A New Class of Error-Pattern-Correcting Codes Capable of Handling Multiple Error Occurrences

Jihoon Park and Jaekyun Moon, *Fellow, IEEE*

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA

A new class of high-rate error-pattern-correcting cyclic codes that correct most single occurrences of target error patterns and a significant portion of their multiple occurrences is proposed. This code is based on first designing a generator polynomial of the lowest possible degree that is tailored to any single occurrence of target error patterns. The generator polynomial is then multiplied by a primitive polynomial so that a cyclic code based on a higher degree generator polynomial can handle all single occurrences of the target error patterns, as well as some highly probable multiple occurrences, using the captured syndrome and reliability measure. A performance comparison is provided for a jitter-dominant perpendicular recording channel.

***Index Terms*—Cyclic codes, dominant error patterns, generator polynomial, syndrome sets.**

I. Introduction

CONVENTIONAL burst-error-correcting cyclic codes are designed to guarantee correction of any single burst-error of length t or less within a received codeword. In contrast to t-random-error-correcting codes, t-burst-error-correcting cyclic codes are not designed based on minimum distance.

In interference-dominant channels, errors tend to occur in specific patterns [1], [2]. While existing burst-correcting codes can algebraically correct these frequently observed, dominant error patterns of length t or less, the correction power is not effectively utilized when some of the target error patterns have very long lengths, or the number of target error patterns is relatively small (but their occurrence frequencies are significant). We have previously constructed a cyclic code that specifically targets a set of L dominant error patterns that make up a very large percentage of all observed occurrences of errors, by constructing a low-degree generator polynomial that produces distinct, nonoverlapping syndrome sets for all L target error patterns [3]. The resulting code is effective in correcting single occurrences of the target patterns within the codeword length.

In this paper, we introduce a new approach to constructing high-rate error-pattern control codes that aims at correcting most single occurrences of L target error patterns and further correcting a significant portion of their multiple occurrences. The new approach is based on combining the low-degree generator polynomial obtained by the method of [3] with a primitive polynomial. The resulting generator polynomial produces not only the distinct syndrome sets for all single occurrences of the target error patterns, but also extra syndrome sets for identifying some of the highly probable multiple occurrences. We show that in addition to the ability to handle certain multiple error pattern occurrences, the probability of miscorrection for single pattern occurrences is further reduced with the present code.

II. Error-Pattern-Correcting Codes Capable of Handling Multiple Error Occurrences

A. Cyclic Codes Targeting a Single Error Occurrence

Let $e_i(x)'\text{s}, i = 1,\ldots,L$ be the targeted, dominant error patterns in the form of polynomials over GF(2). Also, let $p_k(x)'\text{s}, k = 1,\ldots,K$ be the irreducible polynomial factors making up all $e_i(x)$'s. It has been shown in [3] that if the greatest common divisors (GCDs) between a generator polynomial $g(x)$ and $e_i(x)$'s are all different, then the corresponding syndrome sets are guaranteed to be distinct among different $e_i(x)$'s. A syndrome set here refers to the sequence of syndromes that gets generated as the original captured syndrome feeds through the feedback shift register whose connection weights are specified by the given cyclic code generator polynomial. As such, all syndromes in a given syndrome set point to the same error pattern (the syndromes in a given set correspond to all cyclic shifts of the given error pattern). A fairly low-degree $g(x)$ can be obtained from the general form

$$g(x) = p_1^{\gamma_1}(x)p_2^{\gamma_2}(x)\cdots p_K^{\gamma_K}(x) \tag{1}$$

that produces distinct syndrome sets for all L target error polynomials. The search procedure involves increasing each $\gamma_k, k = 1,\ldots,K$, from zero while checking for the appropriate conditions including having distinct GCDs [3].

As an example of a cyclic code design for perpendicular magnetic recording, a hyperbolic tangent transition response is assumed, with an equalizer target response of $1+D$ at a channel density of 1.4. The density is defined as the ratio of the width over -50% to 50% of the transition response's saturation level to the user bit period. The mixed noise contains 10% additive white Gaussian noise (AWGN) and 90% jitter noise. For the signal-dependent noise environment, a pattern-dependent noise predictor (PDNP) [4] with one noise prediction tap per branch is used as the detector on the four-state trellis. This particular detector essentially provides the best detection performance for the assumed channel.

