BOUNDS ON THE CAPACITY OF THE PEAK SHIFT MAGNETIC RECORDING CHANNEL

by

Shlomo Shamai (Shitz) and Ephriam Zehavi*

Department of Electrical Engineering Technion — Israel Institute of Technology Haifa 32000, Israel

ABSTRACT

A simple statistical channel model is suggested to account for single position peak (bit) shifts which were identified to be one of the major impairments in magnetic recording.

Simple, lower and upper bounds on the capacity of this channel with \((d, k)\) channel inputs (commonly used in magnetic recording) are presented. For \(d \geq 2\), this channel is conveniently described in terms of phrase lengths where a bit shift causes a phrase either to shrink or to expand. The inherent correlation present in consecutive shift affected phrases manifests itself in memory introduced into the channel model. The lower bound, when evaluated for a certain phrase distribution turns to be also a lower bound on the zero error capacity for this channel.

I. INTRODUCTION

Peak detectors are a common practice in standard high density magnetic recording systems [1]-[3]. It has been noticed that one of the major impairments in these systems is the so-called bit shift (peak shift) [1]-[5]. The bit shifts are caused by readback peaks being shifted in time due to noise effects, read-out circuitry imperfection and clock jittering. In some cases this phenomena accounts for most of the errors monitored in the system [5]. It has also been noticed [5] that in a well designed recording system most of the bit shifts span for only one bit position and the probabilities of the right-pairs of one symbols, are used to control intersymbol maximum length of consecutive zero symbol runs between interference and self clocking.

The commonly used RLL sequences do not usually possess significant error correcting capabilities [7] and therefore they are unable to cope efficiently with the channel impairments. This has motivated some recent studies of combined error correcting codes having RLL properties [7]-[10]. Codes designed specifically to cope with a single one position bit shift within an encoded block are considered in [11] whereas multi bit shifts of a single position are mentioned in [12].

Naturally, the question of the maximum reliable rate of information, that is capacity, conveyed by RLL sequences, is encountered. For the ideal errorless channel the answer has long been known [13] and capacity equals the logarithm of the largest root of a certain polynomial specified by the \((d, k)\) constraints. See [7] for an overview and further details. Recently, bounds on the capacity have been reported also for the memoryless binary symmetric channel (BSC) and for the discrete time memoryless Gaussian additive channel [6], [14]. These elementary channel models account for the random errors observed in practical recording systems. However, they do not capture the bit shift effects and therefore are of limited significance in predicting the ultimate performance of a large class of practical systems for which bit shifts are the major cause for performance deterioration.

In this paper we propose a simple statistical channel model termed in short PSC to model a single position (possibly asymmetrical) peak (bit) shift and derive lower and upper bounds on the capacity of the proposed model, with the constraint of RLL input sequences. Interesting lower bounds on the zero error capacity of the PSC are also worked out and discussed. Some properties of the bounds are explored and numerical results for a variety of parameters are exhibited. Concluding remarks and extensions of the PSC model are shortly stated at the end of this paper.

II. THE BIT SHIFT CHANNEL MODEL

A single peak (bit) shift channel (PSC) is best formulated in terms of phrase lengths where a phrase is uniquely defined by a consecutive sequence of bits starting with none, one or more zeros ("0") and terminating with the first single one ("1") [6]. Any binary sequence of zeros and ones is uniquely decomposable into a concatenated sequence of phrases. Single position bit shifts cause the "1" symbol terminating the phrase to wander by one single position to its right (right shift) or to its left (left shift). This is demonstrated in Figure 1 where a specific \((d, k) = (2, 5)\) input sequence undergoes bit shifts resulting in the output sequence shown also in Figure 1. The places where left or right shifts occurred are designated with small arrows pointing to the direction of the shifts. The input and output phrases and their respective lengths are explicitly shown. We conclude therefore that the PSC is conveniently formulated in terms of phrase lengths. The bit shift effect, restricted throughout to no more than a single bit position, either shrinks or expands the input phrase. Of course the phrase lengths are not modified if no bit shift has taken place. We restrict our discussion to \(d \geq 2\). Following the example depicted in Figure 1, we conclude that in this case additional phrases are neither generated nor existing phrases are destroyed.

