Correspondence

Conservative Arrays: Multidimensional Modulation Codes for Holographic Recording

Alexander Vardy, *Senior Member, IEEE*,
Mario Blaum, *Senior Member, IEEE*,
Paul H. Siegel, *Senior Member, IEEE*,
and Glenn T. Sincerbox

***Abstract*— In holographic storage, two-dimensional arrays of binary data is optically recorded in a medium via an interference process. To ensure optimum operation of a holographic recording system, it is desirable that the patterns of 1's (light) and 0's (no light) in the recorded array satisfy the following modulation constraint: in each row and column of the array there are at least t transitions of the type $1 \rightarrow 0$ or $0 \rightarrow 1$, for a prescribed integer t. A two-dimensional array with this property is said to be a conservative array of strength t. In general, an n-dimensional conservative array of strength t is a binary array having at least t transitions in each column, extending in any of the n dimensions of the array. We present an algorithm for encoding unconstrained binary data into an n-dimensional conservative array of strength t. The algorithm employs differential coding and error-correcting codes. Using n binary codes—one per dimension—with minimum Hamming distance $d \geq 2t-3$, we apply a certain transformation to an arbitrary information array which ensures that the number of transitions in each dimension is determined by the minimum distance of the corresponding code.**

***Index Terms*—Holographic recording, modulation codes, multidimensional modulation constraints, error-correcting codes.**

I. Introduction

In holographic storage, two-dimensional data arrays (pages) are optically recorded via an interference process and subsequently retrieved by illumination of the hologram and forming an image of a page on a matched two-dimensional array of photodetectors. To ensure optimum operation of the holographic recording system, the patterns of 1's (light) and 0's (no light) have to satisfy certain modulation constraints [5], [6]. For instance, it is desirable to avoid long periodic stretches of contiguous light or dark in both dimensions [5]. To this end, one would like to have as many transitions as possible from light to dark and from dark to light in each row and each column of the recorded data page. This may be achieved by requiring that in each row and column of the recorded array there are at least t transitions of the type $1 \rightarrow 0$ or $0 \rightarrow 1$, for a prescribed integer t. Extending the terminology of [1], [11], [12], we shall say that a binary array with this property is a *conservative array* of strength t. Thus the modulation problem at hand is to encode unconstrained binary data into a conservative array of a given strength.

Manuscript received September 6, 1994; revised July 20, 1995. This work was supported in part by the National Science Foundation under Grant NCR–9409688. The material in this correspondence was presented in part at the 32nd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 1994.

A. Vardy is with the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801 USA.

M. Blaum and G. T. Sincerbox are with IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA.

P. H. Siegel is with the Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093 USA.

Publisher Item Identifier S 0018-9448(96)00015-6.

Although as discussed above, the main application of conservative arrays for holographic storage is in two, or perhaps three, dimensions we shall also present a solution to the modulation problem for any number of dimensions. When referring to multidimensional arrays, we run into a problem of terminology as the English words "row" and "column" have no counterparts beyond two dimensions. Thus given a k-dimensional binary array $V = [v_{i_1 i_2 \cdots i_k}]$ of size $n_1 \times n_2 \times \cdots \times n_k$, we shall say that a *column* of V is obtained by fixing some $k-1$ subscripts and letting the remaining, say lth, subscript vary from 1 to n_l. For instance, the number of columns in a k-dimensional cube with edges of length n bits is kn^{k-1}. More precisely, we shall refer to the one-dimensional sequence of binary values $(x_1, x_2, \cdots, x_{n_l})$ as an *l-column* of V if $x_i = v_{j_1 j_2 \cdots j_l \cdots j_k}$, where $j_l = i$ while all the other subscripts are the same for all $x_1, x_2, \cdots, x_{n_l}$. A k-dimensional conservative array of strength t may be now defined by requiring that for all $l = 1, 2, \cdots, k$, in each l-column of the array there are at least t transitions. Using differential coding, the pigeonhole principle (cf. [7]), and k binary codes—one per dimension—with minimum Hamming distance $d \geq 2t-3$, we derive an efficient algorithm for encoding unconstrained binary data into a k-dimensional conservative array of strength t. The algorithm can be readily implemented using simple logic circuits.

