoped that a redundancy removal system can be placed on an orbiting spacecraft as an experiment or as backup to the primary telemetry system. This would be primarily to develop user confidence and gain operational experience in the handling of compressed data.

Acknowledgements The author is indebted to Mr. Louis Mizell who wrote the computer programs for the study.

REFERENCES

1. D. D. McRae, "Interpolation Errors" Advanced Telemetry Study Technical Report #1. Part 1, 18 May 1961.
2. Lester C. Van Atta, "Interplanetary Commmication" International Science and Technology Nov. 66.
3. Gemini Mid Program Conference Report NASA SP-121.
4. R. S. Simpson, "Evaluation of Redundancy Reduction Algorithms" ITC Proceedings Vol. II 1966.
5. R. S. Simpson, et al "A Study of Redundancy Removal for Saturn Telemetry Data" Systems-Engineering Group, University of Alabama Technical Report #5 Nov. 1964.
6. L. W. Gardenhire., "Redundancy Reduction, The Key to Adaptive Telemetry" NTC Proceedings 1964.
7. R. A. Schomburg, "Computer Simulation of a Data Compressor for Aerospace Telemetry Systems" Proceedings of the 1962 Nat. Symposium on Space Electronics and TM.
8. R. Aylaworth, et al "Techniques and Applications of Data Compression" LAC(LMSD)TR 8-51-63-2 June 1963.

Table I - Spacecraft XII Data Compression Ratios

PERIOD	TOTAL TIME EXAMINED (SEC)	TOTAL SAMPLES EXAMINED	TECH-NIQUE	COMPRESSION RATIO		
				APERTURE ± 1 COUNT	APERTURE ± 3 COUNT	APERTURE ± 7 COUNT
	T_X	R_X		$\frac{R_X}{N_X}$	$\frac{R_X}{N_X}$	$\frac{R_X}{N_X}$
LAUNCH	6092	2,408,553	ZFN	29.6	79.6	120.0
			ZVP	24.3	68.6	109.8
			ZVA	49.1	103.5	144.2
RENDEZVOUS	4000	1,726,695	ZFN	42.9	151.1	234.3
			ZVP	34.3	126.9	216.4
			ZVA	86.5	200.3	307.4
STANDUP EVA	3649	1,557,168	ZFN	47.5	152.2	249.5
			ZVP	36.6	129.5	224.1
			ZVA	89.8	205.8	311.9
UMBILICAL EVA	5401	1,489,503	ZFN	21.2	39.3	51.0
			ZVP	19.0	37.2	48.9
			ZVA	31.4	47.9	59.1
TETHER	3559	2,188,493	ZFN	47.6	136.0	200.2
			ZVP	36.0	117.2	187.5
			ZVA	82.1	175.4	241.4
SLEEP	3408	1,030,481	ZFN	16.8	386.6	609.7
			ZVP	15.3	337.1	583.1
			ZVA	91.5	537.2	891.0
REENTRY	2400	1,442,492	ZFN	13.5	22.9	33.7
			ZVP	12.1	21.7	31.7
			ZVA	18.2	29.5	41.6
EXTRAPOLATED RESULTS (C_D)			ZFN	24.95	130.30	212.66
			ZVP	21.97	123.73	201.76
			ZVA	58.64	187.69	290.84

Table II - RMS Errors

APERTURE	± .3937	± .7874	± 1.1181	± 1.5748	± 2.3622	± 2.7559
PEAK ERROR	.3937	.7874	1.1181	1.5748	2.3622	2.7559
ZFN	.2492	.4128	.5392	.7380	1.0816	1.2095
ZVP	.2373	.3968	.5214	.7693	1.1097	1.3269
ZVA	.2742	.4697	.6065	.8736	1.2085	1.4125
FFP	.1936	.3054	.3941	.5405	.8850	.9952
FFA	.2362	.3988	.5275	.7265	1.0305	1.2594
FVP	.2015	.3190	.4314	.6341	.9792	1.1547
FVA	.2353	.4015	.5446	.7761	1.1939	1.3522