* This work was partially performed while E. Zehavi was a Visiting Researcher at QUALCOMM, Inc., San Diego, CA 92121. The research was partially supported by QUALCOMM, Inc.
Let $x_i$ stand respectively for the $i$-th channel input phrase length and $y_i$ for the corresponding channel output phrase length. An inherent pairwise correlation is introduced into the $(y_i)$; this is clearly demonstrated in Figure 1, where a single left shift at the end of the second input phrase ($x_2 = 5$) has affected both the second and the third output phrases ($y_2$ and $y_3$, respectively). Keeping track of this consecutive inherent correlation, yields,

$$y_i = x_i + e_i \cdot e_{i-1}$$

(1)

where $e_i$ is a ternary random variable taking on $(-1,0,+1)$ values designating whether a left $(-1)$, a right $(+1)$ or no $0$ bit shift has occurred at the end of the $i$-th phrase. This configuration is schematically described in Figure 2b whereas in Figure 2a the PSC is schematically represented by a block, the inputs to which are binary $(d,k)$ codes and the outputs are the corresponding binary symbols. We assume here that $(e_i)$ is an independent identically distributed (i.i.d.) sequence with the probability distribution $P(e_i = 1) = P_j$, $P(e_i = -1) = P_{-j}$ and $P(e_i = 0) = 1 - P_j - P_{-j}$. It is of convenience to introduce the parameters

$$P_0 = 1 - (P_j + P_{-j})$$

$$P_{as} = (P_j - P_{-j})/2$$

(2)

where $P_0$ is interpreted as the probability of no bit shift (in either direction) taking place and $P_{as}$ is the shift asymmetry factor.

III. BOUNDS ON THE PSC CAPACITY

Equations (1) specifies a discrete time channel in which the inputs consisting of phrase lengths of a $(d,k)$ sequence, take on $x_i \in \{d+1, d+2, \ldots, k+1\}$ values and the outputs correspondingly assume the values $y_i \in \{d-1, d, \ldots, k+1, k+2, k+3\}$. We are interested in the capacity of this PSC that is: the ultimate reliable rate of transmission in the presence of the bit shift impairments with $(d,k)$ input sequences. The capacity $C_{PSC}$ measured in bits per input information symbol can be expressed by:

$$C_{PSC} = \lim_{N \to \infty} \sup_{P(X^N) \in \mathbb{P}(d,k)} \left\{ \frac{I(Y^N : X^N)}{E(X^N)} \right\}$$

(3)

where $X^N = (x_1, x_2, \ldots, x_N)$ and $Y^N = (y_1, y_2, \ldots, y_N)$ stand respectively for the channel input and output vectors with the phrase length components $x_i$ and $y_i$. The notation $I(\cdot : \cdot)$ is used for mutual information [15] and $E$ stands for the statistical expectation. The supremum is taken over all $P(X^N)$ - the probability measures of $X^N$ satisfying the $(d,k)$ constraint, that is: $x_i \in \{d+1, d+2, \ldots, k+1\}$ with $i = 1, 2, \ldots, N$. The whole class of the $(d,k)$ constrained probability measures of vectors of length $N$ is explicitly denoted by $\mathbb{P}(d,k)$. The normalization of $I(Y^N : X^N)$ specified by the divisor $E(X^N) \triangleq \sum_{i=1}^{N} E(x_i)$ - the average length of $N$ phrases - is introduced in order to determine the capacity in terms of bits per a channel input symbol that is: per a channel, described in Figure 2a, use (and not per an input phrase where the latter is determined by replacing $E(X^N)$ by $N$). This normalization is in place if a comparison to the capacity of other channels on a bit by bit basis is of interest and it is commonly applied in cases where the phrase representation is natural and lends itself to a relatively simple analysis. See [16] for further details. With no loss of generality we restrict our attention to stationary phrase length sequences $(x_i)$. This follows since the channel described by (1) is discrete time invariant though not memoryless and the capacity achieving distribution of $(x_i)$ is therefore stationary, (see detailed arguments in [14]). For convenience, we use the same notation $P_N(d,k)$ to describe the class of all stationary probability measures of $N$-tuples of phrase lengths satisfying the $(d,k)$ constraints.

