Low-Density Parity Check Codes for High-Density 2D Barcode Symbology

Ramon Francisco Mejia, Yuichi Kaji, and Hiroyuki Seki
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
Email: ramon-m@is.naist.jp, kaji@is.naist.jp, seki@is.naist.jp

***Abstract*—Increasing the data density of two-dimensional barcode symbology is an important area of research in Automatic Identification and Data Capture systems because it provides significant improvements on the range and usefulness of barcodes in different applications. The implementation of error-correcting codes for such symbologies is crucial due to the susceptibility of high-density barcode symbols to damage. In this paper, we study the performance of Low-Density Parity Check (LDPC) codes for a two-dimensional barcode symbology with high data density. A high-density barcode symbology is designed, and the characteristics of the communication channel defined by commonly used printers and scanners are modeled and observed. Additionally, the parameters of the symbology are adjusted and the performance of LDPC codes with different code rates is tested. Performance tests show that LDPC codes are effective for certain parameters of the high-density symbology, and limitations of the chosen printing technology severely affect the robustness of the symbology.**

***Keywords-Two-Dimensional Barcodes*; *Low-Density Parity Check Codes*; *Automatic Identification and Data Capture*; *Image Processing*.**

I. Introduction

A lot of interest has been put on the study of *two-dimensional (2D) barcode systems*, where data is represented as a machine-readable symbol printed on a physical surface called a *barcode symbol* [1]. In such systems, the symbol represents data in a matrix of high and low reflectance regions of the printing surface, and therefore able to carry between 10 to 100 times more data than one-dimensional barcodes. Continued development in this area revolves around improving the *symbology* of 2D barcodes in order to increase its data density; i.e., enhancing the structure and processing methods of a barcode symbol to increase its data capacity while retaining its compact size and portability.

Improvements in the symbology of 2D barcodes further expand its application for different systems. Quick Response (QR) Codes [2], created by Japanese company Denso-Wave in 1994, are used for Medical Information Management systems, where prescriptions are stored in a barcode to reduce human error in the interpretation of handwritten prescriptions and the administration of medicine [3]. In Fu et al. [4], a study of tax-filing methods in Taiwan revealed that the adoption of an electronic filing system which utilizes 2D barcodes increased the processing efficiency of tax returns and reduced error rates versus paper returns. In addition, barcodes also have applications in Mobile Commerce [5], Multimedia Teachware [6] and many others. For these applications, the high data density of 2D barcodes is important as they function not only as an index to external databases, but can hold either files or databases themselves.

A number of 2D barcode symbologies have been proposed so far, and while most are designed with reading speed in mind, some of them are designed for high data capacity [7]. For example, Optar [8] can accommodate 200 kilobytes of data in one A4-paper. PaperDisk [9] can contain 1 megabyte of data in an 8.5 $\times$ 11 inch space. Laboratory tests for High Capacity Color Barcode [10] using eight colors have a capacity of 2,000 bytes of data per square inch in its highest density. However, technical details of these technologies are not disclosed so far. It will be useful if high data capacity barcodes are realized using open technologies only, and if these barcodes can be used with reasonably-priced equipment found in office or home environments.

One disadvantage to increasing the data density of 2D barcodes is its increased susceptibility to errors. Physical damage or inaccuracies caused during the printing or scanning process of the 2D symbol may cause erroneous reading of the stored data bits. To recover from errors, most 2D barcode symbologies use *Reed-Solomon* (RS) codes to encode data prior to generating the symbol [11]. In recent years, it has been shown that well-designed *Low-Density Parity Check* (LDPC) codes [12] perform well compared to RS codes for some communication channels. A remarkable aspect of LDPC codes is that we can perform *soft-decision decoding* for LDPC codes with almost linear-time complexity. Soft-decision decoding is an algorithm for error correction in which inputs to the algorithm can have continuous values. It is more powerful than conventional (referred to as hard-decision) algorithms in which inputs are quantized into two level, but usually requires very large computational complexity. For this reason, soft-decision decoding has been considered as impractical for conventional error-correcting codes including RS codes. On the other hand, this problem can be mitigated with the use of LDPC codes. It has been shown that the performance of appropriately designed LDPC codes asymptotically approaches to Shannon limit as we extend the code length [13], and that LDPC codes with a practical code length show better performance than

 ISBN: 978-1-61208-139-7

conventional error-correcting codes such as RS codes and convolutional codes [14], [15].

