# The Equivalence of Certain Harper Codes

By MORGAN M. BUCHNER, Jr.

(Manuscript received May 12, 1969)

*A class of binary encoding algorithms called Harper codes has been studied previously as a means of encoding numbers for transmission over an idealized binary channel. This paper considers a more general and practical transmission system model. For any Harper code, it presents a technique for obtaining the expression for the average absolute numerical error that occurs during transmission. It shows that all Harper codes do not exhibit the same average absolute numerical error for all transmission systems that satisfy the model. However, there is a subset of Harper codes such that all codes in the subset give identical performance. The paper defines the subset and presents an expression for the average absolute numerical error for any Harper code in the subset. The subset is important because it includes the natural binary representation, the Gray code, and the folded binary code.*

## I. INTRODUCTION

In order to send numerical data over a binary channel, each input number must be encoded into a suitable binary sequence for transmission. For example, when a sampler and quantizer are used, a binary sequence is assigned to each quantization level. For each sample, the number of the appropriate quantization level is transmitted by sending the binary sequence assigned to the level. But how should the binary sequences be assigned? One approach is to use the natural binary representation of each number. Alternatively, a Gray code might be used with the idea that its unit-distance properties are in some sense desirable.

If the transmission system is error-free and if the binary sequences are unique, it does not matter how the sequences are assigned. However, if transmission errors can occur, some assignment algorithms may be preferable to others. In this paper, the performance of certain binary encoding algorithms is considered. The average magnitude by which the number delivered to the destination differs from the transmitted number is used as the criterion of performance.

Previously, Harper presented a class of binary codes that we call Harper codes.[1] The class includes the natural binary representation, the Gray code, and the folded binary code. Reference 2 showed that for any set of $2^k$ input numbers all Harper codes exhibit the same mean magnitude error when used with a specific binary transmission system model (see Section II) and that, when the probability of transmission error is sufficiently small, Harper codes are optimum.

In this paper, a more general transmission system model is considered. For $2^k$ equally spaced input numbers, a means of obtaining the expression for the average absolute numerical error (hereafter called average numerical error) for any Harper code is presented. All Harper codes do not exhibit the same average numerical error except in the special case when the transmission system model reduces to the model used in Ref. 2. However, there does exist a subset of Harper codes such that all codes in the subset are equivalent in performance. The subset is defined and an expression is given for the average numerical error for any Harper code in the subset. The subset is important because it includes the natural binary representation, the Gray code, and the folded binary code.

## II. SYSTEM MODEL AND PREVIOUS RESULTS

A system model is shown in Fig. 1. In general, we wish to send over a binary transmission system† any one of the $2^k$ equally likely numbers of the form $A + Bs$ where $s$ is an integer, $0 \leqq s \leqq 2^k - 1$. At the transmitter, the binary encoder receives $A + Bs$ and, based upon $s$, sends a $k$-bit binary sequence assigned by a Harper code and denoted by $H_k(s)$. At the receiver, a binary decoder receives a $k$-bit binary sequence $H_k(r)$, $0 \leqq r \leqq 2^k - 1$, and generates $A + Br$. Let $\Pr[H_k(r) \mid H_k(s)]$ denote the probability of receiving $H_k(r)$ when $H_k(s)$ is sent. If all $s$ are equally likely, the average numerical error (as in Ref. 3) that occurs is

$$ANE = \frac{B}{2^k} \sum_{r=0}^{2^k-1} \sum_{s=0}^{2^k-1} | r - s | \Pr[H_k(r) \mid H_k(s)]. \tag{1}$$

The average numerical error is dependent upon the binary encoding algorithm and the transmission system through $\Pr[H_k(r) \mid H_k(s)]$.

† It is important to distinguish between the binary transmission system and the channel. The transmission system includes the channel and the encoder and decoder for error control.

Harper codes are defined in terms of the vertices of the $k$-cube[1]. Assign 0 to an arbitrary vertex; that is, $H_k(0)$ is arbitrary. Having assigned $0, 1, 2, \cdots, l-1$, assign $l$ to an unnumbered vertex (not necessarily unique) that has the most numbered one-distant neighbors.† In the remainder of this paper, certain properties of Harper codes presented in Refs. 1 and 2 are used without specific reference.

We can now summarize the results in Ref. 2. In a binary transmission system as shown in Fig. 1, it was assumed that the errors between

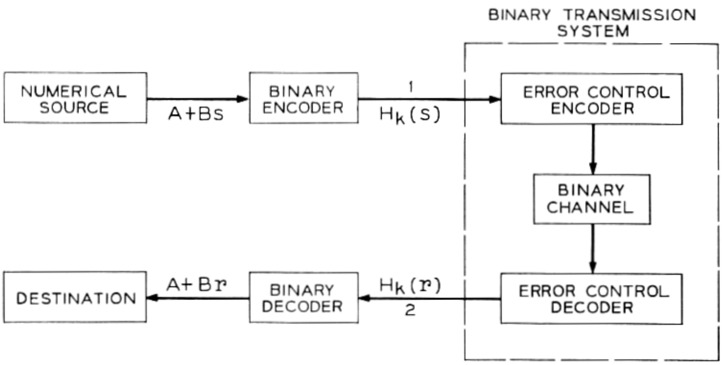


Fig. 1 — System model

locations 1 and 2 are independent of the transmitted bits and occur independently of one another with probability $p_1$. For such a transmission system and for any set of $2^k$ input numbers, it was shown that all Harper codes yield the same mean magnitude error and, thus, are equivalent. Also, it was shown that when $p_1$ is sufficiently small, Harper codes are optimum for any set of $2^k$ input numbers because they minimize the mean magnitude error. Of course, the results in Ref. 2 are applicable to our set of $2^k$ equally spaced numbers and indicate that all Harper codes yield the same average numerical error for a transmission system that satisfies the model in Ref. 2.

