Influence of Data Processing on the Design and Communication of Experiments*

Solomon W. Golomb

Contribution From the University of Southern California, Los Angeles, Calif.

It is possible to define the relative significance of raw data bits in terms of the influence which they exert on the final processed information. In particular, if the data reduction program is specified in advance, then the experimental design and the communication system can be designed for optimum accumulation of the relevant data. Examples are given, involving nonstandard binary coding of telemetry to minimize the variance of the processed information, in terms of a conceptual deep-space experiment. This paper also considers the effect of successive histogramming as a means of data reduction.

1. Introduction

A communications system is essentially antisymmetric about the channel. That is, as one designs the portions of the system at the receiver terminus based on the channel statistics and appropriate engineering considerations, the corresponding portions at the transmitter terminus are necessarily their functional inverses, in reverse order. Thus, the modulation must be demodulatable, the coding must be decodable, the multiplexing must be unravelable, etc. This duality extends outward to signal source as the conceptual inverse of signal destination, and signal preparation as the functional inverse of signal processing. The elusive and seemingly metaphysical notion of relative significance of information bits becomes a precise mathematical concept when determined by the influence of these bits on the ultimate processed data which reaches the user. That is, if one is forthright enough to specify the data reduction techniques which will ultimately be used, it becomes simply an exercise in numerical analysis to determine the relative importance of bits to be transmitted. The concept of *bit significance* furnishes an evaluation criterion for signal preparation schemes (methods of on-board "preprocessing" of the raw-data prior to transmission), and given a criterion, one may look for an optimum.

In the case of deep-space communication, it is important to distinguish signal preparation for the purpose of protecting information bits generally against the distortions of channel noise, from signal preparation for the purpose of weighting the information bits in accordance with their relative significance. Operationally, we may regard "deep space" as the region from which it is easier to add an extra computer on the ground for data reduction then to add an extra power cell on board to increase the signal strength (and hence, the channel capacity). If one entrusts to the experimenter the specification of format for the data bits to be sent, the communicator's job is quite easy. Blocks of these bits are encoded into orthogonal (or "transorthogonal") waveforms of the maximum duration over which coherent detection can be maintained, transmitted over the channel, and then these waveforms are decoded at the receiver by correlation detection. However, it is likely that for every 1 dB improvement available by these methods, there is the possibility of a 5 or 10 dB improvement based on examining the relevance of the raw data bits to the ultimate reduced data.

As an archetypical problem, one may consider the following: our space probe on Mars has obtained a Martian penny, and we on Earth would like to know the probability p with which it lands "heads." The channel is very noisy. Should the probe transmit fewer samples, well-protected against the channel noise, or should it send as many samples as possible (simply transmitting 1 for heads and 0 for tails) without special noise protection? More generally, if our objective is to determine the mean of the distribution of a remote physical phenomenon with minimum variance of the sample mean, should we send fewer samples more accurately or more samples less accurately? The answer, in general, depends on the signal-to-noise conditions, and the state of a priori knowledge concerning the distribution. In particular, if Mars coins are expected to be rather honest ($p \approx \frac{1}{2}$), and if the channel noise is Gaussian, "more samples" is a better strategy than "more protection." However, if all we intended to do with the samples was average them, how much better than either of the two strategies mentioned it would be to sample at the fastest possible rate, average the samples prior to transmission, and send only this average with as much protection (redundancy) as possible!

2. Numerical Analysis and the Communication System

It is generally an oversimplification to believe that data appears in the form of ideal mathematical "bits." Suppose, for example, that an experiment measures the *intensity* of some phenomenon, with 32 levels of quantization. It is customary to assign the binary numbers from 00000 to 11111 as the "code words" for these quantization levels. If all 32 levels are equally likely, and successive samples are independent, then in the mathematical sense, at least, all five binary symbols in the codeword convey full *bits* of information. Yet, with the usual binary numbering system, an error in the first bit of the codeword is sixteen times as big as an error in the last bit. In this sense, the notion of "significant figures" (or "significant bits") is an old one in mathematics.

In examining this concept more closely, we see that it necessarily relates to assumptions about the future use (processing) of the data. If some sort of arithmetic average of the sample values is to be computed, then the usual idea of significant bits is appropriate. However, there are phenomena for which the most interesting question might be whether the sample value is *even* or *odd*. (For example, this could be the case when counting *events* in certain quantum-mechanical situations.) In such a context, the *last* bit would be the *only* significant one.