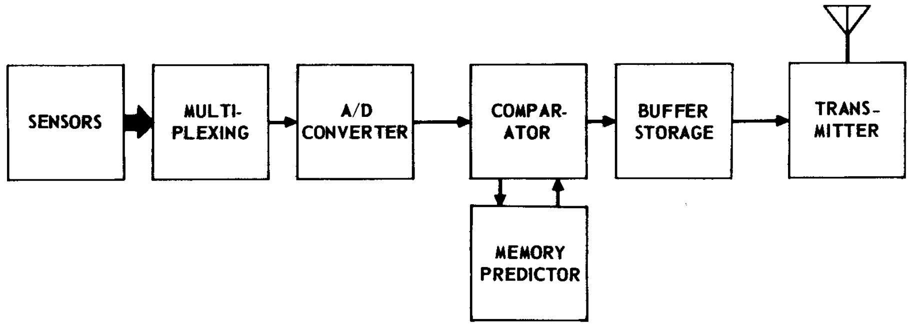

Figure I - Compressed PCM System Block Diagram

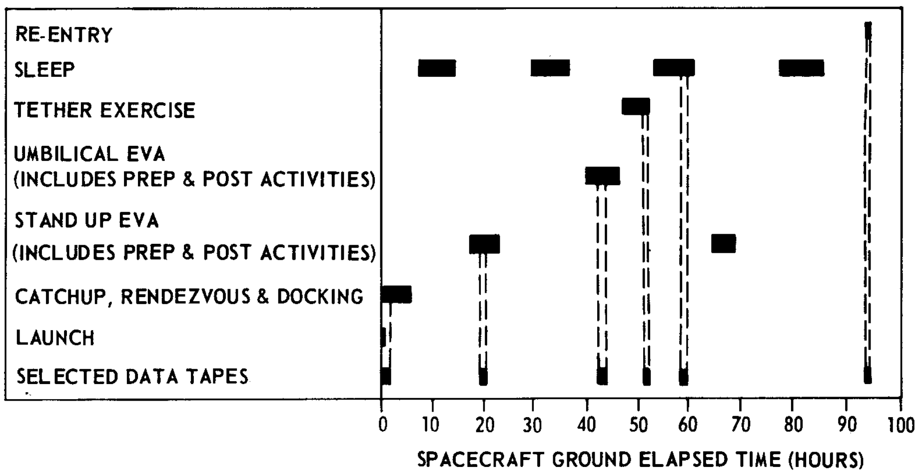

Figure II - Spacecraft XII Activit Periods

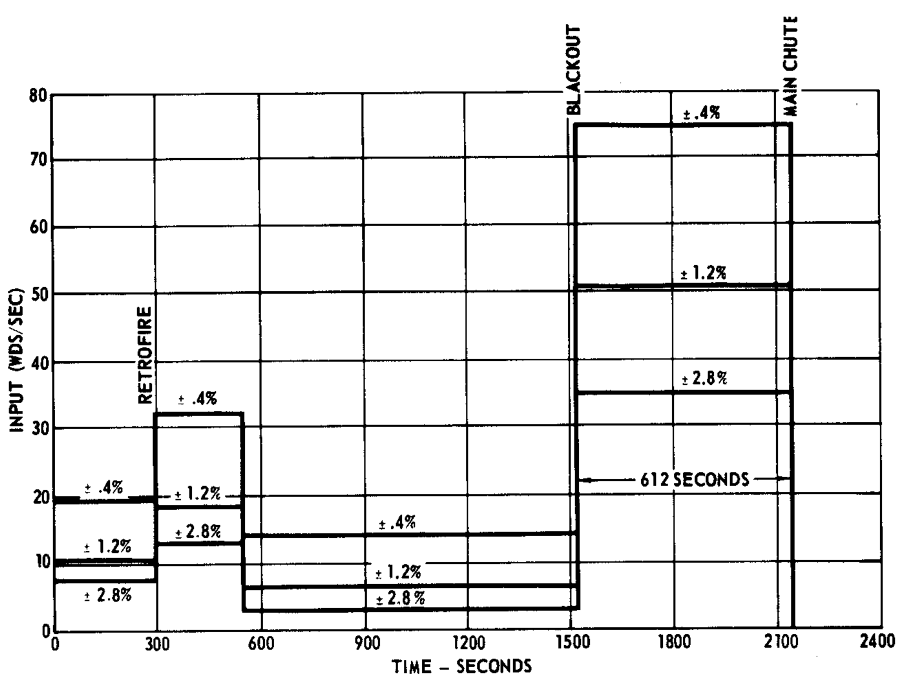

