Optimal, Systematic, q-Ary Codes Correcting All Asymmetric and Symmetric Errors of Limited Magnitude

Noha Elarief and Bella Bose, *Fellow, IEEE*

***Abstract*—Systematic q-ary ($q > 2$) codes capable of correcting all asymmetric errors of maximum magnitude l, where $l \leq q - 2$, are given. These codes are shown to be optimal. Further, simple encoding/decoding algorithms are described. The proposed code can be modified to design codes correcting all symmetric errors of maximum magnitude l, where $l \leq \frac{q-2}{2}$.**

***Index Terms*—q-ary codes, asymmetric channels, error control, error correcting codes, limited magnitude error.**

I. Introduction

Classical error control codes have been designed under the assumption of binary symmetric errors, i.e., both $1 \rightarrow 0$ and $0 \rightarrow 1$ errors can occur during transmission. Nevertheless, errors in some VLSI and optical systems are asymmetric in nature [5], [9]. For example, in VLSI circuits and memories, charges may leak with time but new charges will not be added. Thus, a suitable channel model for such systems is the binary asymmetric channel (Z-channel) where errors are of one dominant type known *a priori*, say $1 \rightarrow 0$ errors. In [10], Varshamov introduced the q-ary asymmetric channel where the channel's input/output symbols are over the alphabet $Q = \{0, 1, \ldots, q-1\}$. Moreover, such channel has the property that, when a symbol $a \in Q$ is transmitted, the corresponding received symbol is in the set $\{0, 1, \ldots, a\}$, assuming a decreasing error.

Similar to the asymmetric channel is the unidirectional channel; the difference is that the type of error is not known *a priori*. Detailed studies of asymmetric/unidirectional error control codes can be found in [3] and [8]. Not until recently has the notion of limited magnitude asymmetric errors been introduced [2]; we say that a vector $(x_{n-1}, x_{n-2}, \ldots, x_0)$ over Z_q suffers an asymmetric error of maximum magnitude/level $l \leq q - 1$ if and only if the corresponding channel output $(x'_{n-1}, x'_{n-2}, \ldots, x'_0)$ is such that $x'_i \in \{x_i - l, x_i - l + 1, \ldots, x_i\}$, with $x'_i \in Q$. Fig. 1 illustrates the difference between the traditional q-ary asymmetric channel and the q-ary asymmetric channel with $l = 1$. In [4], an interesting application for this special case of q-ary asymmetric channel was pointed out: multilevel flash memories. Unlike traditional single-level flash memories where each cell stores only one bit, multilevel flash memories achieve higher storage capacities and thus lower manufacturing costs by programming the cells into one of $q > 2$ threshold voltage thereby storing $\log_2 q$ bits per cell. Nevertheless, increasing the number of threshold levels imposes an important challenge [6]: the voltage difference between states is narrowed since – technically – the voltage window is limited. A natural consequence is that reliability issues such as low data retention and read/write disturbs become more significant [4]; errors in such cases are typically in one dominant direction and of limited magnitude.

In [1], the authors introduced codes capable of correcting all asymmetric errors of limited magnitude l (or l-AEC codes for short). However, the proposed codes are nonsystematic. A systematic code, where the information symbols are separated from the check symbols, is advantageous over a nonsystematic code because, in a systematic code, the data processing and encoding/decoding can be done in parallel. In this paper, we first give a bound on the number of check digits of a systematic l-AEC code. Then, we present a code that uses the minimum possible number of check symbols and is thus optimal. Furthermore, properties of the q-ary *symmetric* channel with level l are explored. We show that the code construction ideas used for l-AEC codes can be applied to design codes correcting all symmetric errors of maximum level l (l-SEC codes).

The rest of the paper is organized as follows. In Section II, we present a background necessary to tackle the problem. In Section III, we give a lower bound on the number of check symbols needed to encode an l-AEC code. A code achieving this bound is proposed in Section IV. We extend the results for l-SEC codes in Section V. Finally, concluding remarks are given in Section VI.

II. Preliminaries

It turns out that the knowledge of the maximum error level gives nice properties that can be used in the design of error correcting codes. We start by introducing a well-known distance metric capturing these properties as mentioned in [1]

Manuscript received August 23, 2008; revised September 09, 2009. Current version published March 10, 2010. This work is supported by the NSF by Grants CCF-0701452 and CCF-0728810.