This study investigates the error-correcting performance of LDPC codes in a high-density 2D barcode symbology. Given the advantageous properties of LDPC codes over RS codes, our goal is to design a symbology capable of storing large amounts of data, and determine appropriate symbology parameters and LDPC codes to ensure robustness against errors. In order to achieve this, the symbology was designed to allow larger data block sizes (i.e., longer LDPC codes) to be placed within the symbol. The performance of LDPC codes were then tested with barcode symbols of different data densities and geometric parameters. This symbology also takes into consideration the limitations of printing and scanning technologies present in most office environments today.

The remainder of this paper is organized as follows. In Section II the design of the symbology is introduced. Next, methods for evaluating the performance of LDPC are described in Section III. Finally, Section IV discusses the results of the evaluation.

II. Symbology

The proposed barcode symbology for this study includes common structural features found in other 2D barcode symbologies, such as finder patterns and timing patterns. However, the focus of the proposed symbology is to increase data density, where the number of bits represented in a given printing area is significantly greater than in other 2D barcodes designed for fast reading. Therefore, instead of using traditional barcode printers and scanners with limited resolutions and computing power, the symbology is designed to work with equipment available in most office environments; namely, a laser printer, a flatbed scanner, and a desktop computer. The following subsections describe the structure of the symbology and the processes involved in generating and scanning symbols.

A. Data Area

A symbol is composed of a $d \times d$ matrix of *data cells* printed as a monochrome image on approximately a 25.4 mm $\times$ 25.4 mm data area, where d is the dimension or the number of cells per row, and d^2 is the density of the symbol. Each data cell represents a single bit of data, printed as a solid square. For the purposes of this study, a white cell is printed when a '0' bit is required and a black cell for a '1' bit.

Because of the size restrictions for the data area, the dimension d of the matrix determines the printing area for a data cell. The *cell size* in millimeters is computed as $(25.4 \text{ mm}/(d + 2 \text{ cells}))^2$. The additional 2 cells are allocated for timing patterns, which will be explained in the following subsection. Since data cells are small (around $0.257\text{mm} \times 0.257\text{mm}$ for $d = 97$), it is probable that the printer toner used to draw black cells "spill out" and blot neighboring white cells. This is called *inter-pixel leakage*, and this is mechanically unavoidable when laser printers are used in printing. This may not be the case when using professional-quality image setters, however as mentioned, the environment considered is that of a typical office setting.

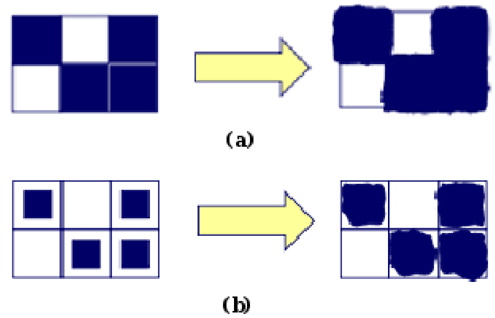


Figure 1. Inter-pixel leakage for (a) black cells printed without adjustments ($mf = 1.0$) and (b) with margin factor ($mf = 0.6$).

To reduce this effect, we propose that the printing size of black cells is reduced to a certain percentage by a *margin factor*. Figure 1 shows the effect of inter-pixel leakage and adding a margin factor.

B. Timing and Finder Patterns

The data area is bordered by four timing patterns. A *timing pattern* is used in most 2D barcodes as a way to calculate the size and location of data cells. In this symbology, the timing pattern is a consecutive series of data cells starting and ending with a black cell, connected with alternating white and black cells. Note that since timing patterns must start and end with a black cell, the dimensions of the data area must be odd.

In order to detect the location of the timing patterns, finder patterns are positioned on the four corners of the symbol. A *finder pattern* consists of three black rings which are co-centric to the first cell of each timing pattern. The rings intersect with the first three black cells of the timing pattern. The diameter and thickness of the rings depend on the dimensions and margin factor of the symbol. To improve the detection of the finder pattern, a *quiet zone* of white cells is placed around the location of the finder patterns. As a result of placing quiet zones in the corners of the symbol, the locations of overlapping data cells are adjusted. Figure 2 illustrates an example of a symbol with timing and finder patterns, and a complete barcode symbol.

