

Enhanced Upper Bound for Erasure Recovery in SPC Product Codes

Ali Muqaibel

Single parity check (SPC) product codes are simple yet powerful codes that are used to correct errors and/or recover erasures. The focus of this paper is to evaluate the performance of such codes under erasure scenarios and to develop a closed-form tight upper bound for the post-decoding erasure rate. Closed-form exact expressions are derived for up to seven erasures. Previously published closed-form bounds assumed that all unrecoverable patterns should contain four erasures in a square. Additional non-square patterns are accounted for in the proposed expressions. The derived expressions are verified using exhaustive search. Eight or more erasures are accounted for by using a bound. The developed expressions improve the evaluation of the recoverability of SPC product codes without the need for simulation or search algorithms, whether exhaustive or novel.

Keywords: Erasure decoding, product codes, erasure recovery, single-parity-check codes, stopping set, binary erasure channel.

I. Introduction

Product codes were introduced by Elias in 1954 (see references in [1]). They represent the first practical method capable of achieving error-free coding with a nonzero code rate [2]. They are powerful codes that can be used to recover erasures or to correct errors. Erasures are encountered when transmission is assumed over a binary erasure channel (BEC) where the bits are never received in error. They are either received correctly or they are missing, that is, “erased.” Product codes are widely used to recover erasures in various types of data storage, such as redundant arrays of inexpensive disks (RAID) [3], [4] and CD and DVD standards [5]. In addition to optical and wireless applications, erasure recovery is important in real-time applications and in certain wireless network systems where packet information with a header is transmitted over a noisy channel, such as ATM cell wireless transmission [4], [6].

In the literature, most of the research work related to product codes and erasure recovery addresses the issue of code design and code performance evaluation. Designing a code for erasure recovery is theoretically equivalent to designing a code for error correction/detection but can be different when practical issues are considered [4]. In erasure-resilient code (ERC) design, the objective is to improve the recoverability without decreasing the code rate. In [4], the design of ERC that recovers all sets with two erasures or less is discussed. Three and more erasures are discussed in [3], [7].

Single-parity-check (SPC) product codes are the simplest form of two-dimensional product codes. A two-dimensional code consists of a data block, a parity check on the rows, a check on the columns, and a check on the checks as shown in Fig. 1. The two-dimensional encoding scheme allows many

Manuscript received Feb. 22, 2009; revised July 14, 2009; accepted July 29, 2009.

This work was supported by King Fahd University of Petroleum & Minerals (KFUPM), Saudi Arabia.

Ali Muqaibel (phone: +966 3 860 1595, email: muqaibel@kfupm.edu.sa) is with the Electrical Engineering Department, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia.

doi:10.4218/etrij.09.0109.0131

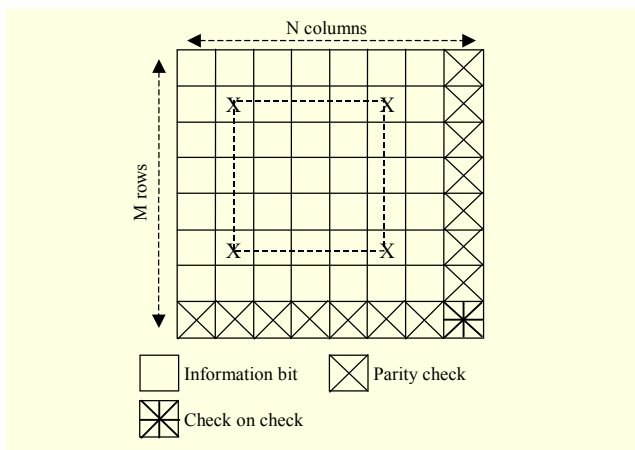


Fig. 1. Two-dimensional encoding matrix with unrecoverable four-erasure pattern.

patterns of lost cells to be recovered. This in turn results in significant reduction in the post-decoding cell loss rate.

An SPC product code (SPCPC) has a minimum distance, d_{\min} , of four; hence, using iterative decoding, all erasures of order three or less can be recovered. The code can recover many erasure patterns beyond those with the number of erasures determined by the minimum distance. The erasure recovery performance of product codes based on their minimum distance is unreliable. The weight enumerating function of the product code which fully specifies the performance of the code cannot be determined from the constituent codes without considering the specific interleaver or arrangement [1], [5]. Some researchers assume a random interleaver [8]. The weight enumerating function of most product codes remains unknown [1]. The iterative decoding continues over rows and columns until no progress can be made. The remaining set of erasures is called a stopping set [5], [7]. The performance of the codes depends on the number of stopping sets and their size. It is important to note that the concept of a stopping set for iterative decoding operating over BEC is different than the concept of stop sets for low-density parity-check (LDPC) codes [5].

