A HEURISTIC METHOD FOR ERROR CORRECTION IN PARALLEL PROBE-BASED STORAGE DEVICES

Maria Varsamou and Theodore Antonakopoulos

Department of Electrical and Computers Engineering
University of Patras, Rio-Patras, 26500, Greece
mtvars@upatras.gr, antonako@upatras.gr

ABSTRACT

This paper presents a new decoding algorithm that improves the reliability of probe-based storage devices, that use multiple, simultaneously accessed parallel fields, when they are affected by burst errors. The presented algorithm exploits the parallelism of the multiple storage fields and the used data allocation method and flags symbols as erasures using the error locations revealed by the initial errors-only decoding attempt. Numerical results demonstrate the performance improvement that is achieved by the proposed algorithm.

Index Terms— Probe storage, Reed-Solomon codes, Burst errors, Erasure decoding.

1. INTRODUCTION

In recent years, due to the constantly increasing need for higher data rates and larger storage densities, the use of multiple, simultaneously accessed channels in storage systems has been investigated [1]. In this case, the original data block is partitioned into several subsets, which are accessed concurrently, each one over a different storage channel. The general architecture of a system with parallel channels is shown in Fig. 1. When the transmission of a new data block is initiated, a preamble is sent for synchronization purposes and then a number of data packets is transmitted. A characteristic example of such a parallel communications system is found in storage on probe-based devices [2], where ultrahigh storage densities and high data rates can be achieved by using atomic force microscopy (AFM) techniques to write and read back data in very thin polymer films with the parallel operation of 2D arrays with multiple tips.

Several coding schemes for parallel communications have been proposed and studied. One scheme with both random and burst error correction capabilities, which is based on Reed-Solomon (RS) codes along with proper interleaving is shown in Fig. 2. The original message is partitioned into a number of datawords which are then encoded using an RS code. The encoded data are symbol-interleaved and then split into smaller blocks, each of which is transmitted over a single channel. An application of this coding scheme is found in data storage with parallel probes [3]. If the channels are statistically independent, then a burst of errors in a single channel will be spread across multiple codewords and the decoder stands good chances of correcting it. However, if an external noise source is applied to the whole system, then it affects all channels concurrently and with the same statistical characteristics. In this case, depending on the number of channels and the duration of the noise effect, a great number of burst errors appear in all codewords and the errors are correlated. On the other hand, it is well known that an RS code that can correct up to t errors, can be extended to correct up to $2t$ errors as long as their positions are known. Knowing the position of corrupted symbols allows us to mark them as erasures. In this work, we propose a heuristic method that exploits the correlation between these burst errors in order to determine erasures for the RS decoder and to enhance the error correction capability without in-

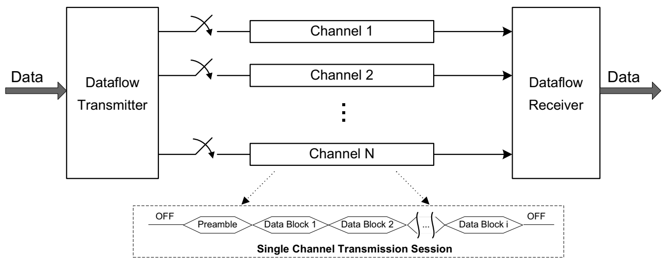


Figure 1. Multiple channels transmission system.

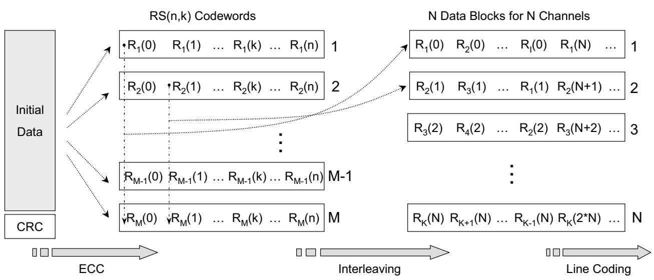


Figure 2. Coding and data allocation in multiple channels.