However, the transmission system model in Ref. 2 is extremely restrictive. Channels with correlated errors are excluded. The model also excludes transmission systems using many types of error-correcting codes even if the actual channel is a memoryless binary symmetric channel with probability of bit error $p$. In fact, even the Hamming

† The weight of an $n$-tuple $v$ is the number of nonzero components in $v$ and is denoted by $w[v]$. The distance between two binary $n$-tuples $u$ and $v$ is $w[u \oplus v]$ where $\oplus$ denotes component by component modulo 2 addition of $n$-tuples.

perfect single error-correcting codes when used with a memoryless binary symmetric channel do not comply with the model in Ref. 2. The reason is that, in a Hamming code, all $H_k(s)$ of a particular weight are not encoded as code vectors of equal weight. Thus, all error patterns of equal weight in the Harper code sequences do not occur with equal probability. However, in order for a transmission system to satisfy the model in Ref. 2, all error patterns of equal weight must occur with equal probability. It follows that the Hamming code violates the model in Ref. 2.

An interesting approach to coding for numerical data transmission is found in unequal error-protection codes[4]. The idea behind unequal error-protection codes is to match the protection provided by the code to the numerical significance of the transmitted bits. Significant-bit codes (a subclass of unequal error-protection codes) have been shown to be effective in reducing the average numerical error and in many cases are easy to implement.[3,5] However, the transmission system model in Ref. 2 excludes unequal error-protection codes which is unfortunate because these codes are directly applicable to the basic problem considered in Ref. 2, that is, reducing the average numerical error.

Accordingly, it is important to examine the performance of Harper codes when a less restrictive and more practical transmission system model is used. For our model, we assume simply that the transmission system is binary and that the errors are independent of the transmitted bits. A binary transmission system satisfies this model if, for every integer $r$, $0 \leqq r \leqq 2^k - 1$, and integer $s$, $0 \leqq s \leqq 2^k - 1$, there exists an integer $t$, $0 \leqq t \leqq 2^k - 1$, such that

$$\Pr[H_k(r) \mid H_k(s)] = \Pr[H_k(t) \mid B_k(0)] \tag{2}$$

where

$$H_k(t) = H_k(r) \oplus H_k(s) \tag{3}$$

and $B_i(j)$ denotes the $i$-bit natural binary representation of the integer $j$, $0 \leqq j \leqq 2^i - 1$. Observe that equation (2) implies that the probability of a particular error pattern $H_k(t)$ in a Harper code sequence is independent of the transmitted sequence.

Because the transmission system model is not very restrictive, the results to be presented are applicable to a wide range of practical systems. For example, the model is satisfied by the important class of binary transmission systems composed of (*i*) a linear block code with a decoding scheme equivalent to Slepian's standard array[6], and

(*ii*) a binary symmetric channel in which the errors are independent of the transmitted bits.

## III. THE AVERAGE NUMERICAL ERROR FOR A HARPER CODE

Let $H'$ be a Harper code in which $s$ is encoded as $H'_k(s)$. From the definition of a Harper code, it is possible that $H'_k(0) \neq B_k(0)$. We first show that if $H'_k(0) \neq B_k(0)$, then a Harper code $H$ [in which $s$ is encoded as $H_k(s)$] can be constructed such that (*i*) $H_k(0) = B_k(0)$ and, (*ii*) the performance of $H'$ is identical to the performance of $H$. The average numerical error for $H'$ is

$$ANE' = \frac{B}{2^k} \sum_{r=0}^{2^k-1} \sum_{s=0}^{2^k-1} | r - s | \Pr[H'_k(r) \mid H'_k(s)]. \tag{4}$$

Let $H$ be a code whose elements are obtained from the elements of $H'$ by the relation

$$H_k(s) = H'_k(s) \oplus H'_k(0). \tag{5}$$

From (5), $H_k(0) = B_k(0)$.

We now show that $H$ is a Harper code. Clearly $H_k(0)$ satisfies the requirements for a Harper code. Suppose that $H_k(0)$ through $H_k(l-1)$ have been determined by (5). Now, if $H'_k(s)$ is distance $d$ from $H'_k(l)$, then $H_k(s)$ is distance $d$ from $H_k(l)$. Thus, if $H'_k(l)$ is assigned to have the most numbered one-distant neighbors, $H_k(l)$ will have the most numbered one-distant neighbors. It follows that $H$ is a Harper code.

The average numerical error for $H$ is given by equation (1). We must show that the expression for $ANE$ is identical to the expression for $ANE'$. From (2),

$$\Pr[H'_k(r) \mid H'_k(s)] = \Pr[H'_k(r) \oplus H'_k(s) \mid B_k(0)].$$

Also, from (2),

$$\Pr[H_k(r) \mid H_k(s)] = \Pr[H_k(r) \oplus H_k(s) \mid B_k(0)].$$

From (5),

$$H_k(r) \oplus H_k(s) = H'_k(r) \oplus H'_k(s).$$

Therefore,

$$\Pr[H_k(r) \mid H_k(s)] = \Pr[H'_k(r) \mid H'_k(s)]$$

and, by (1) and (4),

$$ANE = ANE'.$$

Thus, every Harper code is equivalent in performance to a Harper code in which

$$H_k(0) = B_k(0). \tag{6}$$

For convenience and without loss of generality, we shall consider the performance of Harper codes that satisfy (6). At the end of Section IV, we remove this restriction and give, in general terms, the structure of all Harper codes that are equivalent to the natural binary representation.