In [6], developing exact forms for the post-decoding cell loss recovery for up to five erasures was considered. In [9], Kousa proposed a procedural technique to perform recoverability analysis for SPC product codes and to develop a tight bound on their performance. The idea in [9] is to define some special patterns of erasures called basic patterns from which many other patterns can be generated. Even if the basic pattern is unrecoverable, the generated patterns can have some recoverable cells. There are many conditions and steps to generate all the required patterns and enumerate them. In [10], the authors followed the same approach proposed by Kousa with some closed-form expressions to reduce the complexity of

the algorithm. The work in [9] and [10] is algorithmic in nature and not easy to follow. Moreover, the equations in [10] were derived for a specific matrix size and need to be extended before application to a larger matrix. Al-Shaikhi and Ilow [11], [12] avoided complexity and proposed a closed-form tight upper bound by analyzing the unrecoverable erasure patterns. Their analytical derivation was examined through computer simulation using Hamming and SPC product codes. However, they did not account for unrecoverable patterns which are not made up of the basic unrecoverable four-erasure pattern.

In this paper, we demonstrate that it is possible to have unrecoverable patterns which are not made up of the basic unrecoverable four-erasure pattern. This paper develops exact expressions, which account for all unrecoverable patterns for up to seven erasures. A bound is utilized to account for eight or more erasures. The derived expressions are verified through exhaustive search and supported by Monte-Carlo simulation.

The next section introduces the SPCPC model and its parameters. Performance analysis is detailed in section III. Verification of the analysis, results, and comparison with previous work are discussed next. The paper concludes by summarizing the main findings.

II. SPC Product Code Model

Cells are arranged into M rows and N columns. Parity-check cells are appended to the rows and columns. The resultant code is $(MN, (N-1)(M-1), 4)$. The code rate $(N-1)(M-1)/MN$ can be made close to unity by increasing M and N . Since the objective of this paper is recoverability analysis over BEC, it is assumed that every cell in the matrix has a header for error-detection and the detection process is perfect. Cells that are received in error and cannot be corrected are erased, and their location is stored in the recovery system.

The recovery process is repeated iteratively in rows and then columns until no cells can be recovered. The performance of the recovery system is a function of the following parameters:

- i : number of lost cells in the matrix before recovery
- U_i : number of patterns with i erasures and unrecoverable cells
- e_i : average number of unrecoverable cells
- p : cell loss probability
- P : post-decoding (recovery) cell loss probability

The post-decoding cell loss probability is given as in [9] by

$$P = \frac{1}{MN} \sum_{i=4}^{MN} e_i U_i p^i (1-p)^{MN-i}. \quad (1)$$

We will call a set of i lost cells i -erasures. The details for finding the size of the stopping set, e_i , and the number of

occurrences, U_i , for different values of i are discussed in the next section.

III. Performance Analysis

In this part, we derive closed-form expressions for e_i and U_i for up to seven erasures. An upper bound is developed for more than eight lost cells. In terms of minimum distance, $U_i=0$ for $i < d_{\min}-1$. The number of unrecoverable patterns with 1, 2, or 3 erasures is 0 because all erasures can be completely recovered. Since $U_1 = U_2 = U_3 = 0$, the summation in (1) starts from 4 rather than from 1.

The only unrecoverable four-erasure pattern occurs when the four erasures make a rectangle as shown in Fig. 1. Erasures are represented by X in the figure. All of the four cells are lost. The number of possible unrecoverable patterns is given as in [6] by

$$U_4 = \binom{M}{2} \binom{N}{2}, \quad (2)$$

$$e_4 = 4.$$

For unrecoverable five-erasure patterns, four errors have to lie on the corners of a rectangle. The fifth erasure is recoverable, and it may occur in any of the remaining $(MN-4)$ choices; hence,

$$U_5 = U_4 (MN - 4), \quad (3)$$

$$e_5 = 4.$$

Recoverability analysis of six and seven erasures is presented in detail in the following subsections.