N. Elarief is with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331 USA (e-mail: elarief@eecs.orst.edu).

B. Bose is with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331 USA, and also with King Saud University, Riyadh 11451, Saudi Arabia (e-mail: bose@eecs.orst.edu).

Communicated by G. Seroussi, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2009.2039065

0018-9448/$26.00 © 2010 IEEE

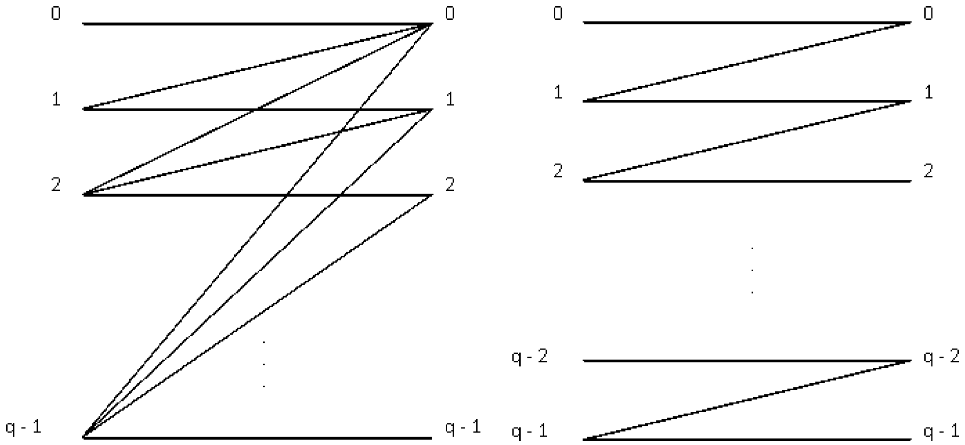

Fig. 1. q-ary asymmetric channel (left) versus q-ary asymmetric channel with level 1 (right).

Definition 1: Let $x = (x_{n-1}, x_{n-2}, \ldots, x_0)$ and $y = (y_{n-1}, y_{n-2}, \ldots, y_0)$ be two vectors over Z_q, then the distance between x and y is defined as

$$d(x,y) = max\{|x_i - y_i| : i \in \{0,1,\ldots,n-1\}\}.$$

It can easily be seen that $d(x,y) \le q-1$.

The following theorem gives the necessary and sufficient conditions on the minimum distance of an l-AEC code.

Theorem 1: **[1]:** A code $C \subseteq Q^n$ is an l-AEC code if and only if, for all distinct codewords, $x, y \in C$, $d(x,y) \ge l+1$.

Let $A_a(n,l,q)$ denote the maximum number of codewords in a q-ary l-AEC code of length n. A bound on $A_a(n,l,q)$ and a nonsystematic l-AEC code achieving this bound are given in [1]:

Theorem 2: **[1]:** $\forall n \in \{1,2,3\ldots\}$ and $l \in Q$, $A_a(n,l,q) = \left\lceil \frac{q}{l+1} \right\rceil^n$.

Theorem 3: **[1]:** Let C be the code of length n over Z_q defined as

$$C = \{(x_{n-1}, x_{n-2}, \ldots, x_0) : x_i \equiv 0 \pmod{(l+1)} \quad \forall i \in \{0,1,\ldots,n-1\}\}.$$

Then, C is an l-AEC code with $\left\lceil \frac{q}{l+1} \right\rceil^n$ codewords.

Finally, for further analysis, $x \pmod a$ denotes the component-wise remainder of a vector x when divided by an integer a.

III. A Lower Bound on the Number of Check Digits

In the following theorem, we investigate the minimum number of check digits needed to encode information vectors of a certain length.

Theorem 4: Let C be a systematic q-ary l-AEC code, such that the number of information digits in a codeword is k. Then, the number of check digits, r, satisfies the following condition:

$$r \ge \frac{k \times \log(l+1)}{\log \left\lceil \frac{q}{l+1} \right\rceil}.$$

Proof: Consider the subset of information vectors

$$V = \{(x_{k-1}, x_{k-2}, \ldots, x_0) : 0 \le x_i \le l, \forall i \in \{0,1,\ldots,k-1\}\}.$$

Vectors of V can be viewed as the set of all vectors of length k over Z_{l+1}. Hence, $\forall x, y \in V$, $d(x,y) \le l$, and $|V| = (l+1)^k$. Therefore, by Theorem 1, the checks assigned to vectors in V must be at least $l+1$ apart for errors to be successfully corrected. Theorem 1 and Theorem 2 together give a bound on the number of vectors satisfying such criterion and we get

$$\left\lceil \frac{q}{l+1} \right\rceil^r \ge (l+1)^k.$$

Taking the log on both sides of the above inequality we get the desired property. □

One important implication of the above theorem is that it is not possible to design a systematic code correcting all errors of maximum magnitude l when $l = q-1$, since

$$\frac{k \times \log(l+1)}{\log \left\lceil \frac{q}{l+1} \right\rceil}$$

goes to infinity in this case. Therefore, for the rest of the paper, we assume $l \le q-2$.