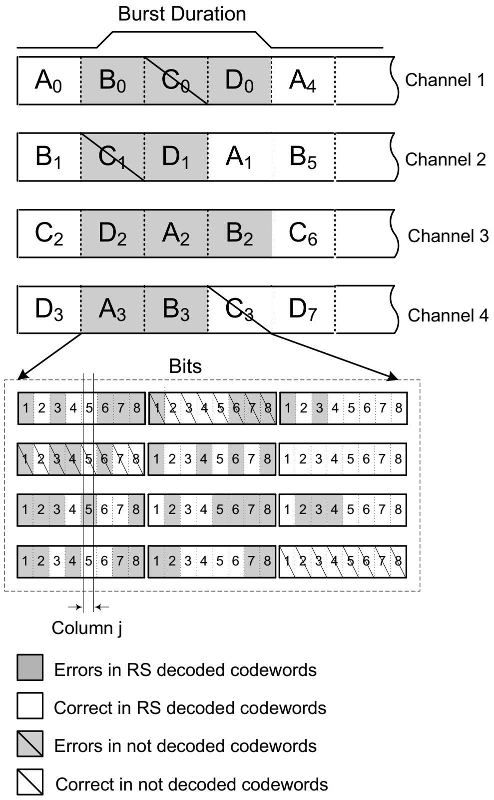


Figure 3. An illustrative example of the errors and erasures decoding algorithm applied to one codeword.

creasing the overhead information. This method can be applied when the standard decoding procedure succeeds in decoding at least a small number of codewords, thus revealing certain error locations. Since even one codeword decoding failure results to a system failure, this method can increase substantially the system's reliability. Simulation results show that this approach yields better performance in terms of overall redundancy rate and implementation complexity.

2. THE ERRORS AND ERASURES DECODING ALGORITHM

Errors-and-erasures decoding of interleaved Reed-Solomon codes has been proved to provide enhanced reliability compared to errors-only decoding on burst error channels. However, it is also well known that mistakenly flagging correct symbols as erasures, seriously deteriorates the performance of the decoder [4]. So, it is very important to have a method that determines in a reliable way the detected symbols as erasures. Several iterative schemes have been proposed that regard as erasures the locations similar to the error locations detected in the neighboring codewords by the errors-only decoder [5], [6]. This approach, depending on the length and the characteristics of the error burst, can lead to misjudgments. We propose a method that assigns an erasure probability to every symbol location of a codeword non-decoded by the errors-only decoder, based on

Algorithm 1 Erasure Estimation Algorithm

for all j **do**
  $pc_j \leftarrow \frac{c_j}{c_j+e_j}$
end for
for all RS symbols in a not decoded codeword **do**
  $P_{er} \leftarrow 1 - \prod_{j=1}^{8} pc_j$
end for
repeat
  do errors-and-erasures decoding using as erasures the symbols with the highest P_{er}
  if *Sector Decoding Success* **then** SUCCESS
  else
    if at least 1 new *Codeword Decoding Success* **then**
      recalculate P_{er}
      if erasures $\neq$ max **then**
        erasures++
      end if
    else
      if erasures $\neq$ max **then**
        erasures++
      else
        FAILURE
      end if
    end if
  end if
until SUCCESS or FAILURE

the information of the erroneous and correct binary symbols that were detected in the same symbol location in the other simultaneously accessed channels. We observe the evolution of the bit error ratio in the sequence of bits that consist each RS symbol and provide accordingly an estimation of the probability that this symbol is in error.

Fig. 3 gives an illustrative example of how the proposed method can be applied. It is assumed that the system includes 4 channels and the data are organized in 4 RS codewords. Due to a burst of errors, one codeword is not decodable by the RS decoder, while the other 3 codewords can be decoded correctly. Column j denotes the position of the bits that are written concurrently on all storage fields. Then c_j corresponds to the number of bits known to be correct as detected by the first decoding attempt, while e_j corresponds to the number of bits known to be initially in error in position j. We use the correct bit proportion pc_j:

$$pc_j = \frac{c_j}{c_j + e_j} \tag{1}$$

as an estimate of the probability that the bit of unknown state in the same position j is correct. Since an RS symbol is correct only when all bits are correct, we can use the quantity P_{er}:

$$P_{er} = 1 - \prod_{j=1}^{8} pc_j \tag{2}$$

as an estimate of the probability that a symbol in a not-decoded codeword is in error. This estimate is called erasure probability. Note that the erasure probability will get