Now, let us consider the expression for the average numerical error for $H$. By substituting (2) into (1) and rewriting,

$$ANE = \frac{B}{2^k} \sum_{t=1}^{2^k-1} \sum_{s=0}^{2^k-1} | r_t - s | \Pr[H_k(t) \mid B_k(0)] \tag{7}$$

where $r_t$ is the value of $r$ in (3), that is,

$$H_k(r_t) = H_k(s) \oplus H_k(t). \tag{8}$$

Now, (7) can be written as

$$ANE = \frac{B}{2^k} \sum_{t=1}^{2^k-1} C_t \Pr[H_k(t) \mid B_k(0)] \tag{9}$$

where

$$C_t = \sum_{s=0}^{2^k-1} | r_t - s |. \tag{10}$$

The expression for the average numerical error is determined by specifying each $C_t$ $(1 \leqq t \leqq 2^k - 1)$.

In order to evaluate $C_t$, we proceed as follows. Divide the $2^k$ elements of $H$ into $k + 1$ sets called levels. The 0-level contains $H_k(0)$ exclusively. For $1 \leqq j \leqq k$, the $j$-level is the set of $H_k(s)$ for which $2^{j-1} \leqq s \leqq 2^j - 1$. Because $H$ is a Harper code, the elements of level $j$ are in the shadow of the $(j - 1)$-subcube† formed by the elements of levels 0 through $j - 1$. From equation (6) and the definition of a Harper code, it follows that each element of the $j$-level has a one in a particular position which we call the $j$-position. Thus, the $j$-level consists of the $k$-tuples that

† A $(j - 1)$-subcube of the $k$-cube is a set of all $k$-tuples that are the same in $k - j + 1$ positions. The shadow of a $(j - 1)$-subcube is obtained by changing one of the $k - j + 1$ fixed positions.

have zeros in positions $j + 1$ through $k$, a one in position $j$, and all possible $(j - 1)$-tuples in positions 1 through $j - 1$.

Notice that the position numbers are determined by the structure of the Harper code and not by the order in which the bits are arranged for transmission. For example, in the Harper code shown in Table I, $\Pr[H_4(2) \mid B_4(0)]$ is the probability that no transmission errors occur in positions 1, 3, and 4 and that a transmission error occurs in position 2. If transmitted in the order shown in Table I, $\Pr[H_4(2) \mid B_4(0)]$ is the probability that the error sequence 0001 occurs.

We must determine $C_t$ for each of the $2^k - 1$ nonzero values of $t$. Thus, we regard $t$ as known and seek to determine $C_t$. Let $\sigma$ be an integer such that

$$2^{\sigma-1} \leqq t \leqq 2^{\sigma} - 1. \tag{11}$$

Because $H$ satisfies equation (6), $H_k(t)$ has a one in position $\sigma$. To evaluate $C_t$, we rewrite (10) to exhibit the levels of $s$ as

$$C_t = \left( r_t + \sum_{j=1}^{\sigma} \sum_{s=2^{j-1}}^{2^j-1} | r_t - s | \right) + \sum_{j=\sigma+1}^{k} \sum_{s=2^{j-1}}^{2^j-1} | r_t - s | \tag{12}$$

TABLE I—A $k = 4$ HARPER CODE

| $s$ | $H_4(s)$ | Level number |
|---|---|---|
| 0 | 0 0 0 0 | 0 |
| 1 | 0 0 1 0 | 1 |
| 2 | 0 0 0 1 | 2 |
| 3 | 0 0 1 1 | 2 |
| 4 | 0 1 1 1 | 3 |
| 5 | 0 1 1 0 | 3 |
| 6 | 0 1 0 1 | 3 |
| 7 | 0 1 0 0 | 3 |
| 8 | 1 0 0 0 | 4 |
| 9 | 1 0 0 1 | 4 |
| 10 | 1 0 1 1 | 4 |
| 11 | 1 0 1 0 | 4 |
| 12 | 1 1 0 0 | 4 |
| 13 | 1 1 1 0 | 4 |
| 14 | 1 1 0 1 | 4 |
| 15 | 1 1 1 1 | 4 |

position 4 position 2

position 3 position 1

where the 0-level is shown individually as $r_t$ and $j$ indexes the levels from 1 to $k$. The parentheses enclose the contribution of levels 0 through $\sigma$. From Appendix A,

$$\left(r_t + \sum_{j=1}^{\sigma} \sum_{s=2^{j-1}}^{2^j-1} | r_t - s |\right) = 2^{2\sigma-1}. \tag{13}$$

Now, consider the set of $H_k(s)$ in the $j$-level where $\sigma + 1 \leqq j \leqq k$ and $2^{j-1} \leqq s \leqq 2^j - 1$. First, we define a run as follows.† In the $j$-level, there is a run in position $m$, $1 \leqq m \leqq j - 1$, that starts at $s_0$ and is of length $R(m, s_0)$ if and only if

(*i*) $R(m, s_0) = 2^l$ for some integer $l \geqq 0$,
(*ii*) the set of $H_k(s)$ for $s_0 \leqq s \leqq s_0 + 2^l - 1$ forms an $l$-subcube of the $k$-cube where $m$ is one of the $k - l$ fixed positions,
(*iii*) the set of $H_k(s)$ for $s_0 + 2^l \leqq s \leqq s_0 + 2^{l+1} - 1$ forms an $l$-subcube that is in the shadow of the subcube in (*ii*),
(*iv*) the subcube in (*iii*) is distinguished from the subcube in (*ii*) by position $m$, and
(*v*) the $H_k(s)$ for $2^{j-1} \leqq s \leqq s_0 - 1$ can be divided into runs of length $2^l$ although perhaps not in position $m$.