It has been observed that ten error patterns of lengths up to ten make up 99.7577% of the observed error patterns, at a bit error rate (BER) of 2.3276×10^{-3}. The ten target error patterns are $1, (1+x), (1+x+x^2), (1+x)^3, (1+x+x^2+x^3+x^4), (1+x)(1+x+x^2)^2, (1+x+x^3)(1+x^2+x^3), (1+x)^7, (1+x+x^2)(1+x^3+x^6)$, and $(1+x)(1+x+x^2+x^3+x^4)^2$ in the form of binary polynomials. It is seen that there are six irreducible polynomial factors: $p_1(x) = (1+x), p_2(x) = (1+x+x^2), p_3(x) = (1+x+x^2+x^3+x^4), p_4(x) = (1+x+x^3), p_5(x) = (1+x^2+x^3)$, and $p_6(x) = (1+x^3+x^6)$. Through the search procedure, it

Digital Object Identifier 10.1109/TMAG.2007.893478

is found that a generator polynomial $g(x)$ that can produce ten different syndrome sets for the ten target error patterns is

$$\begin{aligned} g(x) &= p_1^2(x)p_2^1(x)p_3^1(x)p_4^0(x)p_5^0(x)p_6^0(x) \\ &= 1 + x^3 + x^5 + x^8. \end{aligned}$$

With this $g(x)$ of order 30, an extended $(30s, 30s-8)$ cyclic code can be constructed for any positive integer s, by simply applying the same $g(x)$ to a larger input block of length $30s-8$. This allows a wide range of codeword lengths, depending on the applications. In this extended code consisting of s sub-blocks, a syndrome set repeats itself over the entire length of the code because of the cyclic property. Thus, the captured syndrome, while still pointing to a particular error pattern in the target list, now points to a number of possible starting positions of that error pattern. The decision on the error position is based on the comparison of reliability measures of the possible error event starting positions.

B. Error-Pattern-Correcting Cyclic Codes Capable of Handling Multiple Error Occurrences

While a higher code rate can be achieved with a larger s, multiple error occurrences of target error patterns would limit the code performance, as a $(ps, ps-r)$ cyclic code, based on a degree-r $g(x)$ of order p, does not provide a correction capability for the multiple error occurrences.

Most multiple occurrences are mistaken as a single target error pattern, and subsequently miscorrection is made due to the wrong error-type decision. Miscorrection often creates a "mirror-image error pattern" that consists of two identical error patterns occurring at the same position in two different sub-blocks. The resulting syndrome becomes zero, so the presence of the mirror-image error pattern cannot be detected. Consequently, no attempt for correction is made.

Given a degree-r $g(x)$ of order p that is tailored to L target error polynomials, consider the following polynomial:

$$g'(x) = g(x)p'(x) \tag{2}$$

where $p'(x)$ is a degree-m primitive polynomial that is not a factor of any L target error polynomials. Then, the order p' of $g'(x)$ is the least common multiple (LCM) of p and $(2^m - 1)$. With the higher-degree generator polynomial $g'(x)$, we now have a $(p', p'-r-m)$ cyclic code. A unique mapping between the syndrome sets and the target error patterns is preserved as with $g(x)$, but now that $g(x)$ is multiplied by $p'(x)$, the periods of new syndrome sets are increased by a factor equal to (2^m-1).

For comparison, a generator polynomial for the t-burst-error-correcting Fire code is given by $g_f(x) = (x^{2t-1}+1)p_f(x)$ [5]. Here, $p_f(x)$ is a degree-m primitive polynomial, and the degree m should be greater than or equal to t. Given t, the least redundancy is obtained in the case of $m = t$. The order p_f of $g_f(x)$ is the LCM of $(2t-1)$ and (2^m-1). As a result, a $(p_f, p_f - 2t + 1 - m)$ Fire code can be constructed, and the Fire code can algebraically correct any single burst-error of length t or less. While a factor $(x^{2t-1}+1)$ in $g_f(x)$ is determined only by the maximum length t of burst errors to be corrected, irrespective of dominant error polynomials, in our approach, $g(x)$ in $g'(x)$ is systematically constructed, guaranteeing distinct syndrome sets for any given set of L target error polynomials.

The probability that a mirror-image error pattern will occur in a $(p', p'-r-m)$ cyclic code based on $g'(x)$ can be zero, since the code consists of only one sub-block. Therefore, it is possible to recognize a significant portion of miscorrection by the syndrome re-check. Moreover, whether we have single or multiple error occurrences can also be determined by the syndrome re-check. For a single occurrence, the revised syndrome for the decoded codeword must be zero; otherwise, we know there are multiple error occurrences.

Besides the L syndrome sets for identifying any single occurrence of L target error patterns, many extra syndrome sets are produced by $g'(x)$. This is because the total number of syndrome sets generated by $g'(x)$ is always greater than L. Accordingly, the extra syndrome sets can be mapped to highly probable double-error-pattern events, i.e., $x^\mu[e_i(x) + x^\rho e_j(x)]$ for target error polynomials $e_i(x)$ and $e_j(x), i, j = 1, \ldots, L$, and all possible values of μ and ρ.