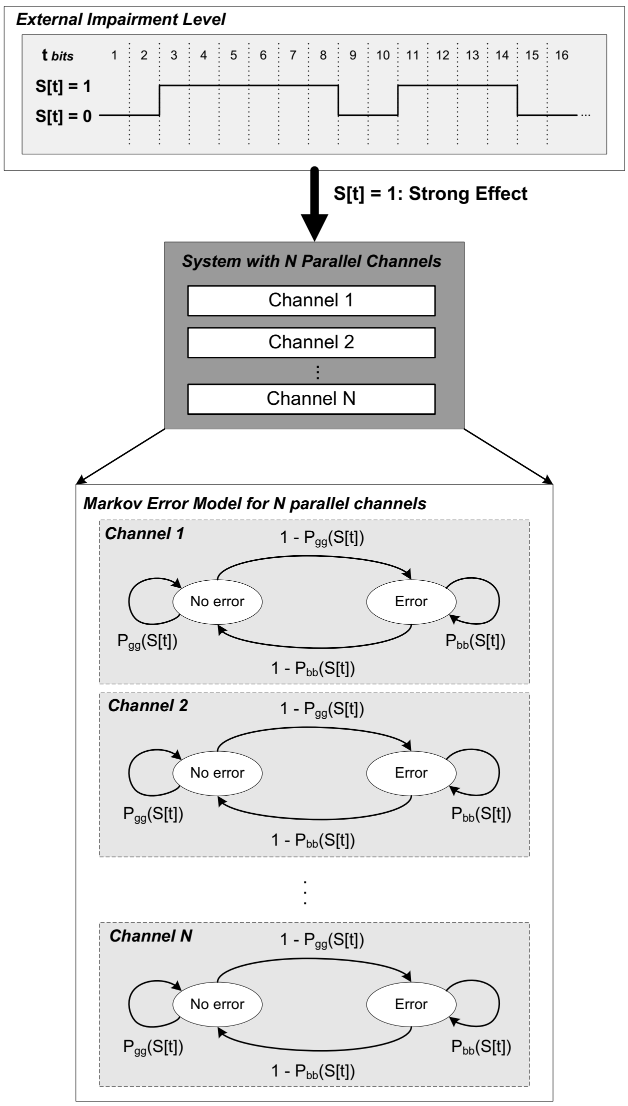


Figure 4. Multiple-channels burst error model based on Markov processes.

a value greater than 0 only for the symbols transmitted during a burst. For all other symbols, the erasure probability equals 0. The symbols with the highest erasure probabilities are flagged as erasures and an additional errors-and-erasures decoding step is performed. The method starts with an initial number of erasures, usually 2, and is executed repetitively by flagging additional symbols as erasures, until either the sector is decoded successfully, or a maximum allowed number of erasures is reached. Algorithm 1 gives a detailed description of the proposed method.

3. APPLICATION ON A PARALLEL PROBE-BASED STORAGE DEVICE

An application of the aforementioned coding scheme is found in the data controller of a probe-based data storage device presented in [1]. In this device, the information is stored by means of thermo-mechanical formation of indentations in thin polymer films [7], using nanometer-sharp tips, similar to those used in atomic force microscopy (AFM) [8]. To increase the achievable data rates, the use of 2D arrays of probes operating in parallel has been proposed [9]. In this case, each probe performs read/write/erase operations on a dedicated area, named a storage or data field, while the storage medium is placed in the x/y plane.

As in conventional storage devices, the data are stored in the form of sectors of fixed length. If N is the number of probes operating in parallel, then each sector is encoded as shown in Fig. 2. N smaller blocks are formed, and each one of them is stored in a single storage field. While reading, the microscanner is responsible for moving the storage fields under their associated tips, such that each tip operates in the center of the line with the sequence of indentations corresponding to the specific sector. The read channel that corresponds to each probe, along with the effects of the various distortions of the read-back signal, has been studied in [10]. Since each probe operates on a distinct storage area, in the absence of external disturbances the N parallel read channels are statistically independent. However, an external shock or vibration that is applied to the device, and consequently to the microscanner, while reading a sector, will cause a displacement to the tips, same to all of them. The effect of the microscanner perturbations on both X- and Y- axes on the positioning error and consequently, on the probability of error in the bit sequence reproduced by the read-back signal in a single channel, is studied in [11].

A complete multiple-channels burst error model that describes the mechanism of burst errors that appear in a set of simultaneously accessed channels is shown in Fig. 4. The model consists of two complementary modules. The first one is related to the modeling of the external noise source, while the second module models the correlated burst errors in the multiple storage channels, using Markov processes. The common part of these two modules is the bit error probability observed on each storage channel due to the external noise effect.