The related works of [1], [11], [12] also deal with constraints on the number of transitions in a binary sequence. Motivated by various applications for bit synchronization, the objective of [1], [11], [12] was to encode information into binary sequences of length n having precisely $\lfloor n/2 \rfloor$ transitions. We point out that using differential coding, as described in this correspondence, this objective can be achieved immediately with any of the known means for constructing balanced codes, say the Henry–Knuth algorithm [8], [9]. However, our modulation problem is much more involved since the constraint on the number of transitions must be satisfied *simultaneously* in all the columns, extending in any of the k dimensions of the array.

The rest of this correspondence is organized as follows. In the next section we describe the main ideas and derive the encoding algorithm for two dimensions. Generalization to k-dimensional conservative arrays for $k \geq 3$ is presented in Section III-A. Finally, combination of the proposed algorithm with additional modulation constraints and error-correcting coding is discussed in Section III-B.

II. Conservative Arrays in Two Dimensions

Suppose that the data to be encoded are presented in the form of an $m \times n$ binary array V. We proceed by constructing a set of modulation arrays $\mathcal{U}$, such that for any $m \times n$ array V there exists at least one element $U \in \mathcal{U}$, with $V \oplus U$ being conservative of strength t. This approach is, in a sense, analogous to the well-known Henry–Knuth algorithm [8], [9] for balancing binary vectors. Indeed, to balance an arbitrary vector of even length n, the Henry–Knuth algorithm essentially employs the set $\mathcal{K}_n = \{(1^i \mid 0^{n-i})\colon\ 0 \leq i \leq n-1\}$, where $(\cdot|\cdot)$ stands for concatenation, and 1^i denotes a vector of length i whose coordinates are fixed at 1. This set has the property that for any vector $\underline{v}$ of length n there is at least one $\underline{u} \in \mathcal{K}_n$, such that $\underline{v} \oplus \underline{u}$ is balanced.

In what follows, we shall count the transitions $0 \rightarrow 1$ and $1 \rightarrow 0$ in a cyclic manner. Thus if a vector begins with a 1 and ends with 0, this would count as a $0 \rightarrow 1$ transition. For example, there are two $0 \leftrightarrow 1$

0018–9448/96$05.00 © 1996 IEEE

transitions in the vector 1110001111 and four $0 \leftrightarrow 1$ transitions in the vector 1110001110. With this definition of $0 \leftrightarrow 1$ transitions, it is clear that in any vector the number of $0 \rightarrow 1$ transitions is equal to the number of $1 \rightarrow 0$ transitions. Hence the total number of such transitions must be even.

The above method of counting transitions is not realistic—there is no physical justification for counting the transitions in a cyclic manner. However, as we shall see, counting the transitions cyclically greatly simplifies the derivation. Obviously, any sequence having at least t cyclic transitions has at least $t-1$ noncyclic "real" transitions. So, the difference between the two methods of counting is negligible for large t. In practical applications of the proposed encoding method, this difference would likely be accounted for by choosing the design parameter t to be one larger than the desired number of "real" transitions.

Let $\mathbb{F}_n$ denote the set of all binary n-tuples, and let $\mathbb{E}_n$ denote the subset of $\mathbb{F}_n$ consisting of all the n-tuples of even weight. For any $\underline{x} = (x_1, x_2, \cdots, x_n) \in \mathbb{F}_n$ let $\sigma(\underline{x}) = (x_n, x_1, \cdots, x_{n-1})$ be the cyclic shift of $\underline{x}$. Define a mapping $\phi : \mathbb{F}_n \rightarrow \mathbb{E}_n$ as follows:

$$\phi(\underline{x}) = \underline{x} \oplus \sigma(\underline{x}).$$

Note that the mapping ϕ is linear and that the number of $0 \leftrightarrow 1$ transitions in $\underline{x}$ is equal to the Hamming weight of $\phi(\underline{x})$ (the action of ϕ on $\mathbb{F}_n$ is analogous to the action of the $1 + D$ channel on semi-infinite sequences, or to the action of the differential on continuous functions). Although ϕ is not invertible we can formally define the inverse mapping. Let $\underline{y} = (y_1, y_2, \cdots, y_n) \in \mathbb{E}_n$ and $\underline{x} = (x_1, x_2, \cdots, x_n) \in \mathbb{F}_n$. Then $\phi^{-1}(\underline{y}) = \underline{x}$ if and only if $\phi(\underline{x}) = \underline{y}$ and $x_1 = 0$. Since $\phi(\underline{x}) = \phi(\underline{x}')$ implies that either $\underline{x} = \underline{x}'$, or $\underline{x}$ and $\underline{x}'$ are complements of each other, it is clear that each element of $\mathbb{E}_n$ will have exactly one inverse.