Note that while $g(x)$ consists of irreducible factors that appear in L target error polynomials, $p'(x)$ is not a factor of any L target error polynomials.

The decoding strategy is as follows: if the syndrome matches one or more highly probable double-error-pattern events, then an attempt is made to correct $x^\mu e_i(x)$ based on reliability measures for possible $e_i(x)$'s among target error polynomials, and possible μ's for each $e_i(x)$. The syndrome re-check confirms whether the correction is successful: The revised syndrome should be among the L syndrome sets for any single error occurrence. Afterwards, the remaining error polynomial $x^{\mu+\rho}e_j(x)$ is corrected, either algebraically without any miscorrection, or based on reliability measures of only a few possible error positions with a high probability of accuracy.

We have obtained $g(x) = 1+x^3+x^5+x^8$ that guarantees ten distinct, nonoverlapping syndrome sets for the ten target error patterns of lengths up to ten. From this $g(x)$, we construct $g'(x)$ with a degree-6 primitive polynomial $p'(x) = 1 + x + x^6$ of order 63 as

$$g'(x) = g(x)p'(x) = (1 + x^3 + x^5 + x^8)(1 + x + x^6).$$

Since the order p' of $g'(x)$ is 630, a (630, 616) cyclic code is constructed, and its code rate is 0.9778.

It turns out that four target error polynomials $1, (1+x+x^2)$, $(1+x+x^3)(1+x^2+x^3)$ and $(1+x+x^2)(1+x^3+x^6)$ can be algebraically corrected without any miscorrection. This is because the periods of the corresponding syndrome sets are equal to the codeword length, and a syndrome element in each syndrome set indicates the exact error event starting position.

While syndrome sets for the remaining six target error polynomials do not have the same periods as the codeword length, the number of possible error positions is substantially reduced. For the target polynomial $(1+x)(1+x+x^2)^2$, there can be a maximum of $126 (= 630/5)$ possible error positions with a (630, 622) cyclic code based on $g(x)$ and $s = 21$, but, at most $2 (= 630/315)$ possible error positions, one of which is an actual error starting position, are obtained with the (630, 616) cyclic code based on $g'(x)$. Therefore, the probability of miscorrection can be reduced by $1/(2^6-1)$.

Since this $g'(x)$ can produce a total of 71 syndrome sets, and ten syndrome sets are assigned for any single occurrence of the ten target error patterns, there are 61 extra syndrome sets. These

TABLE I
TARGETED, HIGHLY PROBABLE DOUBLE ERROR OCCURRENCES AND THEIR IDENTIFICATION RATES BASED ON 61 EXTRA SYNDROME SETS

Targeted double error event	Identification rate
$x^{\mu}[(1+x+x^2)+x^{\rho}(1+x)]$	80.45 %
$x^{\mu}[(1+x+x^2)+x^{\rho}(1+x)^3]$	74.44 %
$x^{\mu}[(1+x)^3+x^{\rho}(1+x)]$	68.86 %
$x^{\mu}[(1+x+x^2)+x^{\rho}(1+x+x^2)]$	80.42 %
$x^{\mu}[(1+x+x^2)+x^{\rho}]$	74.08 %
$x^{\mu}[(1+x+x^2+x^3+x^4)+x^{\rho}(1+x+x^2)]$	67.15 %
$x^{\mu}[(1+x)^3+x^{\rho}]$	67.95 %
$x^{\mu}[(1+x+x^2+x^3+x^4)+x^{\rho}(1+x)]$	39.71 %
$x^{\mu}[(1+x)+x^{\rho}]$	55.27 %
$x^{\mu}[(1+x)+x^{\rho}(1+x)]$	92.48 %
$x^{\mu}[(1+x+x^2+x^3+x^4)+x^{\rho}(1+x)^3]$	40.00 %
$x^{\mu}[(1+x)^3+x^{\rho}(1+x)^3]$	79.87 %

61 extra syndrome sets are utilized to recognize highly probable double occurrences.

Table I lists 12 targeted, highly probable double error events (78.14% of the observed 2 to 7 pattern occurrences) and corresponding identification rates based on the 61 extra syndrome sets. The identification rate is given by the ratio of the number of recognizable $\{\mu, \rho\}$ pairs to the number of all possible $\{\mu, \rho\}$ pairs within a 630-bit block. As an example, if the syndrome is 15 104 in decimals, then $x^{\mu}[(1+x+x^2)+x^{\rho}(1+x)]$ can be recognized with 40 possible μ's, e.g., 0, 26, 40, 46 ..., and 2 possible ρ's for each μ, e.g., 4 and 319 for $\mu = 0$. The most likely error starting position, conditioned on the possible μ's for $x^{\mu}(1+x+x^2)$, can be found using the soft metric with the modification to incorporate a PDNP [3]. Once syndrome re-check confirms the correction, $x^{\mu+\rho}(1+x)$ is then corrected by the revised syndrome.