1. Recoverability Analysis of Six Erasures

There are three scenarios that have six erasures with unrecoverable patterns: unrecoverable six-erasures with a rectangular pattern, unrecoverable six-erasures with a shared missing corner from two rectangles, and only four out of six erasures are unrecoverable.

A. Unrecoverable Six-Erasure Rectangular Pattern, U_{6a}

The first way in which all the six erasures are unrecoverable is when they form two rectangles as shown in Fig. 2. The first rectangle comprises four cells marked with X. The other two erasures can be in the same column and in the existing two rows, denoted by C in Fig. 2(a). They may also occur in the same row, R, and in the existing two columns as shown in Fig. 2(b). In the first case, there are $(N-2)$ possible columns. Alternatively, in the second case, the lost cells may be located on one of the remaining $(M-2)$ rows. In both cases, there are three overlapping rectangles. Any one of them can be considered as the main rectangle; therefore, the final expression should be divided by 3. Starting with U_4 , the number of

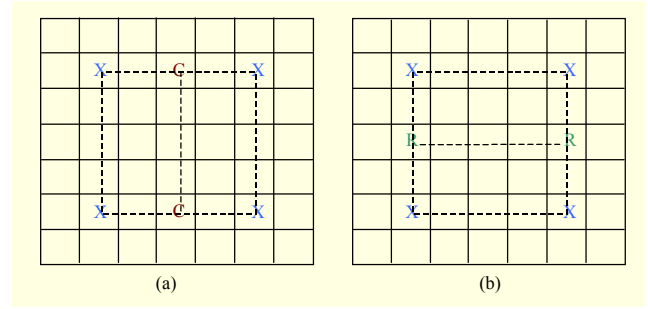


Fig. 2. Unrecoverable six-erasure rectangular pattern.

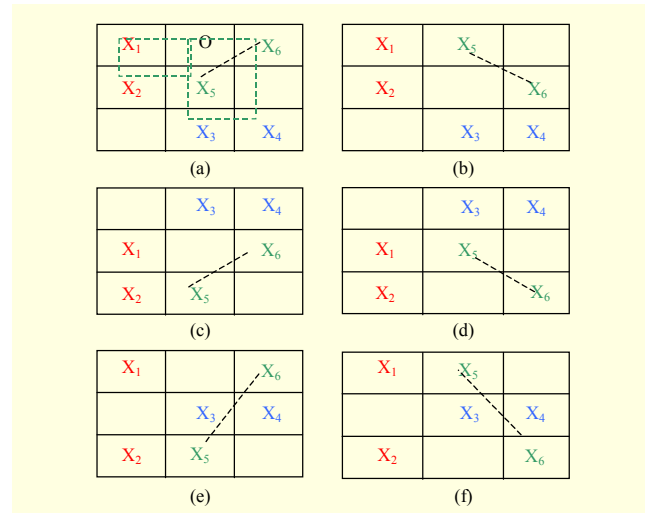


Fig. 3. Unrecoverable six-erasure patterns with two rectangles missing a shared corner.

patterns is given by

$$U_{6a} = \frac{1}{3} U_4 \left[\binom{M-2}{1} + \binom{N-2}{1} \right] = \frac{U_4}{3} [M + N - 4], \quad (4)$$

$$e_{6a} = 6.$$

B. Unrecoverable Six-Erasure Pattern with Two Rectangles Missing a Shared Corner; U_{6b} .

This scenario occurs when the lost cells form two rectangles sharing a missing corner. An example is shown in Fig. 3(a), where the two rectangles are $X_1X_2X_5O$ and $OX_3X_4X_6$. Here, O represents the missing erasure. Figure 3 shows all cases of this scenario for a 3×3 matrix. The lost cells indicated by X_1 and X_2

have to be in the same column and any two rows: $\binom{N}{1} \binom{M}{2}$.

The cells represented by X_3 and X_4 are in the same row and in a row and columns not used by X_1 and X_2 . The total number of scenarios for X_3 and X_4 is $\binom{M-2}{1} \binom{N-1}{2}$. The cells X_5 and X_6 can be located in either of the two remaining diagonals. The

total number of possible scenarios is given in (5). The expression is divided by 3 because the pairs of lost cells, X_1X_2 , X_3X_4 , and X_5X_6 , are interchangeable:

$$\begin{aligned} U_{6b} &= \frac{1}{3} \binom{M}{2} \binom{N}{1} \binom{M-2}{1} \binom{N-1}{2} \cdot 2 \\ &= U_4 \frac{2}{3} (M-2)(N-2), \\ e_{6b} &= 6. \end{aligned} \quad (5)$$