Now let C_1 be a $(t-2)$-error-correcting code of length m, and assume that C_1 contains at least $(n+1)$ codewords of even weight. Similarly, let C_2 be a $(t-2)$-error-correcting code of length n, containing at least $(m+1)$ codewords of even weight. When n and m are powers of 2, such codes exist for all $t \leq \min\{n/4, m/4\} + 1$. These are just the first-order Reed–Muller codes of lengths m and n, respectively. Now, define the sets $\phi^{-1}(C_1)$ and $\phi^{-1}(C_2)$ as follows:

$$\begin{aligned} \phi^{-1}(C_1) &= \{\phi^{-1}(\underline{c}): \quad \underline{c} \in C_1 \text{ and } |\underline{c}| \equiv 0 \bmod 2\} \\ \phi^{-1}(C_2) &= \{\phi^{-1}(\underline{c}): \quad \underline{c} \in C_2 \text{ and } |\underline{c}| \equiv 0 \bmod 2\} \end{aligned} \tag{1}$$

where $|\cdot|$ denotes the Hamming weight. Let $\underline{a}_0, \underline{a}_1, \cdots, \underline{a}_n$ be some $(n+1)$ fixed elements of $\phi^{-1}(C_1)$, and let $\underline{b}_0, \underline{b}_1, \cdots, \underline{b}_m$ be some $(m+1)$ fixed elements of $\phi^{-1}(C_2)$.

Proposition 1: For any $\underline{x} \in \mathbb{F}_m$, if $\underline{x} + \underline{a}_i$ has less than t transitions, then $\underline{x} + \underline{a}_j$ has at least t transitions for all $j \neq i$.

Proof: Otherwise, we would have

$$\begin{aligned} d(\phi(\underline{x}), \phi(\underline{a}_i)) &= |\phi(\underline{x}) + \phi(\underline{a}_i)| = |\phi(\underline{x} + \underline{a}_i)| \leq t-2 \\ d(\phi(\underline{x}), \phi(\underline{a}_j)) &= |\phi(\underline{x}) + \phi(\underline{a}_j)| = |\phi(\underline{x} + \underline{a}_j)| \leq t-2 \end{aligned} \tag{2}$$

where $d(\cdot,\cdot)$ denotes the Hamming distance, and the last inequality follows from the fact that the Hamming weight of $\phi(\underline{x})$ is equal to the number of $0 \leftrightarrow 1$ transitions in $\underline{x}$. Since $\phi(\underline{a}_i), \phi(\underline{a}_j) \in C_1$, it follows from (2) that the vector $\phi(\underline{x})$ is at distance at most $(t-2)$ from two distinct codewords of C_1, which is a contradiction. □

The set of modulation arrays $\mathcal{U}$ may now be obtained as follows. Let $\mathcal{A} = \{A_0, A_1, \cdots, A_n\}$ and $\mathcal{B} = \{B_0, B_1, \cdots, B_m\}$ be two fixed sets of $n \times m$ binary arrays, with the elements of $\mathcal{A}$ and $\mathcal{B}$, namely $A_i = [a_{jk}^{(i)}]$ and $B_i = [b_{jk}^{(i)}]$, defined by

$$a_{jk}^{(i)} = \begin{cases} 1, & \text{if the } j\text{th coordinate of } \underline{a}_i = 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{for } i = 1, 2, \cdots, n$$
$$b_{jk}^{(i)} = \begin{cases} 1, & \text{if the } k\text{th coordinate of } \underline{b}_i = 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{for } i = 1, 2, \cdots, m. \tag{3}$$