An example from Table I will illustrate the definition of a run. Consider the 4-level. Then $H_4(8)$ starts a run of length 1 in position 2, a run of length 2 in position 1, and a run of length 4 in position 3. Thus,

$$R(1, 8) = 2 \qquad R(2, 8) = 1 \qquad R(3, 8) = 4.$$

Let $w[H_k(t)] = \omega$ and let $t_1, t_2, \cdots, t_\omega$ denote the $\omega$ nonzero positions in $H_k(t)$. Then $R(t_m, 2^{j-1})$ is the length of the run in position $t_m$ that starts at $2^{j-1}$ (that is, the length of the first run in position $t_m$ in the $j$-level). Let

$$\gamma_{j,1}(t) = \operatorname*{Max}_{m} R(t_m, 2^{j-1}).$$

From Appendix C,

$$\sum_{s=2^{j-1}}^{2^{j-1}+2\gamma_{j,1}(t)-1} | r_t - s |$$
$$= \sum_{s=2^{j-1}}^{2^{j-1}+\gamma_{j,1}(t)-1} (r_t - s) + \sum_{s=2^{j-1}+\gamma_{j,1}(t)}^{2^{j-1}+2\gamma_{j,1}(t)-1} (s - r_t) = 2\gamma_{j,1}^2(t).$$

† Appendix B contains a more complete discussion of the structure of the $j$-level of a Harper code and the relationship between the structure and the concept of a run. It is shown that runs are basic to the structure of Harper codes and that the definition of a run is meaningful and consistent.

The above process can be extended to obtain $\gamma_{j,i}(t)$ after $\gamma_{j,1}(t)$, $\gamma_{j,2}(t), \cdots, \gamma_{j,i-1}(t)$ are known. Specifically,

$$\gamma_{j,i}(t) = \operatorname*{Max}_{m} R\left(t_m, 2^{j-1} + 2 \sum_{l=1}^{i-1} \gamma_{j,l}(t)\right).$$

Then

$$\sum_{s=2^{j-1}+2\sum_{l=1}^{i-1}\gamma_{j,l}(t)}^{2^{j-1}-1+2\sum_{l=1}^{i}\gamma_{j,l}(t)} | r_t - s | = 2\gamma_{j,i}^2(t).$$

By continuing the process, we eventually exhaust the $2^{j-1}$ values of $s$ in the $j$-level. Let $g_j$ denote the number of $\gamma_{j,i}(t)$ needed to cover the $j$-level, that is,

$$2 \sum_{i=1}^{g_j} \gamma_{j,i}(t) = 2^{j-1}.$$

It follows that

$$\sum_{s=2^{j-1}}^{2^j-1} | r_t - s | = 2 \sum_{i=1}^{g_j} \gamma_{j,i}^2(t). \tag{14}$$

From (12), (13), and (14),

$$C_t = 2^{2\sigma-1} + 2 \sum_{j=\sigma+1}^{k} \sum_{i=1}^{g_j} \gamma_{j,i}^2(t). \tag{15}$$

By substituting (15) into (9),

$$ANE = \frac{B}{2^k} \sum_{t=1}^{2^k-1} \left(2^{2\sigma-1} + 2 \sum_{j=\sigma+1}^{k} \sum_{i=1}^{g_j} \gamma_{j,i}^2(t)\right) \Pr[H_k(t) \mid B_k(0)] \tag{16}$$

where $\sigma$ is given by (11). The expression in (16) is particularly useful because it consists exclusively of error probabilities conditional upon $B_k(0)$ being transmitted and the $\gamma_{j,i}(t)$ can be obtained directly from the Harper code. A numerical example that illustrates the use of (15) and (16) is given in Appendix D.

We now consider the condition under which two Harper codes give identical performance. Let $H'$ be a Harper code that is not $H$ (that is, $H'$ cannot be obtained from $H$ by a relationship of the form $H'_k(s) = H_k(s) \oplus B_k(s_1)$ where $s_1$ is an arbitrary integer, $0 \leqq s_1 \leqq 2^k - 1$).

From (9), for $H'$,

$$ANE' = \frac{B}{2^k} \sum_{t'=1}^{2^k-1} C'_{t'} \Pr[H'_k(t') \mid B_k(0)].$$

Then $H$ and $H'$ exhibit identical performance for any transmission system that satisfies our model only if, for every $t$, $C'_{t'} = C_t$ where $t'$ is determined by $H'_k(t') = H_k(t)$. Conversely, if $C'_{t'} \neq C_t$ for at least one value of $t$, the two codes may or may not give the same performance, depending upon the error statistics of the transmission system.

## IV. CODES EQUIVALENT TO THE NATURAL BINARY REPRESENTATION

Because of the considerable structure in the natural binary representation, it is easy to use (15) to compute each $C_t$, $1 \leqq t \leqq 2^k - 1$. For a given $t$, we first find $\sigma$ by (11), that is, $\sigma - 1$ is the largest power of 2 in $t$. Then, for each $j$, $\sigma + 1 \leqq j \leqq k$, we determine $g_j$ and the $\gamma_{j,i}(t)$. For the natural binary representation,

$$\gamma_{j,i}(t) = 2^{\sigma-1} \tag{17}$$

for $1 \leqq i \leqq g_j$ so $g_j = 2^{j-\sigma-1}$. Therefore, by (15) and (17),

$$C_t = 2^{2\sigma-1} + 2 \sum_{j=\sigma+1}^{k} \sum_{i=1}^{2^{j-\sigma-1}} 2^{2\sigma-2} = 2^{k+\sigma-1}. \tag{18}$$

Notice that each $C_t$, $2^{\sigma-1} \leqq t \leqq 2^\sigma - 1$, is equal to $2^{k+\sigma-1}$. Thus, $C_t$ is determined simply by the largest power of 2 in $t$. Substituting (18) into (9) and rewriting, we obtain

$$ANE_B = B \sum_{\sigma=1}^{k} 2^{\sigma-1} \sum_{t=2^{\sigma-1}}^{2^\sigma-1} \Pr[B_k(t) \mid B_k(0)] \tag{19}$$

where $ANE_B$ denotes the average numerical error for the natural binary representation.