IV. An Optimal Systematic l-AEC Code

In this section, codes which require exactly

$$r = \left\lceil \frac{k \times \log(l+1)}{\log \left\lceil \frac{q}{l+1} \right\rceil} \right\rceil$$

check digits are presented.

A. Encoding Algorithm

Input: The information vector: $(x_{k-1}, x_{k-2}, \ldots, x_0)$
Output: The encoded vector:

$$x = (x_{k-1}, x_{k-2}, \ldots, x_0, c_{r-1}, c_{r-2}, \ldots, c_0)$$

1) Compute

$$a = y_{k-1}(l+1)^{k-1} + y_{k-2}(l+1)^{k-2} + \cdots + y_0(l+1)^0$$

where $y_i = x_i \pmod{(l+1)}, i \in \{0, 1, \ldots, k-1\}$. That is, a is the number with radix $(l+1)$ representation of $(x_{k-1}, x_{k-2}, \ldots, x_0) \pmod{(l+1)}$.

2) Represent a in radix $\left\lceil \frac{q}{l+1} \right\rceil$ with r digits: $(a_{r-1}, a_{r-2}, \ldots, a_0)$.
3) Compute the check part: $c = (c_{r-1}, c_{r-2}, \ldots, c_0)$, where $c_i = (l+1)a_i, \forall i \in \{0, 1, \ldots, r-1\}$.
4) Output the encoded vector

$$x = (x_{k-1}, x_{k-2}, \ldots, x_0, c_{r-1}, c_{r-2}, \ldots, c_0)$$

Example 1: We encode the word (6, 2, 8, 1) over Z_{10}, assuming a maximum error level of 2. The number of check digits needed is

$$r = \left\lceil \frac{4 \log(2+1)}{\log \left\lceil \frac{10}{2+1} \right\rceil} \right\rceil = 4.$$

With notations as above, $a = 0 \times 3^3 + 2 \times 3^2 + 2 \times 3^1 + 1 \times 3^0 = 25$ and, thus, (0, 1, 2, 1) is the representation of a in base 4. Therefore, the encoded codeword is (6, 2, 8, 1, 0, 3, 6, 3).

Theorem 5: The above construction yields codewords of minimum distance $l+1$.

Proof: Given two distinct information vectors: $v_1 = (x_{k-1}, x_{k-2}, \ldots, x_0)$ and $v_2 = (y_{k-1}, y_{k-2}, \ldots, y_0)$, there are two possibilities: $v_1 \pmod{(l+1)} \neq v_2 \pmod{(l+1)}$ in which case each vector is assigned different check digits and, since the check digits are multiples of $(l+1)$, the distance between the resulting codewords is at least $l+1$. For the second case, $v_1 \pmod{(l+1)} = v_2 \pmod{(l+1)}$, v_1 and v_2 are assigned the same check digits. Nevertheless, by distinctness of v_1 and v_2, $\exists i \in \{0, 1, \ldots, k-1\}$ such that $x_i = y_i + m(l+1), m \geq 1$. Therefore, $d(v_1, v_2) \geq l+1$, and the resulting codewords satisfy the desired property. □

B. Decoding

Input: The channel output:

$$x' = (x'_{k-1}, x'_{k-2}, \ldots, x'_0, c'_{r-1}, c'_{r-2}, \ldots, c'_0)$$

Output: The recovered codeword:

$$(x_{k-1}, x_{k-2}, \ldots, x_0, c_{r-1}, c_{r-2}, \ldots, c_0)$$

1) Recover the check symbols $(c_{r-1}, c_{r-2}, \ldots, c_0)$ by rounding each received check symbol which is not multiple of $(l+1)$ upwards to the nearest multiple of $(l+1)$.
2) Compute

$$\begin{aligned} a =& \left(\frac{c_{r-1}}{l+1}\right) \left\lceil \frac{q}{l+1} \right\rceil^{r-1} \\ &+ \left(\frac{c_{r-2}}{l+1}\right) \left\lceil \frac{q}{l+1} \right\rceil^{r-2} \\ &+ \cdots + \left(\frac{c_0}{l+1}\right) \left\lceil \frac{q}{l+1} \right\rceil^{0} \end{aligned}$$

.