a. An Upper Bound on Capacity

The mutual information is conveniently expressed

$$\frac{1}{N} I(Y^N : X^N) = \frac{1}{N} H(Y^N) - \frac{1}{N} H(Y^N | X^N)$$

(4)

where $H(\cdot)$ stands for the entropy or conditional entropy functions [15].

Using now the well-known relation

$$\lim_{N \to \infty} \frac{1}{N} H(Y^N) = \lim_{N \to \infty} \frac{1}{N} H(x_1, \ldots, y_i, \ldots, y_N) \leq H(y_{i+1} | y_i, y_2, \ldots, y_N)$$

for $i = 0, 1, \ldots$ (5)

along with the observation that,

$$\frac{1}{N} H(Y^N | X^N) = \frac{1}{N} H(g^N) = H(\epsilon)$$

$$= -P_{-1} \log P_{-1} - (1 - P_{-1} - P_j) \log (1 - P_{-1} - P_j) - P_j \log P_j$$

and specializing to $i = 0$ where $H(y_1 | y_0) \overset{\Delta}{=} H(y_1)$, yield the upper bound $C_{Uo}$

$$C_{PSC} \leq C_{Uo} = \sup_{P(x) \in \mathbb{P}_j(d,k)} \frac{H(Y) - H(\epsilon)}{E(x)}$$

(7)

where $y_1 = y$ and $x_1 = x$ (omitting for the sake of convenience the index 1) and $E(x)$ stands for the average
phrase length. The supremum in (7) is taken over the probability distribution of the input phrase length \( P(x) \) within the class of stationary \((d, k)\) phrases. We note that \( y = x + v \) where \( v = \varepsilon_j - \varepsilon_{j-1} \) has the probability function

\[
P_y(k) = \text{Prob}(y = k) = \sum_{i=2}^{k+1} f_{i-1} P_y(1)
\]

The average phrase length is, \( E(x) = \sum_{i=d+1}^{k+1} f_i \). Now, expression (7) with \( H(y) = -\sum_{i=0}^{k+1} P_y(k) \log P_y(k) \) should be supremized over the \( k-d+1 \) non-negative probability variables \( 0 \leq f_i \leq 1, i = d+1, d+2, \ldots, k+1 \) satisfying evidently \( \sum_{i=d+1}^{k+1} f_i = 1 \). Note that \( H(\varepsilon) \) in (7) is given by equation (6) and it is not a function of \( f_i \). Following a straightforward Lagrange maximization approach noting that \( \frac{\partial P_y(k)}{\partial f_i} = P_y(k-i) \) and \( \frac{\partial E(x)}{\partial f_i} = i \) yields the following set of \( k-d+1 \) equations

\[
\lambda = \sum_{i=d+1}^{k+1} f_i P_y(1) - i \{ H(y) - H(\varepsilon) \} \]

\[
\frac{\sum_{m=-d}^{k} \log \{ e P_y(m) \} P_y(m) E(x) - i \{ H(y) - H(\varepsilon) \} }{E^2(x)}
\]

where \( i = d+1, d+2, \ldots, k+1 \). The \( k-d+1 \) unknown probability variables \( f_i, i = d+1, d+2, \ldots, k+1 \) and the nonnegative Lagrange multiplier \( \lambda \) are determined by the set of equations (10) and the constraint \( \sum_{i=d+1}^{k+1} f_i = 1 \). The global maximum which is one of the solutions of the equation set (10) should be chosen as the final solution.