Figure III - Re-entry Period Mode 1 for Queue Length

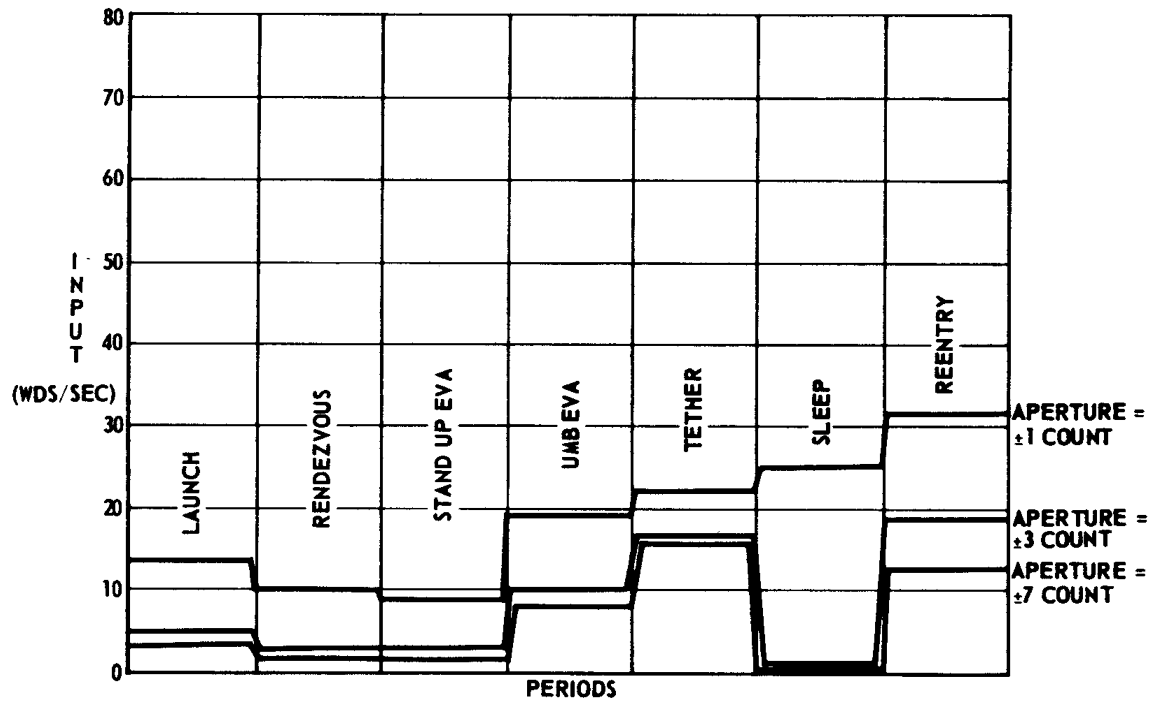

Figure IV - Average Buffer Input Rates

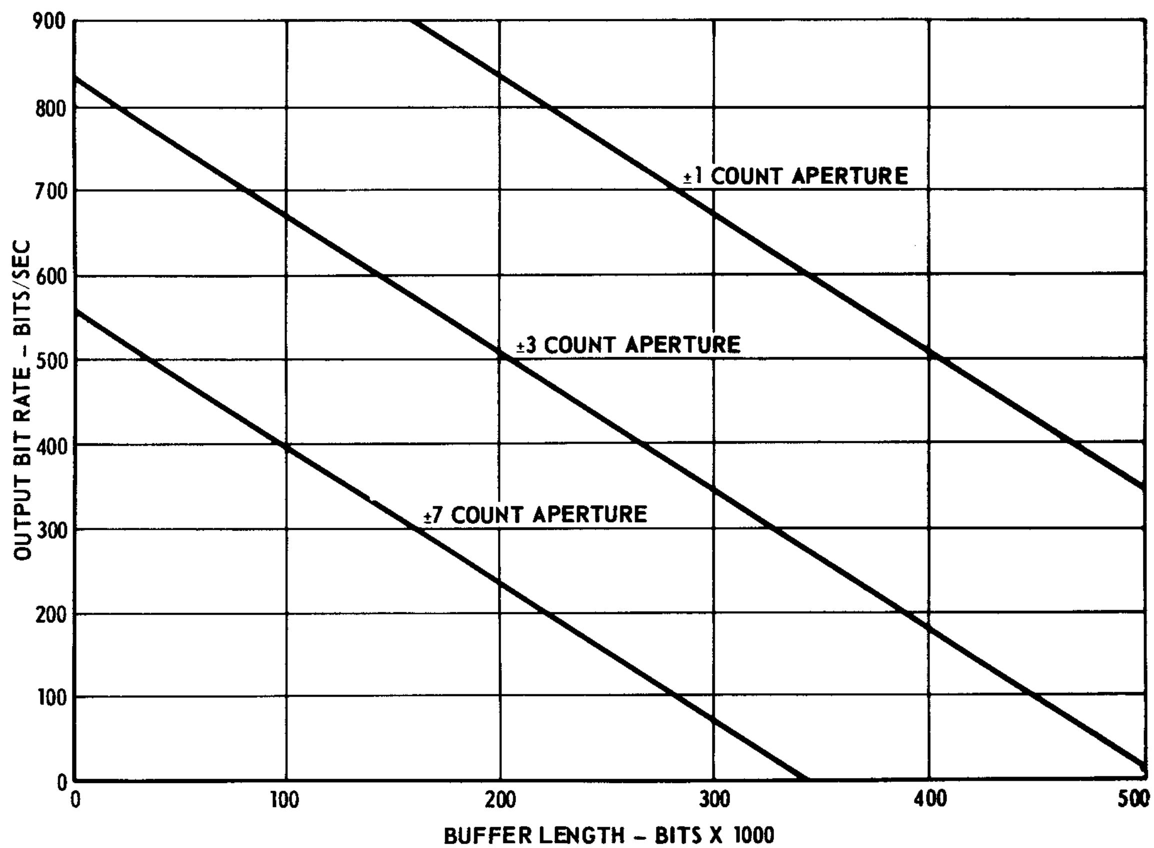

Figure V - Re-entry Period OutT)ut Bit Rate vs Queue Length