C. Encoding Procedure

Data in a barcode symbol is protected by an error-correcting code which is known as an LDPC code. LDPC codes are a class of linear block codes that have very sparse parity check matrices. The sparse structure of the

 ISBN: 978-1-61208-139-7

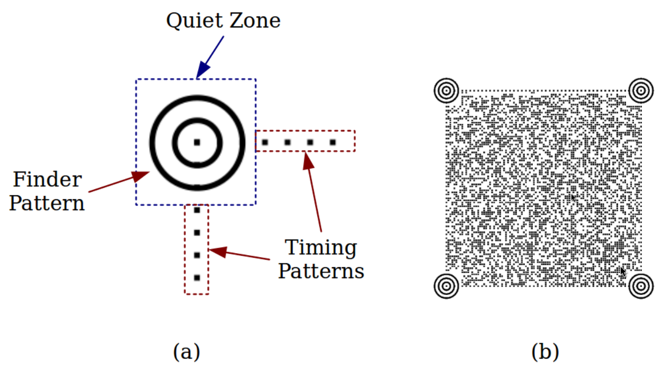


Figure 2. (a) Construction of a finder pattern, quiet zone and two timing patterns. (b) A complete symbol created using the symbology.

check matrix makes the "belief-propagation" principle very effective for finding and correcting errors which are involved in a received signal.

In this study, we consider to use LDPC codes which are designed for IEEE 802.16e standard (also known as mobile WiMAX). Using these codes is advantageous since they have smaller complexity for encoding operation. In general, the encoding operation of an LDPC code requires quadratic-order complexity in the code length; however, the IEEE codes defined in the standard are designed so that they have *quasi-cyclic* structure, which enables the realization a linear-order encoding algorithm.

Another advantage of these codes is that the code parameters can be changed in a flexible manner. The standard defines several classes of LDPC codes with code rates 1/2, 2/3, 3/4 and 5/6, and code length ranges from 576 to 2304 bits. These parameters have a strong relation to the efficiency and the error-correcting capability of the code [16]. Generally, low-rate codes are more powerful than high-rate codes, but more parity bits need to be added for such codes. This implies that, with the same code length, high-rate codes can accommodate more data than low-rate codes. It is also known that long codes usually show better performance than short codes even if they have the same code rate, but the computational cost of longer frame-lengths need to be considered. These parameters are adjusted according to the dimension of barcodes and desired level of reliability.

D. Symbol Processing

We refer to the process of printing and scanning barcodes as *symbol processing*. During symbol processing, the symbol is printed on a piece of paper using a laser printer. A flatbed scanner scans the symbol and the resulting image is passed to an image processing program. The program performs five steps:

1) Finder patterns are located using a basic template matching algorithm [17]. The centers of all finder patterns are then computed.
2) Lines connecting the centers of adjacent finder patterns are connected, and the resulting closed rectangle is masked. This enhances the detection of timing patterns in the following step.
3) The lines connecting adjacent finder patterns are scanned through pixel by pixel (note that one cell consists of several pixels). Given the position of the finder patterns, the path each line passes through is also the location of a timing pattern. When the value of the pixel changes from white to black, this pixel is marked as a *transition point*.
4) Transition points from opposite timing patterns are connected, forming a grid of *sampling cells*. The ratio of black pixels to the total number of pixels from each sampling cell is computed.
5) The ratio r obtained from each sampling cell is mapped to a *soft value* using the function $f(r) = -(2r - 1)$. Soft values are grouped into codewords, which are passed to an LDPC decoder program.

E. Decoding Procedure

The decoding, or error correction, is performed by using a belief-propagation algorithm for LDPC codes [13]. In this algorithm, we consider representing the mathematical structure of the code with a bipartite graph whose incident matrix coincides with the check matrix of the code. The nodes of the bipartite graph are grouped to *variable nodes* and *check nodes*. A variable node receives information from neighbor check nodes, and it attempts to estimate which symbol ('0' or '1' bit) has been transmitted. During the estimation, the statistical information of the communication channel, such as the variance of the Gaussian channel, is considered to derive various probabilities. A check node receives the estimated symbols from neighbor variable nodes, monitors parity constraints, and gives check nodes suggestions for the transmitted symbol. The accuracy of the estimation improves as nodes exchange messages iteratively. It is known that, in most cases, the decoding algorithm reaches the correct codeword with a small number of iterations. The number of iterations needed is rather independent from the code length, thus the decoding algorithm can be regarded as "almost linear-order" complexity.

III. Evaluation

In order to evaluate the performance of the symbology, two experiments were conducted on samples of barcode symbols. The experiments assessed the behavior of the communication channel from different perspectives.