The set of modulation arrays $\mathcal{U}$ is then a direct sum of the sets $\mathcal{A}$ and $\mathcal{B}$, as follows:

$$\mathcal{U} = \{ A \oplus B: A \in \mathcal{A}, B \in \mathcal{B}\}.$$

Note that adding an element of $\mathcal{B}$ to an arbitrary $m \times n$ array V is equivalent to complementing a certain set of columns of V. This does not affect the number of $0 \leftrightarrow 1$ transitions in the columns of V, since a vector and its complement contain the same number of $0 \leftrightarrow 1$ transitions. Similarly, adding an element of $\mathcal{A}$ to V does not affect the number of $0 \leftrightarrow 1$ transitions in the rows of V. However, if $\underline{v}_1, \underline{v}_2, \cdots, \underline{v}_n$ is the set of *columns* of V, then the set of columns of $V \oplus A_i$ is just $\underline{v}_1 + \underline{a}_i, \underline{v}_2 + \underline{a}_i, \cdots, \underline{v}_n + \underline{a}_i$. In view of Proposition 1, for each specific column of V there is at most one element $A \in \mathcal{A}$, such that the number of $0 \leftrightarrow 1$ transitions in the corresponding column of $V \oplus A$ is less than t. Since the cardinality of $\mathcal{A}$ is strictly greater than the number of columns in V, there exists, according to the pigeonhole principle, at least one $A^* \in \mathcal{A}$ such that *all* the columns of $V \oplus A^*$ contain at least t transitions. A similar argument shows that there exists at least one $B^* \in \mathcal{B}$ such that the number of $0 \leftrightarrow 1$ transitions in all the *rows* of $V \oplus B^*$ is at least t. Therefore, $V \oplus U$, where $U = A^* \oplus B^* \in \mathcal{U}$, is a conservative array of strength t.

Example: Let $n = m = 3$, and $t = 2$. Since $t - 2 = 0$, the set of four binary 3-tuples of even weight may be taken as C_1 and C_2. Thus $C_1 = C_2 = \{000, 011, 110, 101\}$ and $\{\underline{a}_0, \underline{a}_1, \underline{a}_2, \underline{a}_3\} = \{\underline{b}_0, \underline{b}_1, \underline{b}_2, \underline{b}_3\} = \{000, 001, 010, 011\}$. Hence we have

$$A_0 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \quad A_1 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{matrix}$$

$$A_2 = \begin{matrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{matrix} \quad A_3 = \begin{matrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$$

$$B_0 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \quad B_1 = \begin{matrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{matrix}$$

$$B_2 = \begin{matrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{matrix} \quad B_3 = \begin{matrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{matrix}$$

The set of modulation arrays $\mathcal{U}$ thus consists of the following 16 arrays

$$\begin{matrix} 0&0&0 \\ 0&0&0 \\ 0&0&0 \end{matrix} \quad \begin{matrix} 0&0&0 \\ 0&0&0 \\ 1&1&1 \end{matrix} \quad \begin{matrix} 0&0&0 \\ 1&1&1 \\ 0&0&0 \end{matrix} \quad \begin{matrix} 0&0&0 \\ 1&1&1 \\ 1&1&1 \end{matrix}$$

$$\begin{matrix} 0&0&1 \\ 0&0&1 \\ 0&0&1 \end{matrix} \quad \begin{matrix} 0&0&1 \\ 0&0&1 \\ 1&1&0 \end{matrix} \quad \begin{matrix} 0&0&1 \\ 1&1&0 \\ 0&0&1 \end{matrix} \quad \begin{matrix} 0&0&1 \\ 1&1&0 \\ 1&1&0 \end{matrix}$$

$$\begin{matrix} 0&1&0 \\ 0&1&0 \\ 0&1&0 \end{matrix} \quad \begin{matrix} 0&1&0 \\ 0&1&0 \\ 1&0&1 \end{matrix} \quad \begin{matrix} 0&1&0 \\ 1&0&1 \\ 0&1&0 \end{matrix} \quad \begin{matrix} 0&1&0 \\ 1&0&1 \\ 1&0&1 \end{matrix}$$

$$\begin{matrix} 0&1&1 \\ 0&1&1 \\ 0&1&1 \end{matrix} \quad \begin{matrix} 0&1&1 \\ 0&1&1 \\ 1&0&0 \end{matrix} \quad \begin{matrix} 0&1&1 \\ 1&0&0 \\ 0&1&1 \end{matrix} \quad \begin{matrix} 0&1&1 \\ 1&0&0 \\ 1&0&0 \end{matrix}$$