In deriving (5), we use the following identity:

$$\binom{k}{1} \binom{k-1}{2} = \binom{k}{2} (k-2).$$

C. Four Out of Six Erasures Are Recoverable, U_{6c}

To be able to recover two out of the six erasures, four erasures should form the unrecoverable rectangular pattern, and the remaining two can be anywhere in the remaining $(MN-4)$ cells, except the ones included in U_{6a} . For this scenario,

$$\begin{aligned} U_{6c} &= U_4 \left[\binom{MN-4}{2} - (M+N-4) \right], \\ e_{6c} &= 4. \end{aligned} \quad (6)$$

Therefore, the total number of unrecoverable patterns, U_6 , is the sum of U_{6a} , U_{6b} , and U_{6c} , and the number of cells in error after decoding, e_6 , is the average of the three cases. U_6 and e_6 are given by

$$\begin{aligned} U_6 &= U_4 \left[\binom{MN-4}{2} + \frac{2}{3} (M-2)(N-2) \right. \\ &\quad \left. - \frac{2}{3} (M+N-4) \right], \\ e_6 &= \frac{1}{U_6} [U_{6a}e_{6a} + U_{6b}e_{6b} + U_{6c}e_{6c}]. \end{aligned} \quad (7)$$

2. Recoverability Analysis of Seven Cells

There are three scenarios that have unrecoverable patterns with seven erasures: all the seven are unrecoverable; six remain erased, and one is recovered; and four remain erased, and three are recovered.

A. Seven Unrecoverable Cells, U_{7a}

All the seven cells will be lost if six cells are located as in U_{6b} and the seventh fills a missing corner. It can be in any of three locations shown Fig. 4. Therefore,

$$\begin{aligned} U_{7a} &= 3U_{6b} = 2U_4(M-2)(N-2), \\ e_{7a} &= 7. \end{aligned} \quad (8)$$

An alternative approach to obtain the same result is shown in

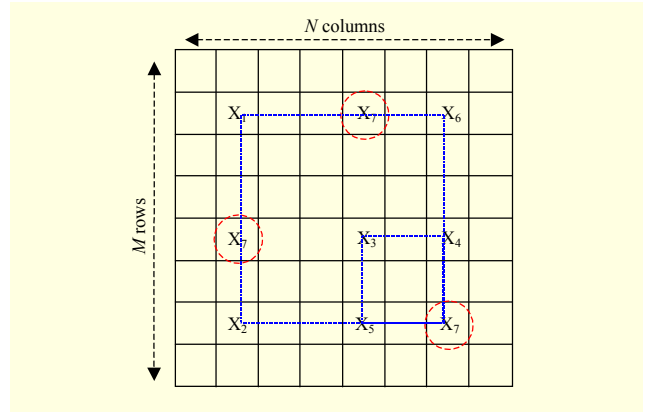


Fig. 4. Seven unrecoverable erasures, U_{7a} .

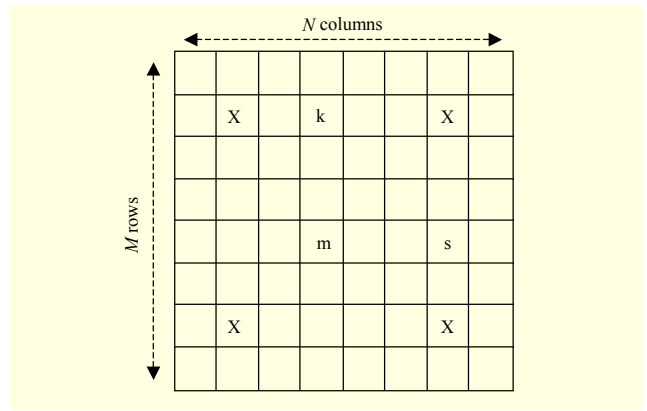


Fig. 5. Another method to derive the number of seven unrecoverable erasures, U_{7a} .

Fig. 5. The four main erasures must make a rectangle, and the other three cells form a rectangle with one of the four main cells. That is, the four Xs form the main rectangle, and k , m , s , and one X form a second rectangle. Note that k needs to be in a free column in one of the two rows occupied by the main rectangle, s should be located opposite to k , and m should be in the cell at the intersection of k 's column and s 's row. The number of possible choices for the main rectangle is U_4 . To build the second rectangle, there are four possibilities for the choice of X. The number of possible cells is $(N-2)$ for k , and it is $(M-2)$ for s . We need to divide by two because there are two rectangles, and either one of them can be the main rectangle, which results in the same expression as in (8).