All experiments used the same equipment and symbol processing steps for testing. First, a set of 24 barcode symbols was generated using an encoder program written in the C programming language. The symbols were then printed on plain white bond paper using a Canon LBP3410 laser printer with the default settings. Next, symbols were scanned using

 ISBN: 978-1-61208-139-7

an Epson GT-F720 flatbed scanner at 720dpi. In both the printing and scanning process, monochrome color settings were used. Finally, each symbol was read using an image processing and decoding program, also written in C.

A. Test 1: Analysis of Channel Characteristics

In the first experiment, the characteristics of the communication channel with respect to symbol processing were investigated. The image data read by the scanner are not identical to the virtually constructed barcode image. Many factors, such as the inter-pixel leakage of printers or blobs on bond paper, causing differences between the scanned image and the ideal image. For reliable data recording, it is essential to eliminate these effects or noises. Furthermore, the communication channel which is defined by a printer and a scanner is different from conventional communication channels, therefore statistical analysis of the channel is vital in determining how to implement LDPC codes for the symbology.

As a result, a set of symbols with dimension $d = 117$ and margin factor $mf = 0.6$ were generated and sampling cells were scanned. The computed ratios (before mapping to a soft value) were separated into two groups, based on the expected value (white or black) of the cell. Histograms for both groups were then plotted to infer observations on the channel model. From the distribution, we can observe how the values in the sampling cells match the encoded data bits after they have gone through symbol processing.

B. Test 2: Performance of LDPC Codes

In the second experiment, the error-correcting performance of LDPC codes was evaluated against changes to parameters of the symbology. To measure performance, the bit-error rate (BER) [18] of each sample set was analyzed, where

$$BER = \frac{\text{Number of erroneous bits after decoding}}{\text{Total number of bits in the set}} \quad (1)$$

There are two phases in this experiment. In the first phase, we wish to observe the error correction performance of LDPC codes as data density increases. Sample sets with increasing dimensions were created using LDPC codes with code rates $1/2$, $2/3$, $3/4$, and $5/6$. As previously mentioned, the LDPC codes used are those defined for IEEE 802.16e, and the code rate indicates the error correction capability of the code (lower code rate means stronger error-correcting capability). The margin factor for all symbols was set at $mf = 0.6$.

In the second phase, sample sets with different margin factors were created using the same LDPC code rates as above. The dimensions of all symbols was set at $d = 127$. As discussed earlier, the printing size of a black cell affects the impact of inter-pixel leakage and the ability of the image processing stage to identify the finder and timing patterns correctly. It is therefore of interest to see how the printing size of black cells affects error-correcting performance.

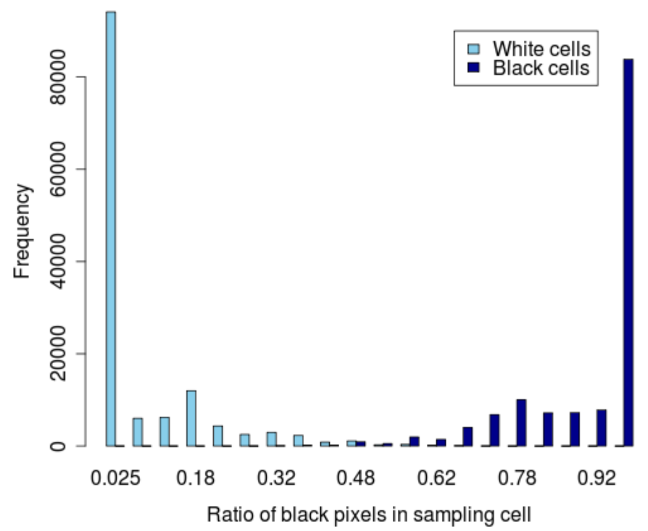
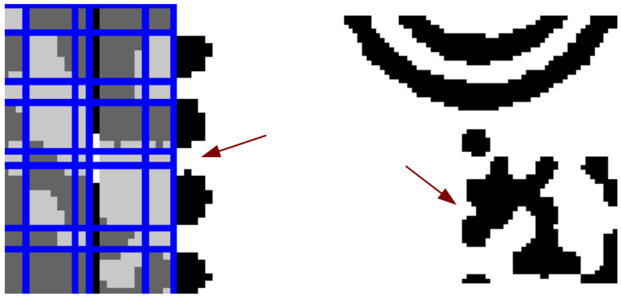


Figure 3. Histogram of ratios for black cells and white cells for symbols with dimension $d = 127$ and margin factor $mf = 0.6$

IV. Results and Discussion

The histograms generated for Test 1 are presented in Figure 3. Statistical analysis of the distribution showed that the channel can be modeled as an *additive white Gaussian noise* (AWGN) channel, which is suitable for the soft-decision decoding of LDPC codes. Also, by analyzing the distribution, the AWGN variance which will be used for the LDPC decoder in the subsequent test was determined to be 0.05637.