More specifically, depending on the external noise characteristics, periods of high ($S[t] = 1$) and low ($S[t] = 0$) error probabilities are observed. In channels with intersymbol interference, it is also common that a change in a bit affects the probability of error of adjacent bits. The time sequence of these periods along with the error probabilities for each $S[t]$, are obtained by extensive measurements on an actual system. Since the various storage channels are affected by the same noise source, a common Markov process models the error probabilities in each channel. The probability that a bit in a channel is altered, resulting to an error in that channel, depends on the read/write mechanism, the statistical characteristics of the external noise source and the current state ('error' or 'no error') of that specific channel. This dependency on the external noise source expresses the spatial correlation among the errors in all channels.

This model can be used to study the effect of various

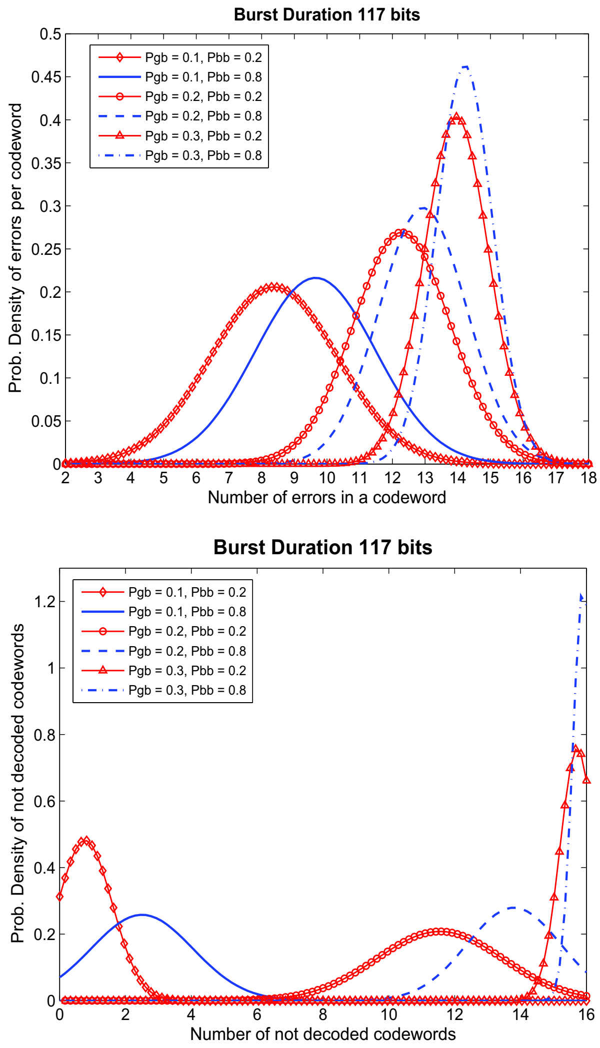


Figure 5. Probability distributions of the number of errors in a codeword and the number of not decoded codewords, when an external disturbance with $B_{av} = 117$ bits duration affects a system with $N = 16$ fields, a sector size of 2048 bytes and RS(151,129) code, for several values of P_{bb} and P_{gb}.