A sequence of non-increasing upper bounds detailed in [16], results by relaxing the specialization to \( l = 0 \) in (5).

b. A Lower Bound on Capacity

To derive a lower bound on capacity we restrict our attention to i.i.d. input sequences \( \{x_i\} \) that satisfy the \((d, k)\) constraints. The mutual information

\[
I(N; X^N) = \frac{1}{N} H(X^N) - H(\varepsilon)
\]

is lower bounded noting that

\[
\frac{1}{N} H(X^N) = \frac{1}{N} \sum_{i=1}^{N} H(y_i | X_i^{i-1}) \]

which results by introducing additional conditions \( (X_i^{i-1}) \) into the conditional entropy function.

The lower bound

\[
C_{PSC} \geq C_{Lo} = \frac{H(x+\varepsilon) - H(\varepsilon)}{E(x)}
\]

follows the observation \( H(y_i | X_i^{i-1}, X_i^{i-1}) = H(y_i | \varepsilon_i) \) where the index \( i \) was omitted for the sake of brevity. The tightest lower bound within this class is found by \( C_{Lo} = \sup(C_{Lo}) \), where the supremum is carried over all \( P(x) \) satisfying the \((d, k)\) constraint. Note that for a deterministic PSC that is for which \( H(\varepsilon) = 0 \) the supremum is achieved by the geometric distribution

\[
\text{Prob}(x=l) = \gamma^{-l} \quad l = d+1, d+2, \ldots, k+1
\]

where \( \gamma \) is determined uniquely by the equation \( \sum_{i=d+1}^{k+1} \gamma^i = 1 \). This specific distribution is known to achieve \( C_* \) [6] the noiseless channel (that is \( P(x) = 1 \)) capacity which is given by: \( C_* \) = log \( \gamma \). A sequence of nondecreasing lower bounds which are tighter than \( C_{Lo} \) is found in [16] by using input sequences \( \{x_i\} \) which are first order Markov satisfying the \((d, k)\) constraints.

c. Zero Error Capacity

A simple lower bound on the zero-error capacity \( C_0 \) of the PSC is established by using i. i. d phrases the inter-distance-length span of which is 3 in the evaluation of the lower bound \( C_{Lo} \). In other words the probability function \( P(x) \) used to evaluate \( C_{Lo} \) (13) is nonzero for \( x \) – the phrase length, taking on the values \( d = 1, d+4, d+7, \ldots \) up to \( k+1 \), since for this specific distribution \( x + \varepsilon \) determines uniquely \( x \) and \( \varepsilon \) and therefore \( H(x + \varepsilon) = H(x) + H(\varepsilon) \), which upon substitution in (13) yields the optimized lower bound.
on the zero error capacity $C_o$. The supremum in (15) taken over all possible probability functions $P(x)$ where $x$ takes on values in $(d+1, d+4, d+7, \ldots, d+1+3 \left\lfloor k_d \right\rfloor)$ (with subscript $[\cdot]$ standing for the closest integer from below) yields,

$$C_{oL} = \log \lambda,$$

(16)