B. Six Lost Erasures and One Recoverable, U_{7b}

For only one cell out of the seven erasures to be recoverable, six erasures will have the same pattern of U_{6a} and the recoverable cell can be anywhere else. There are $(MN-6)$ possible locations for the recoverable cell. Alternatively, the six erasures can have the same pattern as U_{6b} , and the recoverable erasure can be anywhere except these six erasures and the three

missing corners. Therefore,

$$\begin{aligned} U_{7b} &= U_{6a}(MN - 6) + U_{6b}(MN - 9), \\ e_{7b} &= 6. \end{aligned} \quad (9)$$

C. Four Lost Erasures and Three Recoverable, U_{7c}

For three erasures to be recoverable, they can be in any cell except in the four cells forming the rectangle, $\binom{MN-4}{3}$, and they should not be in the above two cases, U_{7a} and U_{7b} . Therefore,

$$\begin{aligned} U_{7c} &= U_4 \left[\binom{MN-4}{3} - 4(N-2)(M-2) \right. \\ &\quad \left. - (M+N-4)(MN-6) \right], \\ e_{7c} &= 4. \end{aligned} \quad (10)$$

The total number of unrecoverable patterns is the sum of U_{7a} , U_{7b} , and U_{7c} . The number of unrecoverable cells after decoding is the average of the three cases, which is given by

$$\begin{aligned} U_7 &= U_{7a} + U_{7b} + U_{7c}, \\ e_7 &= \frac{1}{U_7} (U_{7a}e_{7a} + U_{7b}e_{7b} + U_{7c}e_{7c}). \end{aligned} \quad (11)$$

3. Recoverability of Eight or More Cells

If one of the parity checks precisely involves one of the erasures, then we use this parity to determine the value of the erasure and continue. If no such parity check can be found, then we have a stopping set. Recoverability or stopping set analysis for higher numbers of erasures is tedious, and we resort to using an upper bound on the performance. We assume that all erasure patterns are unrecoverable and that, in this case, none of the cells will be recovered. For this upper bound,

$$\begin{aligned} U_i &= \binom{MN}{i}, \\ e_i &= i, \quad 8 \leq i \leq MN. \end{aligned} \quad (12)$$

For SPCPC, if the number of erasures is greater than or equal to $M+N$, there will be unrecoverable cells. Therefore, under this condition, the expression for U_i given in (12) becomes exact. A lower bound can be readily derived by assuming that $U_i=e_i=0$ for all values of i greater than 7.

IV. Results and Analysis

To verify the previous expressions, an exhaustive simulation was performed where all the possible permutations for a specific number of erasures were generated and decoded. The

Table 1. Summary of post-decoding lost cells.

$M \times N$	i	e_i	U for $e=4$	U for $e=5$	U for $e=6$	U for $e=7$	U_i
3×4	4	4	18				18
	5	4	144				144
	6	4.1707	450	0	42		492
	7	4.7273	540	0	180	72	792
4×4	4	4	36				36
	5	4	432				432
	6	4.1212	2232	0	144		2,376
	7	4.4314	5904	0	1,152	288	7,344
4×5	4	4	60				60
	5	4	960				960
	6	4.0939	6900	0	340		7,240
	7	4.313	27960	0	4,040	720	32,720
5×5	4	4	100				100
	5	4	2100				2,100
	6	4.0755	20400	0	800		21,200
	7	4.2417	118000	0	13,400	1800	133,200

number of lost cells after recovery agrees with the developed algorithms for all examined scenarios. Table 1 summarizes the results for different values of M and N . The table also details the numbers of patterns with specific numbers of lost cells. These verified expressions can be used to judge the performance of SPCPC.

Examining the performance expression given in (1), we can identify three main components, namely, number of unrecoverable patterns, U_i ; the average number of unrecoverable cells, e_i ; and the probability term, $p^i(1-p)^{MN-i}$. At a low probability of error, scenarios with a high number of erasures contribute to the bound at a very low level because of the $p^i(1-p)^{MN-i}$ factors. At a high probability of error, the probability term becomes less dominant, and the bounding of U_i and e_i results in a loose bound.