Is it possible to find a Harper code $H$ that is not the natural binary representation but that exhibits performance that is identical to the natural binary representation for all transmission systems that satisfy our model? The answer is yes. We now show that a necessary and sufficient condition is that

$$\gamma_{j,i}(t) = 2^{\sigma-1} \tag{20a}$$

for $1 \leqq i \leqq g_j$ and

$$g_j = 2^{j-\sigma-1} \tag{20b}$$

for each $t$, $1 \leqq t \leqq 2^k - 1$, and for each $j$, $\sigma + 1 \leqq j \leqq k$ (where $\sigma$ is chosen so that $2^{\sigma-1} \leqq t \leqq 2^\sigma - 1$).

If (20) is satisfied, then by (15), $C_t = 2^{k+\sigma-1}$. The average numerical error for $H$ (denoted by $ANE_H$) is

$$ANE_H = B \sum_{\sigma=1}^{k} 2^{\sigma-1} \sum_{t=2^{\sigma-1}}^{2^\sigma-1} \Pr[H_k(t) \mid B_k(0)]. \tag{21}$$

By the definition of a Harper code and the definition of a level,

$$\sum_{t=2^{\sigma-1}}^{2^\sigma-1} \Pr[H_k(t) \mid B_k(0)] = \sum_{t=2^{\sigma-1}}^{2^\sigma-1} \Pr[B_k(t) \mid B_k(0)]. \tag{22}$$

Therefore, by (19), (21), and (22), $ANE_B = ANE_H$. It follows that (20) is a sufficient condition.

We now show by contradiction that (20) is a necessary condition. Consider the set of coefficients $C_{2^{\sigma-1}}$ for $1 \leqq \sigma \leqq k$. From (15),

$$C_{2^{\sigma-1}} = 2^{2\sigma-1} + 2 \sum_{j=\sigma+1}^{k} \sum_{i=1}^{\sigma j} \gamma_{j,i}^2(2^{\sigma-1}).$$

The term $2^{2\sigma-1}$ is independent of the particular Harper code used. Thus, we need only consider the summation part. Suppose that it is possible to arrange the $\gamma_{j,i}(2^{\sigma-1})$ so that they are not all equal to $2^{\sigma-1}$ but keep $C_{2^{\sigma-1}} = 2^{k+\sigma-1}$. If this is done, at least one $\gamma_{j,i}(2^{\sigma-1})$ will be less than $2^{\sigma-1}$ and at least one $\gamma_{j,i}(2^{\sigma-1})$ will be greater than $2^{\sigma-1}$. However, in order for one $\gamma_{j,i}(2^{\sigma-1})$ to be less than $2^{\sigma-1}$, there must exist a $\sigma' < \sigma$ such that $\gamma_{j',i'}(2^{\sigma'-1}) > 2^{\sigma'-1}$. But in order for $C_{2^{\sigma'-1}} = 2^{k+\sigma'-1}$, there must be at least one $\gamma_{j',i'}(2^{\sigma'-1}) < 2^{\sigma'-1}$. The argument continues until we reach $\gamma_{j'',i''}(2^0)$ where there must be at least one

$$\gamma_{j'',i''}(2^0) > 2^0. \tag{23}$$

However, in order for $C_{2^0} = 2^k$, (23) implies that there must be at least one $\gamma_{j'',i''}(2^0) < 2^0$, which is impossible. It follows that (20) must hold in order for a Harper code to be equivalent to the natural binary representation.

We can show the existence of a great many Harper codes other than the natural binary representation that satisfy (20) by presenting explicitly the structure implied by (20). At this point, we no longer assume that $H_k(0) = B_k(0)$ but state the structure for any Harper code. List the $H_k(s)$ sequentially as $s$ runs from 0 to $2^k - 1$. For position $i$, $1 \leqq i \leqq k$, divide the $s$ into $2^{k-i+1}$ consecutive intervals each of length $2^{i-1}$. Let $j$ index the intervals where $0 \leqq j \leqq 2^{k-i+1} - 1$.

A Harper code is equivalent to the natural binary representation if and only if, for every odd numbered interval ($j$ odd), the binary digit in position $i$ is the complement of the binary digit in position $i$ in the immediately preceding even numbered interval ($j$ even). The digit in position $i$ in the even numbered intervals is arbitrary.

The structure is presented graphically in Table II for $k = 5$. The symbol $b_{i,j}$ denotes the binary digit in position $i$ in the $j$th interval. For odd $j$, $b_{i,j} = b^*_{i,j-1}$ (where $b^*_{i,j-1} = 1 \oplus b_{i,j-1}$) and, thus, $b^*_{i,j-1}$ is shown in Table II for odd $j$. For all even $j$, $b_{i,j}$ can be assigned arbitrarily for each $i$.

The expression for the average numerical error of the Harper codes that are equivalent to the natural binary representation is interesting. From (21), the set of error probabilities $\Pr[H_k(t) \mid B_k(0)]$ for $2^{\sigma-1} \leqq t \leqq 2^{\sigma} - 1$ (that is, for $t$ in the $\sigma$-level) all have the weighting coefficient $2^{\sigma-1}$. Thus, the cost of a particular error pattern is the numerical significance of the most significant bit in error. When one considers unequal error-protection codes, the structure in (21) is very convenient because the protection against transmission errors can be matched to the significance of the bit positions. However, for a Harper code that is not equivalent to the natural binary representation, the average numerical error does not exhibit the above structure. Therefore, unequal error-protection codes appear to be less applicable.

## V. THE GRAY CODE AND THE FOLDED BINARY CODE

The Gray code and the folded binary code are of interest because of their possible applicability to numerical data transmission.[7,8] This section shows that both of these codes exhibit performance that is identical to the performance of the natural binary representation for all binary transmission systems that satisfy our model.