During the execution of the first experiment, it was observed that the printing process had another effect on the printing of the data cells. In addition to inter-pixel leakage, cells along the same row are printed with similar sizes, but cells in other rows may be printed with different heights. The same effect was also observed for the widths of cells from the same columns and cells in different columns. It is not clear if this effect is caused by the mechanical constraints of the printer or scanner, or other known factors; the printed barcode image is not as precise and uniform as stated in the specifications of the devices. Nevertheless, the timing pattern considered in this study can cope with this effect; since timing patterns are also printed on the same row or column as data cells, the grid of sampling cells formed by transition points adjusts to the changes in the printing size of the rows or columns. However, this also means that



(a) (b)

Figure 4. (a) Timing patterns are too close and image sampling cells are too small. (b) Timing patterns have no clear separation and creation of the sampling cells fail.

the accuracy of timing pattern detection is crucial to the performance of the symbology.

In Test 2, the parameters for sample sets were varied and the BER for each set was analyzed. The results of the BER analysis for phase 1, where symbols with increasing dimensions were tested, are presented in Table I. From the results, it can be concluded that LDPC codes with different code rates have similar performance for lower dimensions of the symbol. All codewords from the set of symbols were decoded correctly. For $d = 117$, errors appeared for the symbols with LDPC code rate 5/6. This was expected, as this code rate has the weakest error-correcting capability. However, there is a rapid increase of bit-errors for $d = 127$. Inter-pixel leakage of black cells caused timing patterns to become too close to each other, thus the sampling cells formed were too small and there were not enough pixels to compute accurate soft values. Moreover, note that the BER for $d = 137$ is "Undetermined". Due to the high density of cells in the symbol, some images were distorted by inter-pixel leakage in such a way that transition points for adjacent timing patterns could not be detected by the image processing algorithms, and the creation of sampling cells failed. Figure 4 shows some examples of image processing errors encountered during the experiment.

Table I
BER FOR SYMBOLS WITH MARGIN FACTOR $mf = 0.6$ AND INCREASING DIMENSIONS

	Code rate			
d	1/2	2/3	3/4	5/6
97	0.000%	0.000%	0.000%	0.000%
107	0.000%	0.000%	0.000%	0.000%
117	0.000%	0.000%	0.000%	0.847%
127	0.709%	2.143%	5.674%	1.429%
137	Undetermined			

Table II shows the BER analysis for phase 2, where symbols with different margin factors were tested. We can see that the error-correcting performance degrades if the margin factor is too small or too large. If the margin factor is too small, then a small blot on the paper can cause more white cells to be recognized as black cells. More importantly, the printing process may fail to draw data cells and timing patterns, as was observed in some cases. On the other hand, if the margin factor is too large, then problems related to inter-pixel leakage may arise. If these undesired phenomena occur, then error-correcting codes would not be effective. Indeed, there seems to be little relation between the BER and the capability of codes for margin factors 0.50 and 0.60, which are too small and too large, respectively. The margin factor 0.55 is the best for this resolution, and we can see a monotonic relation between the BER and the code rate.

Table II
BER FOR SYMBOLS WITH DIMENSION $d = 127$ AND INCREASING MARGIN FACTORS

	Code rate			
mf	1/2	2/3	3/4	5/6
0.45	Undetermined			
0.50	2.128%	1.429%	8.511%	7.143%
0.55	2.128%	2.857%	3.546%	3.571%
0.60	0.709%	2.143%	5.674%	1.429%
0.65	Undetermined			

Another remark is that the parameters chosen in this test are close to the performance limit of commonly used laser printers. For the symbols of $d = 127$, the printing size of each cell is around 0.197mm × 0.197mm. Setting the margin factor to 0.50 means that each black cell is drawn by the size 0.098mm × 0.098mm. Due to the small size, a printer may not be able to control the image. In a 600dpi setting, the minimum unit a printer can control is $25.4/600 = 0.042$ mm. Therefore, a black cell with size 0.098mm × 0.098mm is composed of two by two units. Furthermore, with such a small scale, the size of the toner particles used in laser printers cannot be ignored. It is said that the diameter of toner particles is around 0.005mm to 0.010mm, which is not very small compared to the cell size. In addition, the toner particle is firmly pressed on paper surface during the printing process, and it is difficult to fully control the position of toner particles on the paper.