◇

Since the modulation set $\mathcal{U}$ consists of $(m+1)(n+1)$ elements, in order to indicate which modulation array was applied to V we need exactly $\lceil \log_2 (m+1)(n+1) \rceil$ redundant bits. This is logarithmic in the size of the input array V, and therefore the rate of the modulation encoder approaches its bounding value of 1 quite rapidly. After recording a sequence of about $mn/\log_2 (m+1)(n+1)$ data pages, the redundant bits for all these pages can be again encoded into a conservative array of strenght t, and so on recursively. Alternatively, the redundancy bits can be appended as an extra row or column to each recorded page. For more details on implementation issues see [5].

III. Concluding Remarks

This section deals with generalizing the construction presented in Section II to higher dimensions, combining this construction with the balancing algorithm of Henry–Knuth [8], [9], and using the resulting balanced conservative arrays for efficient error detection and correction in holographic storage.

A. Multidimensional Conservative Arrays

We presently show that the modulation method of the previous section easily extends to more than two dimensions. Suppose that the data is presented in the form of a k-dimensional array $V = [v_{i_1 i_2 \cdots i_k}]$ of size $n_1 \times n_2 \times \cdots \times n_k$. As discussed in the Introduction, an l-column of V is obtained by letting the lth subscript vary from 1 to n_l and fixing the remaining $k-1$ subscripts. Conversely, a hyperplane of V is obtained by fixing one subscript and letting the other $k-1$ subscripts range over all the possible values. Thus an l-hyperplane of V is a $(k-1)$-dimensional array given by $[v_{j_1 j_2 \cdots j_l \cdots j_k}]$, where j_l is fixed at a certain specific value, while for all $i \neq l$ the subscripts j_i vary from 1 to n_i.

We now construct a set of k-dimensional arrays $\mathcal{U}$, such that for any $n_1 \times n_2 \times \cdots \times n_k$ array V there exists at least one $U \in \mathcal{U}$ with $V \oplus U$ being conservative. Again, let $C_1, C_2, \cdots, C_k$ be $(t-2)$-error-correcting codes of lengths $n_1, n_2, \cdots, n_k$, respectively. Further, suppose that for $l = 1, 2, \cdots, k$ the code C_l contains at least $(\prod_{i=1, i\neq l}^{k} n_i) + 1$ codewords, all of even weight. Note that $m_l = \prod_{i=1, i \neq l}^{k} n_i$ is precisely the number of l-columns in V. Just as in the two-dimensional case, the requirement that $|C_l| \geq m_l + 1$ for each l determines the highest attainable value of t. That is, t must be such that the codes $C_1, C_2, \cdots, C_k$ exist for all $l = 1, 2, \cdots, k$. Now let the sets $\phi^{-1}(C_1), \phi^{-1}(C_2), \cdots, \phi^{-1}(C_k)$ be defined as in (1), and let $\underline{a}_0^l, \underline{a}_1^l, \cdots, \underline{a}_{m_l}^l$ be some $m_l + 1$ fixed elements of $\phi^{-1}(C_l)$. We construct the set of $m_l + 1$ k-dimensional arrays $\mathcal{A}_l = \{A_0^l, A_1^l, \cdots, A_{m_l}^l\}$ of size $n_1 \times n_2 \times \cdots \times n_k$ in a manner analogous to (3). Namely, for $i = 0, 1, \cdots, m_l$, each A_i^l is composed of n_l l-hyperplanes, with the entries in each such l-hyperplane being either all 0's or all 1's. More specifically, all the entries in the jth l-hyperplane of A_i^l are taken to be equal to the jth entry of the vector $\underline{a}_i^l$. Once the sets $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k$ have been constructed, the required set of $(m_1+1)(m_2+1)\cdots(m_k+1)$ k-dimensional modulation arrays $\mathcal{U}$ is obtained as the direct sum of these sets

$$\mathcal{U} = \{A_1 \oplus A_2 \oplus \cdots \oplus A_k \colon A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, \cdots, A_k \in \mathcal{A}_k\}.$$

Proposition 1 may be now used to show that for any $n_1 \times n_2 \times \cdots \times n_k$ input array V there exists at least one array $U \in \mathcal{U}$, such that $V \oplus U$ is conservative of strength t.