That is, a is the number with radix $\left\lceil \frac{q}{l+1} \right\rceil$ representation of $\left(\frac{c_{r-1}}{l+1}, \frac{c_{r-2}}{l+1}, \ldots, \frac{c_0}{l+1}\right)$:

3) Represent a in radix $(l+1)$ with k digits: $y = (y_{k-1}, y_{k-2}, \ldots, y_0)$.
4) Let $e_i = (y_i - x'_i) \pmod{(l+1)}$. The information symbols are: $(x_{k-1}, x_{k-2}, \ldots, x_0)$, such that $x_i = x'_i + e_i$, $i = 0, 1, \ldots, k-1$.
5) Output the codeword:

$$(x_{k-1}, x_{k-2}, \ldots, x_0, c_{r-1}, c_{r-2}, \ldots, c_0)$$

.

Example 2: Let the encoded word be as in example 4.1 and the channel output be $x' = (4, 2, 7, 1, 0, 3, 5, 1)$. Rounding the check symbols which are not multiple of 3 upwards to the nearest multiple of 3, we get (0, 3, 6, 3). As in Steps 2 and 3 of the algorithm, we compute $a = 0121_4 = 25$, and $y = (0, 2, 2, 1)$. Thus, the correct information symbols are (6, 2, 8, 1).

Theorem 6: Let x be a codeword encoded using the algorithm given in Section IV-A, and let x' be the l-asymmetric channel output. Then, the above decoding algorithm successfully recovers x.

Proof: Let $x = (x_{k-1}, x_{k-2}, \ldots, x_0, c_{r-1}, c_{r-2}, \ldots, c_0)$, then, by the channel properties

$$x' = (x'_{k-1}, x'_{k-2}, \ldots, x'_0, c'_{r-1}, c'_{r-2}, \ldots, c'_0)$$

is such that $x_i - l \leq x'_i \leq x_i$ and $c_i - l \leq c'_i \leq c_i$. Moreover, the encoding algorithm yields check symbols that are multiples of $l+1$, i.e., c'_i lies between two successive multiples of $l+1$. Therefore, the first step of the decoding algorithm successfully recovers the check symbols. It can also be seen that Steps 2 and 3 of the above algorithm do the reverse operations of Steps 1 and 2 of the encoding algorithm. Hence, with notations as above, $y = (y_{k-1}, y_{k-2}, \ldots, y_0) = (x_{k-1}, x_{k-2}, \ldots, x_0) mod (l+1)$. At Step 4, e_i can be seen as the magnitude of the error at the i^{th} information symbol, such that

$$x_i = x'_i + e_i \equiv y_i \pmod{(l+1)}.$$

Thus

$$e_i \equiv (y_i - x'_i) \pmod{(l+1)}.$$

Since the maximum error magnitude is l (i.e., $0 \leq e_i \leq l$) the value of e_i is successfully computed at Step 4 of the decoding algorithm, recovering the information symbols. □

V. An Optimal Systematic l-SEC Code

In this section, we explore some of the properties of the l limited magnitude symmetric error correcting codes. For a vector $x = (x_{n-1}, x_{n-2}, \ldots, x_0)$ over Z_q we say that the corresponding channel output $x' = (x'_{n-1}, x'_{n-2}, \ldots, x'_0)$ suffers a symmetric error of maximum magnitude l if and only if $x_i - e_i \leq x'_i \leq x_i + e_i$, where $0 \leq e_i \leq l, \forall i \in \{0, 1, \ldots, n\}$. As shown below, the similarity between the properties of l-SEC codes and l-AEC codes allows to extend the code construction idea we presented in the previous section to design a family of l-SEC codes.

Theorem 7: A code C is capable of correcting all symmetric errors of maximum magnitude l if and only if C has minimum distance $2l+1$.

Proof: Let S_x be the set of all words obtained from a codeword $x \in C$, where $|x| = n$, due to n or less symmetric errors of maximum magnitude l, i.e., $S_x = \{(x'_{n-1}, x'_{n-2}, \dots, x'_0) : x'_i = x_i \pm e_i, e_i \in \{0, 1 \dots, l\}\}$. Then, it is easy to see that, if C has minimum distance $2l+1$, then $\forall x, y \in C$, $S_x \cap S_y = \emptyset$. Therefore, C is an l-SEC code.