Let the $k$-bit binary representation of $s$ be $B_k(s) = (b_k, b_{k-1}, \cdots, b_1)$ where $b_i$, $1 \leqq i \leqq k$, is the binary digit in position $i$ and

$$s = \sum_{i=1}^{k} b_i 2^{i-1}.$$

As in Section III, the position numbers are defined in terms of the structure of the code, not the order in which the bits are transmitted. From Ref. 7, the Gray code representation of $s$, denoted by $G_k(s)$, is $G_k(s) = (b_k, b_k \oplus b_{k-1}, \cdots, b_2 \oplus b_1)$. We show that the Gray code is equivalent to the natural binary representation by showing that the structure of the Gray code conforms with the structure in Table II. Consider position $i$. As in the construction of Table II, divide the range for $s$ into consecutive intervals each of length $2^{i-1}$ and number the intervals sequentially from 0 to $2^{k-i+1} - 1$. The binary digit in position $i$ of $G_k(s)$ in an even numbered interval is $b_{i+1} \oplus b_i$ and the

TABLE II—STRUCTURE FOR A HARPER CODE EQUIVALENT TO THE NATURAL BINARY REPRESENTATION; $k = 5$

| | $H_5(s)$ | | | | |
|---|---|---|---|---|---|
| | Position Number | | | | |
| $s$ | 5 | 4 | 3 | 2 | 1 |
| 0 | $b_{5,0}$ | $b_{4,0}$ | $b_{3,0}$ | $b_{2,0}$ | $b_{1,0}$ |
| 1 | | | | ↓ | $b^*_{1,0}$ |
| 2 | | | | $b^*_{2,0}$ | $b_{1,2}$ |
| 3 | | | ↓ | ↓ | $b^*_{1,2}$ |
| 4 | | | $b^*_{3,0}$ | $b_{2,2}$ | $b_{1,4}$ |
| 5 | | | | ↓ | $b^*_{1,4}$ |
| 6 | | | | $b^*_{2,2}$ | $b_{1,6}$ |
| 7 | | ↓ | ↓ | ↓ | $b^*_{1,6}$ |
| 8 | | $b^*_{4,0}$ | $b_{3,2}$ | $b_{2,4}$ | $b_{1,8}$ |
| 9 | | | | ↓ | $b^*_{1,8}$ |
| 10 | | | | $b^*_{2,4}$ | $b_{1,10}$ |
| 11 | | | ↓ | ↓ | $b^*_{1,10}$ |
| 12 | | | $b^*_{3,2}$ | $b_{2,6}$ | $b_{1,12}$ |
| 13 | | | | ↓ | $b^*_{1,12}$ |
| 14 | | | | $b^*_{2,6}$ | $b_{1,14}$ |
| 15 | ↓ | ↓ | ↓ | ↓ | $b^*_{1,14}$ |
| 16 | $b^*_{5,0}$ | $b_{4,2}$ | $b_{3,4}$ | $b_{2,8}$ | $b_{1,16}$ |
| 17 | | | | ↓ | $b^*_{1,16}$ |
| 18 | | | | $b^*_{2,8}$ | $b_{1,18}$ |
| 19 | | | ↓ | ↓ | $b^*_{1,18}$ |
| 20 | | | $b^*_{3,4}$ | $b_{2,10}$ | $b_{1,20}$ |
| 21 | | | | ↓ | $b^*_{1,20}$ |
| 22 | | | | $b^*_{2,10}$ | $b_{1,22}$ |
| 23 | | ↓ | ↓ | ↓ | $b^*_{1,22}$ |
| 24 | | $b^*_{4,2}$ | $b_{3,6}$ | $b_{2,12}$ | $b_{1,24}$ |
| 25 | | | | ↓ | $b^*_{1,24}$ |
| 26 | | | | $b^*_{2,12}$ | $b_{1,26}$ |
| 27 | | | ↓ | ↓ | $b^*_{1,26}$ |
| 28 | | | $b^*_{3,6}$ | $b_{2,14}$ | $b_{1,28}$ |
| 29 | | | | ↓ | $b^*_{1,28}$ |
| 30 | | | | $b^*_{2,14}$ | $b_{1,30}$ |
| 31 | ↓ | ↓ | ↓ | ↓ | $b^*_{1,30}$ |

binary digit in position $i$ in the immediately following odd numbered interval is $b_{i+1} \oplus b_i^* = (b_{i+1} \oplus b_i)^*$. Therefore, from the structure in Table II, the Gray code is equivalent to the natural binary representation.

It is also interesting to consider the folded binary code[8]. Let $F_k(s)$ denote the representation of $s$. Then $F_k(s) = (b_k, b_k^* \oplus b_{k-1}, \cdots, b_k^* \oplus b_1)$ where $b_k^* = b_k \oplus 1$. As in the case of the Gray code, consider position $i$ and divide the range for $s$ into intervals of length $2^{i-1}$. The binary digit in position $i$ of $F_k(s)$ in an even numbered interval is $b_k^* \oplus b_i$. The binary digit in position $i$ in the immediately following odd numbered interval is $b_k^* \oplus b_i^* = (b_k^* \oplus b_i)^*$. Therefore, from the structure in Table II, the folded binary code is equivalent to the natural binary representation.

## VI. CONCLUSIONS

The model used in this paper for the binary transmission system is quite general and is satisfied by a wide range of practical systems including many that utilize error-correcting codes. A technique is presented for determining the average numerical error for any Harper code. All Harper codes do not exhibit equal performance for all transmission systems that satisfy the model. Because the performance of a given Harper code is closely related to the error statistics of the transmission system, it does not appear possible to specify a Harper code that is best for all applications. However, a subset of Harper codes is defined such that all codes in the subset give identical performance for all transmission systems covered by the model. The subset is important because it includes the natural binary representation, the Gray code, and the folded binary code. Unequal error-protection codes appear to be particularly applicable to Harper codes in the subset.