The number of redundant bits required by the multidimensional modulation encoder described above is just the logarithm of the cardinality of $\mathcal{U}$ given by

$$\lceil \log_2 (m_1+1)(m_2+1)\cdots(m_k+1) \rceil \approx (k-1)\log_2 N$$

where $N = n_1 n_2 \cdots n_k$ is the total number of information bits in V. Thus the redundancy of the multidimensional modulation scheme is still logarithmic in the input length.

We note that the idea of inverting entire hyperplanes allows us to break the inherent dependence between the columns of a multidimensional array. This, along with differential coding $\phi(\cdot)$, essentially transforms the modulation problem at hand into the following: find the smallest subset $\mathcal{A} \subset \mathbb{E}_n$, such that given any m-subset of $\mathbb{E}_n$, there is at least one element of $\mathcal{A}$ at Hamming distance $\geq t$ from all the elements of the m-subset. It is easy to see that the smallest possible cardinality of $\mathcal{A}$ is $m+1$ and that $(t-2)$-error-correcting codes provide the optimal solution to this problem.

B. Further Constraints and Error Correction

Another common requirement in holographic recording is that the number of 1's and 0's in a data page should be balanced. In this way, during recording, the amount of signal light is independent of the data content of a page and, during retrieval, the light coming from a stored page is divided between the same number of beams, again independent of data content. Thus it is desirable to balance the recorded arrays, which may be accomplished using, for instance, the Henry–Knuth algorithm [8], [9]. We now show that the proposed modulation encoder can be readily combined with that of Knuth, provided the latter proceeds in a horizontal row-by-row order. Specifically, this means that to apply the one-dimensional modulation vector $\underline{u} \in \mathcal{K}_{mn}$ to a two-dimensional $m \times n$ array $V = [v_{ij}]$, we represent V as a vector

$$\underline{v} = (v_{11}, v_{12}, \cdots, v_{1n}, v_{21}, v_{22}, \cdots, v_{2n}, \cdots, v_{m1}, v_{m2}, \cdots, v_{mn}). \tag{4}$$

Now suppose that V is a conservative array of strength $t+2$ and that several bits of V are complemented so that the resulting array, $W \oplus V$ say, is balanced. It is easy to see that if the bits of V are complemented in the readout order of (4) then $W \oplus V$ is necessarily a conservative array of strength at least t. Thus we have a modulation encoder from unconstrained binary data into balanced conservative arrays in two dimensions. A similar argument applies to k-dimensional arrays, provided the bits of V are complemented in the readout order analogous to (4), that is, recurrently hyperplane-by-hyperplane.

The balancing of recorded pages also provides an effective means for error detection, since the errors in holographic storage tend to be asymmetric—it is much more likely that a 0 will be detected as a 1 than conversely. The latter condition occurs because scattered light and/or crosstalk light from adjacent pages can always be detected if it illuminates a bit location that was originally a 0, hence a 0 bit could become a 1 bit if there is sufficient noise light. Conversion of a 1 bit to a 0 bit requires that the data light and the noise light interfere destructively; that is, they must have the same amplitude and be 180° out of phase. This is much less likely, thereby giving rise to the error asymmetry. The asymmetric nature of the errors will be detected by the balanced code, which can be used to declare erasures at the affected blocks. The information in the erased blocks can then be retrieved using an outer error-correcting code. Since the errors also tend to occur in large two-dimensional clusters [10], the outer error-correcting code should be capable of handling very large symbols with relatively low decoding complexity. Thus the

MDS codes with large symbols developed in [3], [4] are ideally suited to holographic recording. All this, possibly combined with two-dimensional interleaving [2], provides an extremely powerful coding scheme for holographic memory systems.

REFERENCES

[1] S. Al-Bassam and B. Bose, "Conservative codes," in *IEEE Int. Symp. on Information Theory* (San Antonio, TX, 1993).

[2] M. Blaum and J. Bruck, "Correcting two-dimensional clusters by interleaving of symbols," in *IEEE Int. Symp. on Information Theory* (Trondheim, Norway, 1994).

[3] M. Blaum, J. Bruck, and A. Vardy, "MDS array codes with independent parity symbols," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, Mar. 1996, to be published.

[4] M. Blaum and R. M. Roth, "New array codes for multiple phased burst correction," *IEEE Trans. Inform. Theory*, vol. 39, pp. 66–77, 1993.