where $\lambda$ is uniquely determined by the equation

$$\sum_{i=0}^{\left\lfloor k_d \right\rfloor} \lambda - (d+i+1) = 1.$$

IV. NUMERICAL RESULTS AND SOME PROPERTIES OF THE BOUNDS

The upper bound $C_{uo}$ (7) and the optimized lower bound $C_{Lo,opt}$ (13) are depicted in Figures 3-5. In Figure 3 the (2,5) codes are examined while in Figures 4 and 5 the (2,7) codes are discussed. The bounds in Figures 3-5 are shown versus the asymmetry factor $P_{as}$ with the no shift probability $P_0$ used as a parameter. The code capacity $C_w$ and the lower bound on the zero-error capacity $C_{oL}$ (16) are also shown. Note that the upper bound $C_{uo}$ is symmetric with respect to $P_{as}$ while the lower bound $C_{Lo,opt}$ is not. This is attributed to the fact that the probability function of $v - P_0(k)$ (8), which rises in the evaluation of $C_{uo}$ is symmetric with respect to $P_{as}$. However, $P(\epsilon)$ (2) the probability function of $\epsilon$ which is intimately involved in determining the expression for the lower bound $C_{Lo}$ (13) is definitely not symmetric with respect to $P_{as}$ which stands explicitly for the asymmetry factor. This clearly leads to the observed fact that the one dimensional probability $P(x)$ of the phrase lengths that maximizes $C_{Lo}$ is also asymmetric with respect to the average phrase length. The capacity $C_{PSC}$ itself is a symmetric function of $P_{as}$ as concluded in [16] by reversing the time axis at the channel input. This argumentation enables to tighten the lower bounds by taking the max $[C_{Lo}(P_{as}), C_{Lo}(-P_{as})]$ as the improved bound, where $C_{Lo}(P_{as})$ is used to explicitly denote the functional dependence of $C_{Lo}$ on $P_{as}$. Note that the optimized lower bound $C_{Lo,opt}$ is always not smaller than $C_{as}$ as is evident by the derivation techniques.

Observe that the upper bound $C_{uo}$ is not necessarily a convex function of $P_{as}$ while the lower bound $C_{Lo}$ is a property demonstrated in Figures 3-5. The upper and lower bounds examined here coincide $C_{Lo} = C_{uo} = C_*$ to give the exact capacity in cases of no channel impairment present ($P_{as} = 1$) or a deterministic impairment ($P_0 = 0$ and $P_{as} = \pm 1/2$) which evidently causes no degradation in capacity.

V. DISCUSSION AND CONCLUSIONS

A simple channel model has been introduced in an effort to account for the single position peak (bit) shifts which were identified as one of the major error generating mechanisms in a certain class of magnetic recording systems [1]-[5]. The effect of this channel on the commonly used $(d,k)$ codes with $d \geq 2$, is conveniently characterized in terms of phrase lengths. The peak shifts, modeled here by i.i.d. ternary sequence $\{E_i\}$ (1), (2), either expand or shrink the respective phrases imposing therefore an inherent correlation. This correlation manifests itself in the proposed model in terms of the channel memory.

A simple upper bound $C_{uo}$ (7) and a lower bound $C_{Lo}$ (13) on the PSC capacity were evaluated and explored for interesting values of the parameters $P_0$ - the probability of no shift, and $P_{as}$ - the asymmetry factor (2). A lower bound $C_{oL}$ (16) on the zero error capacity is given as well. In [16] sequences of improved lower and upper bounds on the PSC capacity are evaluated which turn out to be more complicated as compared to the elementary bounds presented here.

A single peak shift causes evidently at most two consecutive errors, thus codes with short burst correction properties may turn efficient in this scenario. Codes having short-burst error correction capabilities [10] are a natural selection to be used in case. Block codes designed specifically to completely recover a single of few single position bit shifts within the encoded block have recently been reported [11]-[12].

The basic PSC models introduced here, is easily extendible to account for a multi-position peak (bit) shifts. However, the $d$ parameter in the $(d,k)$ run length limited input, should be chosen not less than twice the maximal span of the allowed peak shift, to guarantee simple analysis [16].

The channel model is further extended to account for errors generated by a peak shift mechanism as well as in a random fashion. This is accomplished [16] by connecting in tandem the proposed peak shift channel (PSC) followed by the binary symmetric channel (BSC). Lower and upper bounds on the capacity of the concatenated channel (abbreviated in [16] by PSC/BSC) with $(d,k)$ constrained inputs are reported in [16].

ACKNOWLEDGEMENT

The authors are grateful to J.K. Wolf for interesting discussions and to P.H. Siegel and to K.A.S. Immink for mentioning some relevant references.

REFERENCES