## APPENDIX A

### *Contribution of Levels* 0 *through* $\sigma$ *to* $C_t$

To determine the contribution of levels 0 through $\sigma$ to $C_t$, we must evaluate

$$r_t + \sum_{j=1}^{\sigma} \sum_{s=2^{j-1}}^{2^j-1} | r_t - s | = \sum_{s=0}^{2^\sigma-1} | r_t - s |.$$

From equation (8), for every $s$ in the range $0 \leqq s \leq 2^{\sigma-1} - 1$, there exists a unique $r_t$ in the range $2^{\sigma-1} \leqq r_t \leqq 2^\sigma - 1$. As $s$ runs from 0 through $2^{\sigma-1} - 1$, every $r_t$ in the range $2^{\sigma-1} \leqq r_t \leqq 2^\sigma - 1$ occurs once and only once. Similarly, as $s$ runs from $2^{\sigma-1}$ through $2^\sigma - 1$, every $r_t$ in the range $0 \leqq r_t \leqq 2^{\sigma-1} - 1$ occurs once and only once. Accordingly,

$$\sum_{s=0}^{2^\sigma-1} | r_t - s | = \sum_{s=0}^{2^{\sigma-1}-1} (r_t - s) + \sum_{s=2^{\sigma-1}}^{2^\sigma-1} (s - r_t) = 2^{2\sigma-1}.$$

## APPENDIX B

### *The Structure of the j-Level of a Harper Code*

Consider the set of $H_k(s)$ in the $j$-level of a Harper code where $2^{j-1} \leqq s \leqq 2^j - 1$. For clarity, Table III illustrates the ideas presented here by applying the ideas to the 4-level of the Harper code in Table I.

Let $\rho$ be an integer, $1 \leqq \rho \leqq j - 1$. For each value of $\rho$, the $j$-level can be divided into $2^{j-\rho}$ sets of consecutive values of $s$ each set of length $2^{\rho-1}$. The sets are numbered consecutively from 0 through $2^{j-\rho} - 1$ as follows. Let $\xi$ be an integer, $0 \leqq \xi \leqq 2^{j-\rho-1} - 1$. For each value of $\xi$, there will be two sets; an even numbered set whose number is of the form $2\xi$ and an odd numbered set whose number is of the form $2\xi + 1$.

An even numbered set contains the $H_k(s)$ for $2^{j-1} + 2\xi 2^{\rho-1} \leqq s \leqq 2^{j-1} + (2\xi + 1)2^{\rho-1} - 1$ and forms a $(\rho - 1)$-subcube because $H$ is a Harper code. Similarly, an odd numbered set contains the $H_k(s)$ for $2^{j-1} + (2\xi + 1)2^{\rho-1} \leqq s \leqq 2^{j-1} + (2\xi + 2)2^{\rho-1} - 1$ and forms a $(\rho - 1)$-subcube. The important point is that for each value of $\xi$, a useful relationship exists between set $2\xi$ and set $2\xi + 1$. Specifically,

TABLE III—DETAILS OF 4–LEVEL OF HARPER CODE IN TABLE I

| $s$ | $H_4(s)$ | $\rho = 1$ Set | $\rho = 1$ $\xi$ | $\rho = 2$ Set | $\rho = 2$ $\xi$ | $\rho = 3$ Set | $\rho = 3$ $\xi$ |
|---|---|---|---|---|---|---|---|
| 8 | 1 0 0 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 1 0 0 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 10 | 1 0 1 1 | 2 | 1 | 1 | 0 | 0 | 0 |
| 11 | 1 0 1 0 | 3 | 1 | 1 | 0 | 0 | 0 |
| 12 | 1 1 0 0 | 4 | 2 | 2 | 1 | 1 | 0 |
| 13 | 1 1 1 0 | 5 | 2 | 2 | 1 | 1 | 0 |
| 14 | 1 1 0 1 | 6 | 3 | 3 | 1 | 1 | 0 |
| 15 | 1 1 1 1 | 7 | 3 | 3 | 1 | 1 | 0 |

position 4 position 2

position 3 position 1

the $(\rho - 1)$-subcube formed by set $2\xi + 1$ is in the shadow of the $(\rho - 1)$-subcube formed by set $2\xi$. Accordingly, all $H_k(s)$ in set $2\xi + 1$ differ in exactly one position from all $H_k(s)$ in set $2\xi$. Denote the position that distinguishes the subcubes by $m$. Therefore, the $2\xi$ set consists of $2^{\rho-1}$ elements each of which has the same binary digit in position $m$. Similarly, the $2\xi + 1$ set consists of $2^{\rho-1}$ elements each of which has in position $m$ the complement of the binary digit in position $m$ in the elements of set $2\xi$.

The above sets form what we call a run in position $m$ of length $2^{\rho-1}$ that starts at $2^{j-1} + 2\xi 2^{\rho-1}$ (the first $H_k(s)$ in set $2\xi$). The definition in Section III follows from the preceding sentence.