[5] M. Blaum, P. H. Siegel, G. T. Sincerbox, and A. Vardy, "Method and apparatus for modulation of multi-dimensional data in holographic storage," filed with the US Patent Office, Aug. 1994.

[6] D. Brady and D. Psaltis, "Control of volume holograms," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1167–1182, 1992.

[7] R. A. Brualdi, *Introductory Combinatorics*. New York: North-Holland, 1977.

[8] P. S. Henry, "Zero disparity coding system," US Patent 4 309 694, 1982.

[9] D. E. Knuth, "Efficient balanced codes," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 51–53, 1986.

[10] S. A. Lis and P. D. Henshaw, "Ultra-dense optical mass storage," AD-A232 767, report prepared for USAF, AFSC, Feb. 1991.

[11] Y. Ofek, "A family of conservative codes with block delimiters for decoding without a phase-locked loop," in *ACM Comput. Sci. Conf.* (Baltimore, MD, 1987).

[12] ——, "The conservative code for bit synchronization," *IEEE Trans. Commun.*, vol. 38, pp. 1107–1113, 1990.

Detection of Binary Markov Sources Over Channels with Additive Markov Noise

Fady Alajaji, Nam Phamdo, *Member, IEEE*, Nariman Farvardin, *Senior Member, IEEE*, and Thomas E. Fuja, *Member, IEEE*

***Abstract*— We consider maximum *a posteriori* (MAP) detection of a binary asymmetric Markov source transmitted over a binary Markov channel. Here, the MAP detector observes a long (but finite) sequence of channel outputs and determines the most probable source sequence. In some cases, the MAP detector can be implemented by simple rules such as the "believe what you see" rule or the "guess zero (or one) regardless of what you see" rule. We provide necessary and sufficient conditions under which this is true. When these conditions are satisfied, the exact bit error probability of the sequence MAP detector can be determined. We examine in detail two special cases of the above source: i) binary independent and identically distributed (i.i.d.) source and ii) binary symmetric Markov source. In case i), our simulations show that the performance of the MAP detector improves as the channel noise becomes more correlated. Furthermore, a comparison of the proposed system with a (substantially more complex) traditional tandem source-channel coding scheme portrays superior performance for the proposed scheme at relatively high channel bit error rates. In case ii), analytical as well as simulation results show the existence of a "mismatch" between the source and the channel (the performance *degrades* as the channel noise becomes more correlated). This mismatch is reduced by the use of a simple rate-one convolutional encoder.**

***Index Terms*— Markov source, Markov channel, source redundancy, MAP detection.**

I. INTRODUCTION AND MOTIVATION

A source with memory as well as a memoryless source with a nonuniform distribution are sources with *redundancy*. For a finite alphabet of size J, a uniformly distributed independent and identically distributed (i.i.d.) random process contains a maximal amount of information and exhibits no redundancy. Its entropy rate is equal to $\log_2 J$ bits/sample. The total redundancy a stationary ergodic J-ary alphabet source $\{X_n\}_{n=1}^{\infty}$ possesses is equal to the difference between $\log_2 J$ and its entropy rate $H_\infty(X)$ [9]: $\rho_T = \log_2 J - H_\infty(X)$, where

$$H_\infty(X) \triangleq \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \cdots, X_n).$$

The redundancy may be attributed to the nonuniform source distribution or to the source memory (or both). More specifically, we can write $\rho_T = \rho_D + \rho_M$ where $\rho_D \triangleq \log_2 J - H(X_1)$ denotes the redundancy in the form of a nonuniform distribution and

Manuscript received March 27, 1995; revised June 26, 1995. The work of F. Alajaji, N. Farvardin, and T. E. Fuja was supported in part by the National Science Foundation under Grant NSFD CD-8803012. The work of N. Phamdo was supported in part by NTT Corporation. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Trondheim, Norway, June 1994.

F. Alajaji is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada.

N. Phamdo is with the Electrical Engineering Department, State University of New York, Stony Brook, NY 11794-2350 USA.

N. Farvardin and T. E. Fuja are with the Electrical Engineering Department, University of Maryland, College Park, MD 20742 USA.

Publisher Item Identifier S 0018-9448(96)00016-8.