In general, then, it is the data processing routine which determines the relative significance of incoming data bits, and this can be measured quantitatively in terms of the size of the error in the final processed output due to an error in a particular data bit.

From the viewpoint of Information Theory, it is easy to reconcile the fact that not all "true bits" of information have the same *significance*. Specifically, *data reduction* generally involves *information destruction*, and only *part* of the information in each bit is utilized. When arithmetic averages are taken, one part of the information in the bits is used; when values are observed to be even or odd, another part is used.

This fact has obvious implications for the design of spacecraft experiments. On the one hand, if some of the data processing can take place on board the spacecraft prior to transmission, there will be considerably fewer information bits requiring transmission. On the other hand, if only the reduced data are sent, it will be impossible to arrive at various conclusions inherent in the raw data, but not specifically sought for by the data processing routine. To see this conflict in its proper perspective, it should be pointed out that it is in fact extremely rare that a spacecraft experimenter processes his data in ways other than he had originally intended. The resulting moral dilemma is: Is it worth the extra channel capacity to send the raw data in order to leave the experimenter with an option he is almost certain not to exercise?

3. Nonstandard Coding for Telemetry Data

The conventional assignment of the binary n-tuples from $00 \ldots 0$ to $11 \ldots 1$ for the numbers from 0 to $2^n - 1$ is of course somewhat arbitrary. Of course, it is systematic, fairly easily implemented, and universally familiar. But none of these reasons would indicate that is the best assignment for transmitting quantization levels from a spacecraft experiment.

One well-known family of nonstandard binary codes are the "Gray Codes," with the property that between consecutive integers, only a single bit of the codeword changes. This has certain switching advantages in the mechanization of binary counters. Specifically, no allowance need be made for the propagation time required for "carry" bits. Thus, for switching purposes, if the numerical values are close (only one apart in numerical distance) then their codewords should be close (only one apart in "Hamming distance"). For telemetry purposes, the emphasis should be reversed. That is if the codewords are close (only one apart in Hamming distance), then the corresponding numerical values should be close (as close in numerical distance as possible). In other words, if a *single error* occurs in the transmission of a data word, its effect, on the average, should be minimized. Since the number 20 has only two immediate numerical neighbors (19 and 21), while the codeword 10100 has *five* immediate Hamming neighbors (10101, 10110, 10000, 11100, and 00100), it is impossible to assign codewords in such a way that immediate Hamming neighbors are also immediate numerical neighbors.

The following rather surprising theorem was conjectured by the author and proved by Mr. Larry Harper.[1]

THEOREM. *Consider any assignment of* n-*bit binary codewords to the numbers from 0 to* 2^n-1, *and add up the absolute value of the numerical error produced by every possible single error in every possible codeword. The minimum possible value for this total is* $2^n(2^n-1)$, *which is attained by the standard binary coding, as well as various nonstandard codes.*

Thus, relative to a "mean absolute first power" error criterion, it is not possible to improve on the standard binary numbering system! However, this theorem ceases to be valid if the first power criterion is replaced by any higher-power criterion. In particular, in the rather common situation that the appropriate criterion is a mean-square-error one, it *is* possible to improve on the standard binary numbering system. Table 1 lists the ordinary binary code, a Gray code, and a minimum-mean-square-error code, for the case $n=5$. The minimum-mean-square-error code illustrates the important fact that "uncoded codes" (codes which add no redundancy) are capable of improving communications performance, because of the phenomenon of "bit significance."

[1] L. H. Harper, Optical assignments of numbers to vertices, J. Soc. Ind. and Appl. Math. **12**, No. 1, 131–135 (March 1964). *See also* J. H. Lindsey, II, Assignment of numbers to vertices, Am. Math. Monthly, **71**, No. 5, 508–516 (May 1964).