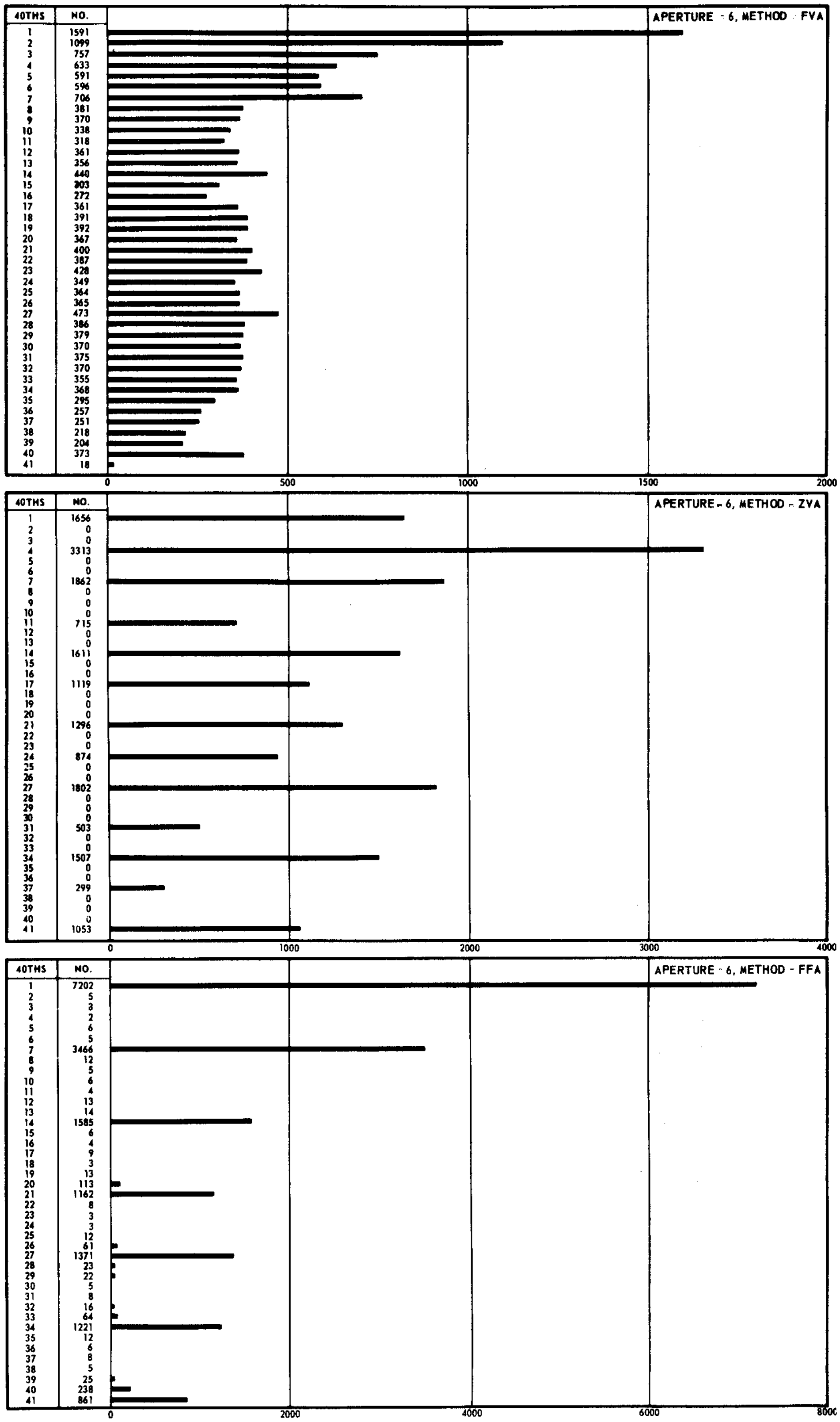

Figure VI - Error Distribution Histogram for ZVA, FFA and FVA Graph of the number of occurances of magnitudes of error in 40th's of aperture