V. Conclusion and Future Work

In order to evaluate the effectiveness and robustness of LDPC for high-density 2D barcodes, we first designed a new symbology. Two types of experiments were conducted on the symbology; these were done to observe the characteristics of the channel, and to analyze the changes in the BER of different LDPC codes as the parameters of the symbology were adjusted.

This study has concluded that LDPC codes are feasible for our proposed symbology. The results showed that the communication channel defined by printing and scanning the barcode symbols can be modeled as an AWGN channel, which is suitable for soft-decision decoding of LDPC

 ISBN: 978-1-61208-139-7

codes. Also, the error correction performance of LDPC codes defined in the IEEE 802.16e standard are effective on the channel for certain symbology parameters. Finally, properties of laser printing technology severely affect the performance of the symbology, particularly with regards to the small printing size of data cells. In these cases, LDPC codes were not effective in error correction.

Future work includes performance evaluation of RS codes for the same high-density barcode symbology, and rigorous testing of the robustness of the symbology against physical damage to the symbol.

REFERENCES

[1] Japanese Standards Association, "JIS X 0500-2:2009 – information technology – automatic identification and capture $AIDC$ techniques – harmonized vocabulary – part 2 Optically readable media ORM," 2009.

[2] ——, "JIS X 0510:1999 – two dimensional symbol – QR code – basic specification," 1999.

[3] Y. L. Yeh, J. C. You, and G. J. Jong, "The 2D barcode technology applications in medical information management," *Intelligent Systems Design and Applications, International Conference on*, vol. 3, pp. 484–487, 2008.

[4] J. Fu, C. Farn, and W. Chao, "Acceptance of electronic tax filing: A study of taxpayer intentions," *Information & Management*, vol. 43, no. 1, pp. 109–126, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/B6VD0-4H0BSXY-1/2/eb9f5899522023887bc24f3e111ea556

[5] J. Gao, V. Kulkarni, H. Ranavat, L. Chang, and H. Mei, "A 2D barcode-based mobile payment system," in *Proceedings of the 2009 Third International Conference on Multimedia and Ubiquitous Engineering*, ser. MUE '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 320–329. [Online]. Available: http://dx.doi.org/10.1109/MUE.2009.62

[6] D. Kim and Y. Mun, "Design and performance analysis of multimedia teachware making system using 2D barcode," in *Computational Science and Its Applications - ICCSA 2006*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, vol. 3981, pp. 195–203. [Online]. Available: http://dx.doi.org/10.1007/11751588_21

[7] Y. Kaji, "Two-dimensional barcode system with extreme density," Institute of Electronics, Information and Communication Engineers, Tech. Rep. IT2009-56, January 2010.

[8] Twibright Labs, "Optar," http://ronja.twibright.com/optar/ 14.04.2011.

[9] Cobblestone Software, "Paperdisk technology," http://www.paperdisk.com/id1.html 14.04.2011.

[10] Microsoft Research, "High capacity color barcode technology," http://research.microsoft.com/en-us/projects/hccb/about.aspx 14.04.2011.

[11] K. Tan, D. Chai, and H. Kato, *Barcodes for Mobile Devices*, 1st ed. New York: Cambridge University Press, 2010.

[12] R. Gallager, "Low-density parity-check codes," *Information Theory, IEEE Transactions on*, vol. 8, no. 1, pp. 21–28, 1962. [Online]. Available: http://dx.doi.org/10.1109/TIT.1962.1057683

[13] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *Information Theory, IEEE Transactions on*, vol. 45, no. 2, pp. 399–431, 1999.

[14] N. Andreadou, C. Assimakopoulos, and F. N. Pavlidou, "Performance evaluation of LDPC codes on PLC channel compared to other coding schemes," in *Power Line Communications and Its Applications, 2007. ISPLC '07., IEEE International Symposium on*, March 2007, pp. 296–301.

[15] J. Chen, L. Wang, and Y. Li, "Performance comparison between non-binary LDPC codes and reed-solomon codes over noise bursts channels," in *Communications, Circuits and Systems, 2005. Proceedings. 2005 International Conference on*, vol. 1, May 2005, pp. 1–4.

[16] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 585–598, February 2001.

[17] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. Pacific Grove: PWS Publishing, 1999.

[18] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*, 2nd ed. Upper Saddle River: Prentice Hall, 2003.