TABLE 1. *Numerical binary codes*

Standard Binary code	Gray code	Min. square error code
00000	00000	00000
00001	00001	00001
00010	00011	00010
00011	00010	00100
00100	00110	01000
00101	00111	10000
00110	01111	10001
00111	01110	10010
01000	01010	10100
01001	01011	11000
01010	11011	01001
01011	11111	01010
01100	11101	01100
01101	11100	00101
01110	11110	00110
01111	10110	00011
10000	10111	00111
10001	10101	01011
10010	10001	10011
10011	10011	10101
10100	10010	01101
10101	11010	11001
10110	11000	11010
10111	11001	11100
11000	01001	10110
11001	01000	01110
11010	01100	01111
11011	01101	10111
11100	00101	11011
11101	00100	11101
11110	10100	11110
11111	10000	11111

4. Histograms

A standard method of data reduction is the use of *histograms*. From the raw data, the histogram indicates what sample values occurred and with what frequencies, but it destroys the information concerning the sequential order in which the values occurred. To estimate the *rate* of data reduction effected by taking histograms, we may iterate the histogramming process until we have reduced the data to nothing. For finite data samples, the rate of reduction is found to be exponential. However, as a mathematical curiosity, we can exhibit an infinite data sample (*i.e.*, a function $f(n)$ defined for $n = 1,2,3,4, \ldots$), which is its own histogram, as in table 2. The rule whereby $f(n)$ is constructed is as follows: We set $f(1)=1$. If $f(n)$ is to be its own histogram, then it must take on the value "1" exactly once—which means that $f(n) \neq 1$ for $n > 1$. We now set $f(2)$ equal to the smallest available positive integer—thus, $f(2)=2$. By the self-histogramming property, $f(n)$ must now take on the value "2" a total of *twice*, so we also set $f(3)=2$. This then requires that $f(n)$ also assume the value "3" exactly *twice*, so we set $f(4)=f(5)=3$. This in turn requires that the values "4" and "5" be assumed three times each, so we set $f(6)=f(7)=f(8)=4$ and $f(9)=f(10)=f(11)=5$. Then the values "6," "7," and "8" must each occur *four* times, while the values "9," "10," and "11" must each occur *five* times, and the table continues to generate itself. Strictly speaking, this function is only *uniquely* specified if we require $f(1)=1$, $f(2)=2$, and that $f(n)$ be monotonic nondecreasing. If we define $f^{-1}(n)$ to be the *smallest* integer m such that $f(m)=n$, then we have the curious identity $f(n)+f^{-1}(n)=f^{-1}(n+1)$.

The finite truncations of the function $f(n)$ correspond to finite data samples for which the rate of convergence of iterated histogramming is slowest. The reader is invited to truncate table 2 after $n=23$, and observe the effect of repeated histogramming.

TABLE 2. *A self-histogramming function*

n	$f(n)$	n	$f(n)$	n	$f(n)$
1	1	13	6	25	9
2	2	14	6	26	9
3	2	15	6	27	9
4	3	16	7	28	9
5	3	17	7	29	10
6	4	18	7	30	10
7	4	19	7	31	10
8	4	20	8	32	10
9	5	21	8	33	10
10	5	22	8	34	11
11	5	23	8	25	11
12	6	24	9	36	11 etc.

5. Parametrization of Experiments

It can be argued cogently that space probe experiments should *not* be restricted to measuring phenomena which deviate only slightly from their earth-based a priori values. The real payoff, according to this reasoning, occurs when the truly unexpected is observed. From this viewpoint, a data communication and processing system incapable of handling the "5σ" events is like a life insurance policy which remains in force at all times except in the "highly unlikely event" that something fatal befalls the insured, or a gambling game that pays off except on big bets. On the other hand, one cannot put all one's resources into the long shots. The ideal is to transduce and preprocess the data in such a way that a priori improbable events can be observed and reported, without sacrificing efficiency in the communication of more prosaic data.

Several "obvious" steps in this direction have gradually been incorporated into the standard body of space technology. One procedure is to obtain an initial reading, from a sensor with as wide a dynamic range as possible, and transmit this value; then communicate only the departures from this value, with a new initial fix derived at infrequent intervals. A closely related method is to transmit only the *first differences* of the sequence of sample values after the initial reading has been communicated.

When Explorer I was launched, in January 1958, the cosmic ray intensities far exceeded their anticipated values leading to saturation of the sensing tubes, which gave false readings of "zero." In that case, the "solution" was to include sensing tubes in the subsequent Explorers which were better calibrated for the phenomenon at hand.

It is reasonable to contend that the strategy of experimental design should be different for the *first of a series*, for a *one-of-a-kind shot*, and for a *follow-up shot*. That is, the *first of a series* should get *order-of-magnitude* impressions, and bring in data indicating where the really interesting (and unexpected) results may lie. These indications are then explored to greater precision in the follow-up shots. The one-of-a-kind craft is hardest to design.