Conversely, if $\exists x, y \in C$ such that $d(x,y) \leq 2l$, then it is possible to obtain a word, say z, from both x and y due to symmetric errors of magnitude l or less. Hence, a decoder for C cannot correct z. Therefore, the minimum distance should be no less than $2l+1$ for C to be capable of correcting all symmetric errors of maximum magnitude l. □

The above theorem implies that, when $l > \frac{q-2}{2}$, any l-SEC code can have at most one codeword. Thus, we assume $l \leq \frac{q-2}{2}$.

Theorem 8: Let $A_s(n,l,q)$ denote the maximum number of words in a q-ary l-SEC code of length n. Then

$$A_s(n,l,q) = \left\lceil \frac{q}{2l+1} \right\rceil^n.$$

Proof: Similar to the proof of Theorem 2. □

Theorem 9: Let C be a code of length n over Z_q defined as

$$C = \{(x_{n-1}, x_{n-2}, \dots, x_0) : x_i \equiv 0 \pmod{(2l+1)} \quad \forall i \in \{0, 1, \dots, n-1\}\}.$$

Then, C is an l-SEC with $A_s(n,l,q)$ codewords.

Proof: Similar to the proof of Theorem 2. Digits of the channel output can be decoded in this case by rounding downwards or upwards (whichever is closer) to the nearest multiple of $2l+1$. □

Theorem 10: Let C be a systematic q-ary l-SEC code, such that the number of information digits in a codeword is k. Then, the number of check digits, r, satisfies the following condition:

$$r \geq \frac{k \times \log(2l+1)}{\log \left\lceil \frac{q}{2l+1} \right\rceil}.$$

Proof: Similar to the proof of Theorem 4. □

Now that we have identified the similarities between l-SEC and l-AEC codes, it can easily be seen that the proposed encoding/decoding algorithms for l-AEC codes can be modified in the following way to construct an l-SEC code. In both algorithms, simply replace all computations including the value $l+1$ with $2l+1$. Moreover, in Step 1 of the decoding algorithm, the check digits are recovered by rounding the received check digits either upwards or downwards to the nearest multiple of $2l+1$, whichever is closer. Finally, at Step 4 of the decoding algorithm, the information digits are recovered as $x_i = x'_i \pm e_i$, $\forall i \in \{0, 1, \dots, k-1\}$, where $0 \leq e_i \leq l$ such that $x_i \equiv y_i \pmod{(2l+1)}$.

TABLE I

q	l	$R_{non-systematic}$	$R_{systematic}$
8	1	0.66	0.5
	2	0.52	0.5
	3	0.33	0.33
	4	0.33	0.25
	5	0.33	0.25
	6	0.33	0.25
16	1	0.75	0.5
	2	0.64	0.5
	3	0.5	0.5
	4	0.5	0.33
	5	0.39	0.33
	6	0.39	0.33
	7	0.25	0.25
	8	0.25	0.25
	9	0.25	0.25
	10	0.25	0.25
	11	0.25	0.25
	12	0.25	0.25
	13	0.25	0.25
	14	0.25	0.25

VI. Conclusion

In this paper, we proposed a family of optimal codes correcting all q-ary asymmetric errors of limited magnitude l,where $l \leq q-2$. These codes are advantageous over other l-AEC codes given in the literature, namely in [1] and [4]. The code constructions given in [4] start with t symmetric error correcting codes over Z_{l+1} to construct codes that can correct t asymmetric errors of maximum magnitude l. It is known that no such code exists when $t = n$ since the minimum distance of any t symmetric error correcting codes is $2t+1$ [7]. Moreover, the codes given in [4] are not optimal in general which makes the use of the l-AEC codes presented in this paper more favorable as t approaches n.