In [11], although $e_i \leq i$, when evaluating the bound, making $e_i=i$ was assumed to have a minimal effect on the bound as compared to a large value of U_i . This is not true because the terms are multiplicative rather than additive. The first enhancement to the bound in [11] is evident because we calculated e_i up to $i=7$. For the case of $M=N=5$, Table 1 shows that for $i=7$ the average number of lost cells is 4.24 rather than 7. The second enhancement is the correction added to the number of unrecoverable patterns U_i up to $i=7$. We have shown that unrecoverable patterns can be made without the basic square pattern.

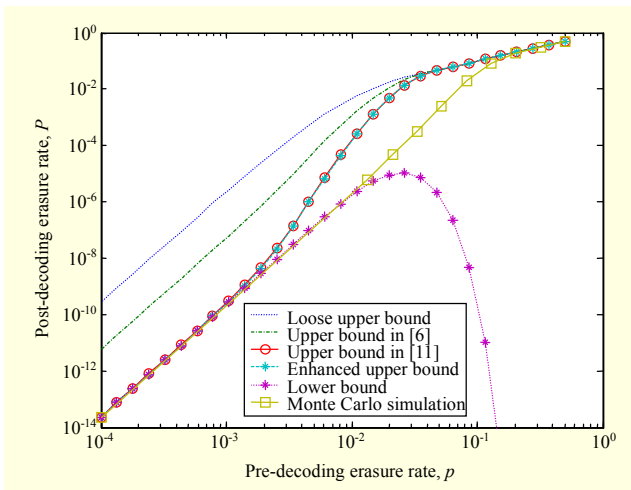


Fig. 6. Simulation and various bounds on the post-decoding erasure rate of a 16×16 SPCPC.

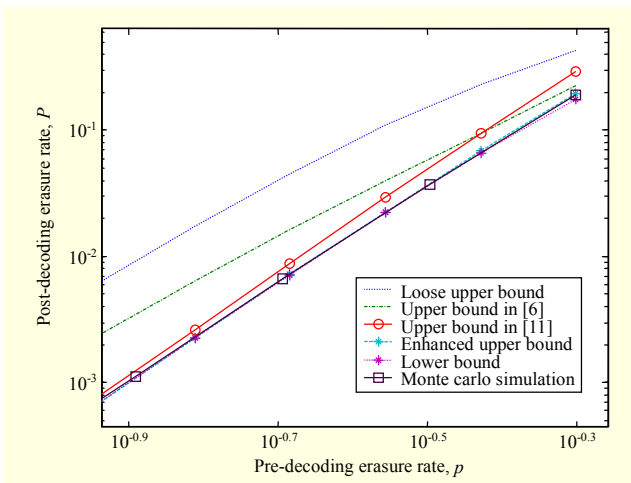


Fig. 7. Simulation and various bounds on the post-decoding erasure rate of a 3×3 SPCPC.

The enhanced bound is plotted in Fig. 6 for a 16×16 matrix. The figure also shows the Monte Carlo simulation points. Other curves are plotted for comparison. The loose bound represents the straightforward approach to evaluate (1), where any pattern of erasures greater than four is assumed to be left intact, and no recovery is attempted. Equation (1) therefore reduces to the following loose bound:

$$P < \frac{1}{MN} \sum_{i=4}^{MN} i \binom{MN}{i} p^i (1-p)^{MN-i}. \quad (13)$$

The bound in [6] considers the exact number of unrecoverable patterns and the exact number of lost cells with up to five erasures. Our enhanced bound considers the exact expressions for up to seven erasures. Note that, on the logarithmic scale, the performance of SPCPC can be fitted to a straight line at low pre-decoding erasure rates. The Monte

Carlo simulation converges to the enhanced bound at low pre-decoding erasure rates. Figure 6 shows a tight fit between the proposed bound and the simulation points for p less than 0.001. This range for p is the operating range for most of the practical applications. The enhanced upper bound and the derived lower bound agree for small values of p .

In the previous example of 16×16 SPCPC, it is difficult to distinguish our bound from the bound in [11]. This is because the bound in [11] does two opposite things. It is looser than the enhanced bound because it assumes that if a pattern contains unrecoverable cells then all of them are lost, $e_i = i$. On the other hand, it mistakenly ignores some unrecoverable patterns which tighten the bound. To illustrate that the bound in [11] is not generally true and tight, simulation results of another 3×3 SPCPC are shown in Fig. 7. The difference between the bound in [11] and our new enhanced bound is more pronounced.