Even the highly successful Mariner II disappointed those who hoped to see some totally unexpected phenomenon or measurement established. On the other hand, had such an event occurred, it would quite possibly have required another space probe for confirmation and accurate interpretation.

An interesting approach to the communication of space experiments is to determine the statistical distribution of the data points on board the space-craft, and to transmit the relevant parameters of this distribution. One set of parameters which may be used is the mean, the variance, and the higher moments. (If the phenomenon is gaussian, it is already specified by the mean and variance of the distribution.) Another family of statistical parameters, which are often more useful than the *moments*, are the *quantiles*, a generic name for the median, the quartiles, the percentiles, etc., of the distribution. (For example, the *first quartile* is a numerical value such that 25 percent of the sample values are *larger* while the remaining 75 percent of the sample values are smaller.) It is also possible to compute and transmit statistical parameters which indicate the degree of dependence between successive sample values. These statistical parametrization techniques make it possible to transmit all the information which is normally significant (i.e., which is required for the usual data reduction routines) at a small fraction of the capacity needed to communicate each individual sample point.

Typically, one spacecraft contains numerous scientific experiments, as well as many devices to monitor the engineering performance of the space-craft; and all these data compete for allocation on a common commutated telemetry link. Carrying the notion of relative significance of bits to its logical conclusion, only those measurements exceeding a certain threshold level of unexpectedness should be allowed over the link, while all the prosaic results go unreported.

6. Data Reduction Limitations

For the Gaussian space channel, the ideal encoding for the purpose of combating channel noise makes use of a large family of waveforms with a high degree of mutual uncorrelation. (Examples include orthogonal waveforms, biorthogonal waveforms, and simplex, or transorthogonal, waveforms.) The optimum detection scheme consists of a matched filter correlation detector for each of the possible transmitted waveforms; and at the receiver, the incoming signal is compared (by correlation) with each of the possible waveshapes which might have been transmitted.

Ideally, such a telemetry system should make use of 2^{10} or more different waveforms. However, the problem of constructing so many correlation detectors is quite formidable. Even with the economics of space communication (where an extra computer on the ground is usually cheaper than an extra fuel cell in the spacecraft) the temptation is to back off in one of several directions. For example, if only 2^5 waveforms are used, the processing becomes more tractable, but much of the potential savings in channel capacity is lost. Also, decoding can be performed on a bit-by-bit basis (using the waveforms as error-correcting codes), although such methods are often inferior to no coding at all. Somewhat better than this, accurate correlation may be performed on the incoming waveform one segment at a time with a sequential decoding algorithm used to aline incoming waveforms with their corresponding "most likely" codewords.

There is a clear-cut instance here where the inadequacy of readily available *ground equipment* makes us back off from the optimal communication techniques. (The onboard *encoding* equipment remains remarkably simple in any case.) However, it may not be very long before improved computer components allow the construction of special purpose telemetry processors with a multiplicity of parallel operations, thereby allowing the telemetry correlation and decision process to perform in real time.

This problem of optimal signal processing is probably the most important instance of computer processing techniques as a restraining influence on the design of optimum space communications, but there are other such limitations as well. For example, the ability of computers to extract *pattern* information from pictorial data is still quite limited.

7. Outlook for the Future

There has been much talk about the spacecraft-borne robot which surveys the extraterrestrial situation, digests all the salient features, decides what aspects are most important, determines what further experiments to perform, and communicates his findings back to earth in an optimally encoded manner of his own choosing. Since this type of speculation began several years ago, I have seen no real progress whatever towards its realization, and in my judgment, we can safely forget about it for the next decade or so of space exploration.

I do not wish to seem too skeptical on the subject of pattern recognition and adaptive systems and machine learning. However, I have a deep respect for the difficulty inherent in these problems, and expect progress to be somewhat labored. We will have to learn to recognize patterns with the large computer systems available on earth before we can hope to do so in the relatively tiny systems capable of being space-borne. As for learning by machines, I believe we must teach them all we can as a foundation for whatever subsequent learning they may be capable of on their own. It is certainly easier to build a machine which chooses between alternatives anticipated by its designer than one which can make "intelligent" choices in situations never previously envisioned. It will be a big breakthrough indeed when there is a general-purpose program for the extraction of patterns and significant information from the raw data received on earth from spacecraft transmitters. When that has been achieved, it may be time to worry about more ambitious objectives.