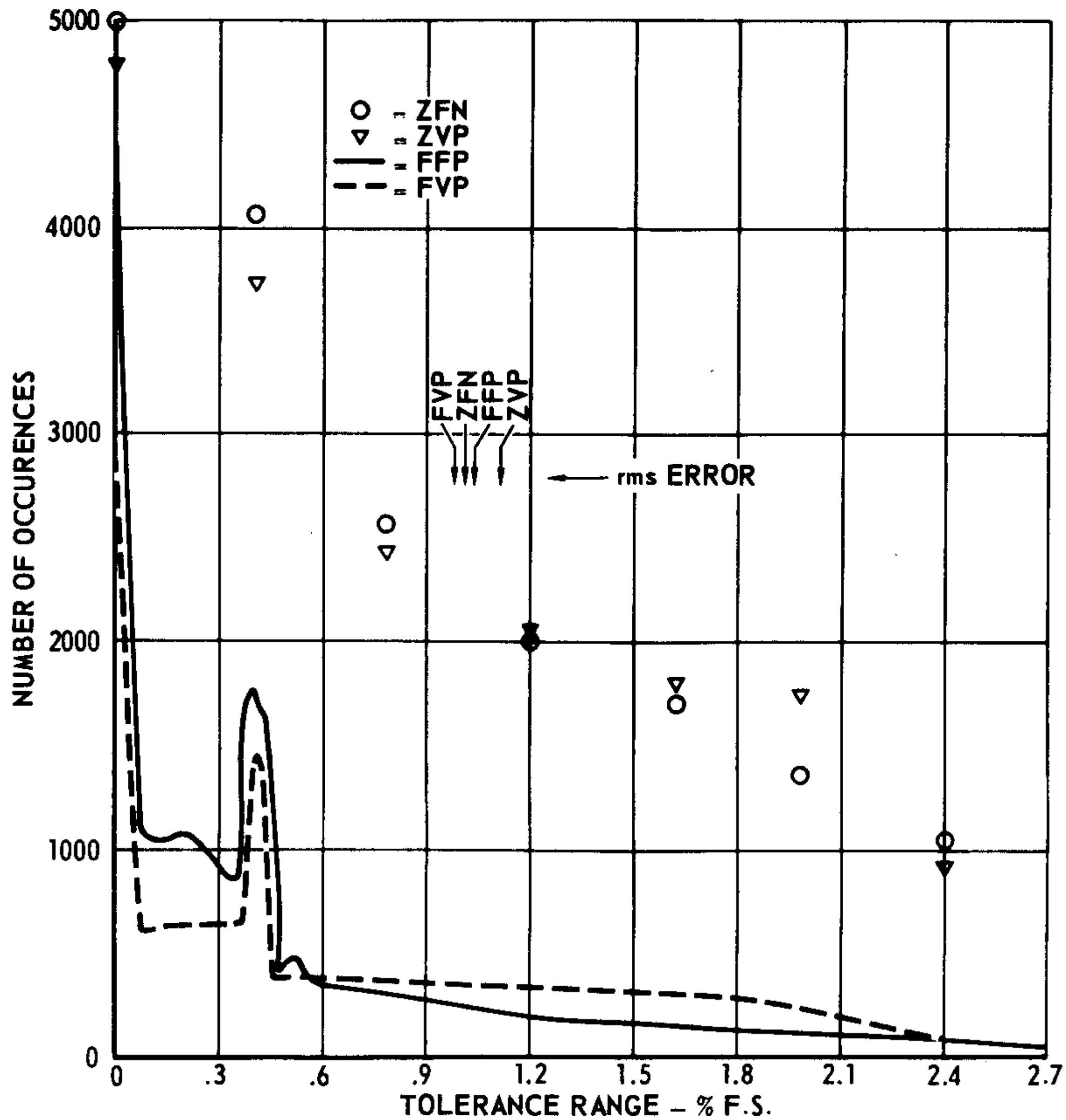

Figure VII - Error Distribution for ZFN, ZVP, FFP and FVP