V. Conclusion

Mathematical expressions for the recoverability of up to seven erasures were derived and verified through exhaustive study. An upper bound is used for the case of more than seven erasures. The bound developed is easy to implement and can be used for code rate calculation in an adaptive packet-level FEC application where the code rate is a function of p , that is, the raw probability of cell loss. Based on the generated performance curves for various array sizes (for example, 4×4 , 8×8 , and 16×16), the post-decoding performance can be accurately estimated, and the code rate can be optimized. The developed bound can help to reduce the cell loss rate by evaluating the proper code for future data storage and real-time applications in wireless and optical networks. The paper also demonstrated a lower bound which, when used with the enhanced upper bound, can accurately estimate the performance of the code at low pre-decoding error rates.

Acknowledgements

The author thanks Mr. Khalid Darwish and Mr. Sohaib Al-Baset for their constructive ideas. The author also thanks Dr. Muhammad Taher Abuelma'atti and Mr. David Birkett for their editorial comments.

References

- [1] M. El-Khamy, "The Average Weight Enumerator and the Maximum Likelihood Performance of Product Codes," *Int. Conf. Wireless Networks, Communications and Mobile Computing*, Maui, Hawaii, USA, June 2005, pp. 1587-1592.

- [2] S. Lin and D.J. Costello, *Error Control Coding: Fundamentals and Applications*, Englewood Cliffs, New Jersey: Prentice Hall, 1983.
- [3] Y.M. Chee, C.J. Colbourn, and A.C. Ling, "Asymptotically Optimal Erasure-Resilient Codes for Large Disk Arrays," *Discrete Applied Mathematics*, vol. 102, 2000, pp. 3-36.
- [4] K. Srinivasan and C. Colbourn, "Failed Disk Recovery in Double Erasure RAID Arrays," *Journal of Discrete Algorithms*, vol. 5, no. 1, Mar. 2007, pp. 115-128.
- [5] E. Rosnes, "Stopping Set Analysis of Iterative Row-Column Decoding of Product Codes," *IEEE Trans. Information Theory*, vol. 54, no. 4, Apr. 2008, pp. 1551-1560.
- [6] M.A. Kousa and A.H. Muqaibel, "Cell Loss Recovery Using Two-Dimensional Erasure Correction for ATM Networks," *Seventh Int. Conf. Telecommunication Systems*, Mar. 1999, pp. 85-89.
- [7] H.D. Hollmann and L.M. Tolhuizen, "Generic Erasure Correcting Sets: Bounds and Constructions," *Journal of Combinatorial Theory*, vol. A113, no. 8, Nov. 2006, pp. 1746-1759.
- [8] D.M. Rankin and T.A. Gulliver, "Single Parity Check Product Codes," *IEEE Trans. Communications*, vol. 49, no. 8, Aug. 2001, pp. 1354-1362.
- [9] M.A. Kousa, "A Novel Approach for Evaluating the Performance of SPC Product Codes under Erasure Decoding," *IEEE Trans. Communications*, vol. 50, Jan. 2002, pp. 7-11.
- [10] R. Amutha, K. Verrarghavan, and S. Srivatsa, "Recoverability Study of SPC Product Codes under Erasure Decoding," *Information Sciences*, 2005, pp. 169-179.
- [11] A. Al-Shaikhi and J. Ilow "Erasure Rate Analysis and Tighter Upper Bound for Binary Product Codes," *IEEE Communications Letters*, vol. 10, no. 7, July 2006, pp. 546-548.
- [12] A. Al-Shaikhi and J. Ilow, "Improved Upper Bound for Erasure Recovery in Binary Product Codes," *IEEE 16th Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, Berlin, Germany, vol. 2, Sept. 2005, pp. 1086-1090.



Ali Muqaibel received the BSc and MSc degrees from King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 1996 and 1999, respectively, and the PhD from Virginia Polytechnic Institute and State University, USA, in 2003. He worked as an engineer and then as a lecturer at KFUPM from 1996 to 2000. During his study at Virginia Tech, he worked both for the Time Domain and RF Measurements Laboratory and for the Mobile and Portable Radio Research Group (MPRG). He is currently an assistant professor with the Electrical Engineering Department at King Fahd University of Petroleum and Minerals. His main areas of interest include coding, measurements for channel characterization, and ultra-wideband communication.