As opposed to the l-AEC codes proposed in [1], our code construction is systematic; information symbols are separable from the check symbols resulting in faster encoding/decoding operations. Fortunately, the cost of having systematic code is low: we show that the rate of our l-AEC code is very close to the one given in [1]. The rate R of an error correcting code is given as $R = \frac{k}{n}$ where k is the number of information digits and n is the length of the code. For the code given in [1], the rate $R_{\text{nonsystematic}}$ is

$$\begin{aligned} R_{\text{nonsystematic}} &= \frac{\log_q \lceil \frac{q}{l+1} \rceil^n}{n} \\ &= \log_q \lceil \frac{q}{l+1} \rceil. \end{aligned}$$

The rate of the proposed code $R_{\text{systematic}}$ is

$$\begin{aligned} R_{\text{systematic}} &= \frac{k}{k+r} \\ &= \frac{k}{k + \left\lceil \frac{\log_q (l+1)}{\log_q \lceil \frac{q}{l+1} \rceil} \right\rceil} \end{aligned}$$

which can be approximated to

$$\begin{aligned} R_{\text{systematic}} &\approx \frac{1}{1+\frac{\log_q(l+1)}{\log_q \frac{q}{l+1}}} \\ &= \frac{\log_q \frac{q}{l+1}}{\log_q \frac{q}{l+1}+\log_q(l+1)} \\ &= \frac{\log_q \frac{q}{l+1}}{\log_q q-\log_q(l+1)+\log_q(l+1)} \\ &= \log_q \frac{q}{l+1}. \end{aligned}$$

Table I illustrates the closeness of the approximate values of $R_{\text{systematic}}$ and $R_{\text{nonsystematic}}$ for $q = 8$ and $q = 16$.

REFERENCES

[1] R. Ahlswede, H. Aydinian, L. H. Khachatrian, and L. M. G. M. Tolhuizen, "On q-ary codes correcting all unidirectional errors of a limited magnitude," in *Proc. 9th Int. Workshop on Algebr. Combinator. Coding Theory*, Kranevo, Bulgaria, Jun. 2004, pp. 20–26.

[2] R. Ahlswede, H. Aydinian, and L. H. Khachatrian, "Unidirectional error control codes and related combinatorial problems," in *Proc. 8th Int. Workshop on Algebr. Combinator. Coding Theory*, Russia, Sep. 2002, pp. 6–9.

[3] M. Blaum, *Codes for Detecting and Correcting Unidirectional Errors*. Los Alamitos, CA: IEEE Comput. Soc. Press, 1993.

[4] Y. Cassuto, M. Schwartz, V. Bohossian, and J. Bruck, Codes for Asymmetric Limited-Magnitude Errors with Application to Multi-Level Flash Memories Calif. Inst. Technol., Pasadena, CA, 2008, Tech. Rep. CaltechPARADISE:2008.ETR088.

[5] S. D. Constantin and T. R. N. Rao, "On the theory of binary asymmetric error correcting codes," *Inf. Computat.* , vol. 40, pp. 20–36, 1979.

[6] B. Eitan, R. Kazerounian, A. Roy, G. Crisenza, P. Cappelletti, and A. Modelli, "Multilevel flash cells and their trade-offs," in *Proc. Int. Electron Devices Meet.*, Dec. 1996, pp. 169–172.

[7] R. W. Hamming, "Error detecting and error correcting codes," *Bell Syst. Tech. J.*, vol. 29, no. 2, pp. 147–160, 1950.

[8] T. Klove, Error Correcting Codes for the Asymmetric Channel Univ. Bergen, Dep. Math., Bergen, Norway, 1995, Tech. Rep. 1809-0781.

[9] J. R. Pierce, "Optical channels: Practical limits with photon counting," *IEEE Trans. Commun.*, vol. 26, no. 12, pp. 1819–1821, Dec. 1978.

[10] R. R. Varshamov, "A class of codes for asymmetric channels and a problem from the additive theory of numbers," *IEEE Trans. Inf. Theory*, vol. 19, pp. 92–95, Jan. 1973.

Noha Elarief received the B.S. degree in computer science from Ain Shams University, Cairo, Egypt, in 2004, and the M.S. and Ph.D. degrees in computer science from Oregon State University, Corvallis, in 2008 and 2010, respectively.

She is currently a Post-Doctoral Researcher with the School of Electrical Engineering and Computer Science, Oregon State University. Her research interests include error-control codes, fault-tolerant computing, and computer networking.

Bella Bose (F'95) received the B.E. degree in electrical engineering from Madras University, India, in 1973, the M.E. degree in electrical engineering from the Indian Institute of Science, Bangalore, in 1975, and the M.S. and Ph.D. degrees in computer science and engineering from Southern Methodist University, Dallas, TX, in 1979 and 1980, respectively.

Since 1980, he has been with Oregon State University, Corvallis, where he is a Professor and the Associate Director of the School of Electrical Engineering and Computer Science. His current research interests include error-control codes, fault-tolerant computing, parallel processing, and computer networks.

Dr. Bose is a Fellow of the ACM.