We finally note that certain double error events may have a higher probability of occurrences than some of the single events. In this case, it is possible to allocate more syndrome sets for correcting double error pattern occurrences at the expense of reduced single error pattern correction capability.

III. PERFORMANCE EVALUATION

For performance comparison, we compute the sector error rate (SER), under the assumption that an outer t-symbol-correcting Reed-Solomon (RS) code is applied. The SER computation is based on the block multinomial distribution, utilizing estimated probabilities of symbol error events of various weights within a codeword [6] as well as a Gaussian tail extrapolation of the captured statistics. To compute the SER for a 512-information-byte sector, a $(410+2t, 410, t)$ shortened RS code over $\mathrm{GF}(2^{10})$ is considered without interleaving.

Fig. 1 compares the SERs of the proposed schemes, i.e., the (630, 622) cyclic code based on $g(x) = 1 + x^3 + x^5 + x^8$ and the (630, 616) cyclic code based on $g'(x) = g(x)(1 + x + x^6)$, at a fixed overall user density of $D'_u = 1.247$, where outer RS codes with varying byte-error correction capabilities are considered. The dotted lines shown in the high SER region indicate the SERs based on the direct counts. The overall user density D'_u is defined as $D'_u \triangleq D_s \times R \times R'$, where D_s is the channel density, and R and R' are the rates of the inner code and outer RS code, respectively. The signal-to-noise ratio (SNR) has been defined as the energy of the first derivative of the transition response E_{dt} to the noise spectral density N_{α}, which signifies $\alpha\%$ jitter noise [3]. The proposed schemes respectively achieve SNR gains of 0.12 and 0.36 dB at SER $= 10^{-8}$, relative to a stronger RS code without an inner code.

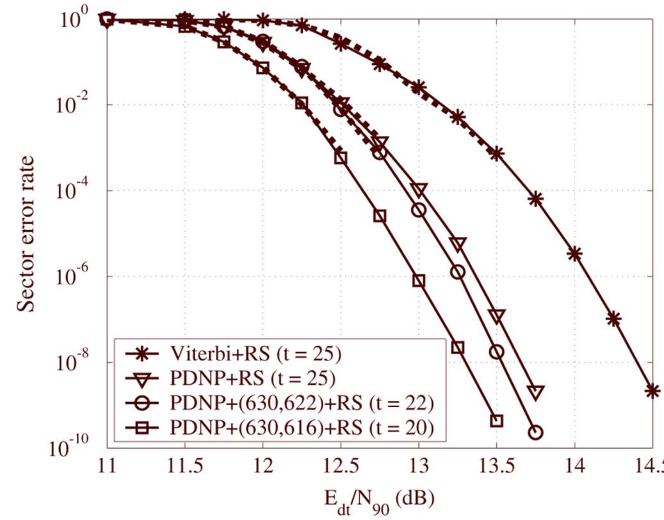

Fig. 1. Sector error rates at a fixed user density of 1.247

IV. CONCLUSION

A new class of high-rate single/multiple-error-pattern-correcting cyclic codes is developed that provides an efficient correction capability for a specific set of target error patterns. The generator polynomial tailored to the target error patterns produces distinct syndrome sets for any single occurrence of the target error patterns, and also for highly probable double error occurrences. When applied to a jitter-dominant channel, the performance gains become significant, with a marked reduction in miscorrection for single error pattern occurrences.

REFERENCES

[1] A. Patapoutian, B. Shen, and P. A. McEwen, "Event error control codes and their applications," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2595–2603, Sep. 2001.
[2] P. Chaichanavong and P. H. Siegel, "Tensor-product parity code for magnetic recording," *IEEE Trans. Magn.*, vol. 42, no. 2, pp. 350–352, Feb. 2006.
[3] J. Park and J. Moon, "Cyclic codes tailored to a known set of error patterns," in *Proc. IEEE GLOBECOM*, Nov. 2006, pp. 1–5, GEN03-4.
[4] J. Moon and J. Park, "Pattern-dependent noise prediction in signal-dependent noise," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 4, pp. 730–743, Apr. 2001.
[5] P. Fire, "A class of multiple-error-correcting binary codes for non-independent errors," Sylvania Electronic Defence Lab., Mountain View, Sylvania Rep. RSL-E-2, Mar. 1959.
[6] Z. A. Keirn, V. Y. Krachkovsky, E. F. Haratsch, and H. Burger, "Use of redundant bits for magnetic recording: Single-parity codes and Reed–Solomon error-correcting code," *IEEE Trans. Magn.*, vol. 40, no. 1, pp. 225–230, Jan. 2004.

Manuscript received February 16, 2007 (e-mail: jhpark@umn.edu).