## APPENDIX C

### *Contribution of First $2\gamma_{j,1}(t)$ Values of s in Level j to $C_t$*

From equation (8), as $s$ runs from $2^{j-1}$ through $2^{j-1} + \gamma_{j,1}(t) - 1$, every $r_t$ in the range $2^{j-1} + \gamma_{j,1}(t) \leqq r_t \leqq 2^{j-1} + 2\gamma_{j,1}(t) - 1$ occurs once and only once. Similarly, as $s$ runs from $2^{j-1} + \gamma_{j,1}(t)$ through $2^{j-1} + 2\gamma_{j,1}(t) - 1$, every $r_t$ in the range $2^{j-1} \leqq r_t \leqq 2^{j-1} + \gamma_{j,1}(t) - 1$ occurs once and only once. Therefore,

$$\sum_{s=2^{j-1}}^{2^{j-1}+2\gamma_{j,1}(t)-1} | r_t - s |$$

$$= \sum_{s=2^{j-1}}^{2^{j-1}+\gamma_{j,1}(t)-1} (r_t - s) + \sum_{s=2^{j-1}+\gamma_{j,1}(t)}^{2^{j-1}+2\gamma_{j,1}(t)-1} (s - r_t) = 2\gamma_{j,1}^2(t).$$

## APPENDIX D

### *Numerical Example to Illustrate Equations* (15) *and* (16)

Consider the Harper code given in Table I. We show how to use equation (15) when $t = 2$ and $t = 3$ to find $C_2$ and $C_3$, respectively. For $t = 2$, $\sigma = 2$ so, from (15)

$$C_2 = 8 + 2 \sum_{j=3}^{4} \sum_{i=1}^{g_j} \gamma_{j,i}^2(2).$$

In the 3-level, $\gamma_{3,1}(2)$ and $\gamma_{3,2}(2)$ are shown in Table IV. Therefore, $g_3 = 2$. Also, in the 4-level, $\gamma_{4,1}(2)$, $\gamma_{4,2}(2)$ and $\gamma_{4,3}(2)$ are given in Table IV. Thus, $g_4 = 3$. It follows that

$$C_2 = 8 + 2\,(1^2 + 1^2 + 1^2 + 1^2 + 2^2) = 24.$$

Table IV—Illustration of Equation (15) Applied to the Harper Code in Table I

| $s$ | | $H_4(s)$ | $\gamma_{j,i}(2)$ | $\gamma_{i,j}(3)$ |
|---|---|---|---|---|
| 0 | 0-level | 0 0 0 0 | | |
| 1 | 1-level | 0 0 1 0 | | |
| 2 | 2-level | 0 0 0 1 | | |
| 3 | | 0 0 1 1 | | |
| 4 | 3-level | 0 1 1 1 | $\gamma_{3,1}(2) = 1$ | $\gamma_{3,1}(3) = 2$ |
| 5 | | 0 1 1 0 | | |
| 6 | | 0 1 0 1 | $\gamma_{3,2}(2) = 1$ | |
| 7 | | 0 1 0 0 | | |
| 8 | 4-level | 1 0 0 0 | $\gamma_{4,1}(2) = 1$ | $\gamma_{4,1}(3) = 2$ |
| 9 | | 1 0 0 1 | | |
| 10 | | 1 0 1 1 | $\gamma_{4,2}(2) = 1$ | |
| 11 | | 1 0 1 0 | | |
| 12 | | 1 1 0 0 | $\gamma_{4,3}(2) = 2$ | $\gamma_{4,2}(3) = 2$ |
| 13 | | 1 1 1 0 | | |
| 14 | | 1 1 0 1 | | |
| 15 | | 1 1 1 1 | | |

position 4 position 2

position 3 position 1

Similarly, for $t = 3$, $\sigma = 2$ so, from (15),

$$C_3 = 8 + 2 \sum_{j=3}^{4} \sum_{i=1}^{g_j} \gamma_{j,i}^2(3).$$

In Table IV, $\gamma_{3,1}(3)$, $\gamma_{4,1}(3)$ and $\gamma_{4,2}(3)$ are given. Thus,

$$C_3 = 8 + 2\,(2^2 + 2^2 + 2^2) = 32.$$

By similar reasoning, the remaining $C_t$ can be found. The expression for the average numerical error of the Harper code in Table I is

$$\begin{aligned} ANE = \frac{B}{16}(&24\Pr[1 \mid 0] + 24\Pr[2 \mid 0] + 32\Pr[3 \mid 0] + 64\Pr[4 \mid 0] \\ &+ 64\Pr[5 \mid 0] + 64\Pr[6 \mid 0] + 64\Pr[7 \mid 0] + 128\Pr[8 \mid 0] \\ &+ 128\Pr[9 \mid 0] + 128\Pr[10 \mid 0] + 128\Pr[11 \mid 0] + 128\Pr[12 \mid 0] \\ &+ 128\Pr[13 \mid 0] + 128\Pr[14 \mid 0] + 128\Pr[15 \mid 0]). \end{aligned}$$

REFERENCES


1. Harper, L. H., "Optimal Assignments of Numbers to Vertices," J. Soc. Ind. Appl. Math., *12*, No. 1 (March 1964), pp. 131–135.
2. Bernstein, A. J., Steiglitz, K., and Hopcroft, J. E., "Encoding of Analog Signals for Binary Symmetric Channels," IEEE Trans. Inform. Theory, *IT-12*, No. 4 (October 1966), pp. 425–430.
3. Buchner, M. M., Jr., "Coding for Numerical Data Transmission," B.S.T.J., *46*, No. 5 (May-June 1967), pp. 1025–1041.
4. Masnick, B., and Wolf, J. K., "On Linear Unequal Error-Protection Codes," IEEE Trans. Inform. Theory, *IT-13*, No. 4 (October 1967), pp. 600–607.
5. Buchner, M. M., Jr., "A System Approach to Quantization and Transmission Error," B.S.T.J., *48*, No. 5 (May 1969), pp. 1219–1247.
6. Slepian, D., "A Class of Binary Signaling Alphabets," B.S.T.J., *35*, No. 1 (January 1956), pp. 203–234.
7. Gray, F., "Pulse Code Communication," U. S. Patent 2,632,058, March 17, 1953.
8. Dostis, I., "The Effects of Digital Errors on PCM Transmission of Compandored Speech," B.S.T.J., *44*, No. 10 (December 1965), pp. 2227–2243.