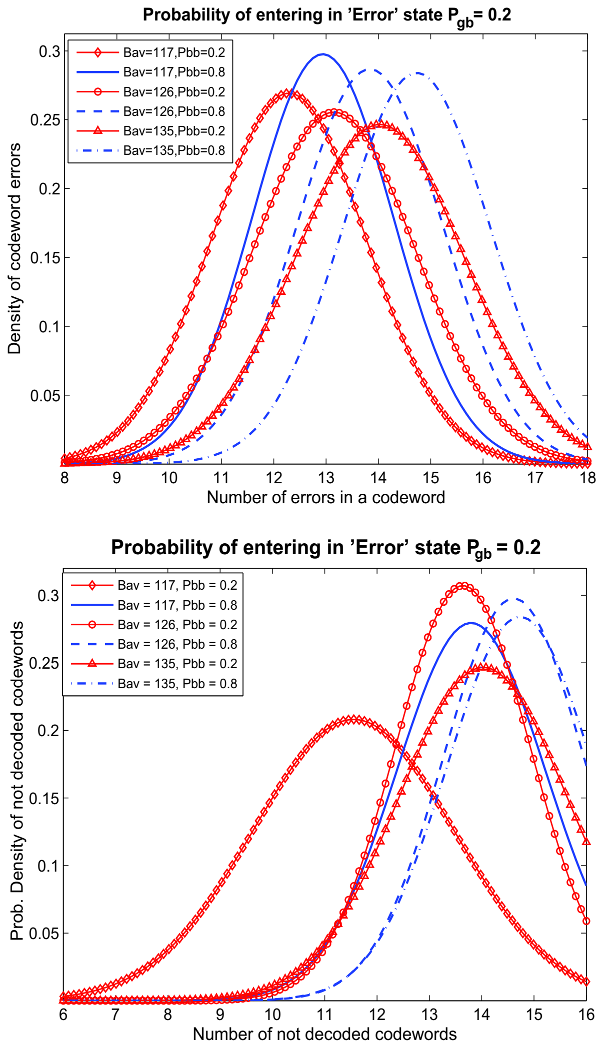
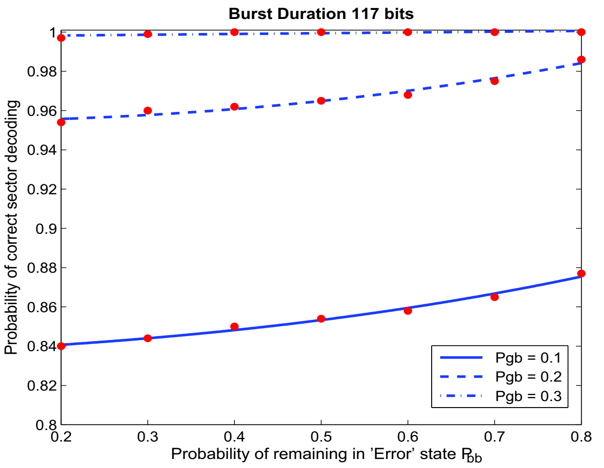


Figure 6. Probability distributions of the number of errors per codeword and the number of not decoded codewords, when an external disturbance affects a system with $N = 16$ fields, a sector size of 2048 bytes and RS(151,129) code, for $P_{gb} = 0.2$ and several values of P_{bb} and burst durations B_{av}.

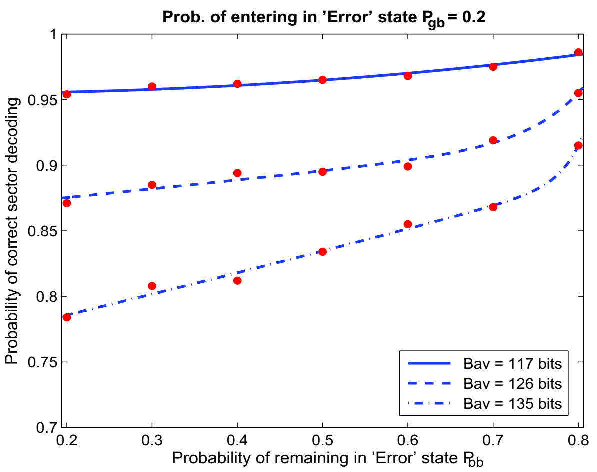
different realizations of an external disturbance to a storage device with multiple probes that operate in parallel, and derive various probabilistic distributions, such as, the distribution of errors in the codewords and the number of codewords that cannot be decoded. It can also be used to reproduce the burst errors that occur in the storage channels when an external disturbance with certain characteristics affects a probe-based storage device while reading a sector. This way we can evaluate the performance of the proposed erasure estimation approach, when the disturbance leads to a sector decoding failure, but at least one codeword was decoded successfully at the initial decoding phase.

3.1. Simulation Results

We consider a storage device with $N = 16$ storage fields, a sector size of 2048 bytes and RS(151,129) code. This means that a sector is divided in $M = 16$ codewords and the RS code can correct up to $t = 11$ errors. We study the case where, while reading a sector, an external shock or vibration has caused a displacement in the positioning of the tips that led to a burst of errors in all storage fields, at least one codeword was not decoded by a typical RS decoder and a few codewords have been decoded. The typical coding scheme results to a sector decoding failure, but we can apply the proposed heuristic method to estimate the locations of errors in the not-decoded codewords and extend the error correction capability of the RS code from t errors to at most $2t$ errors, using the errors-and-erasures approach.

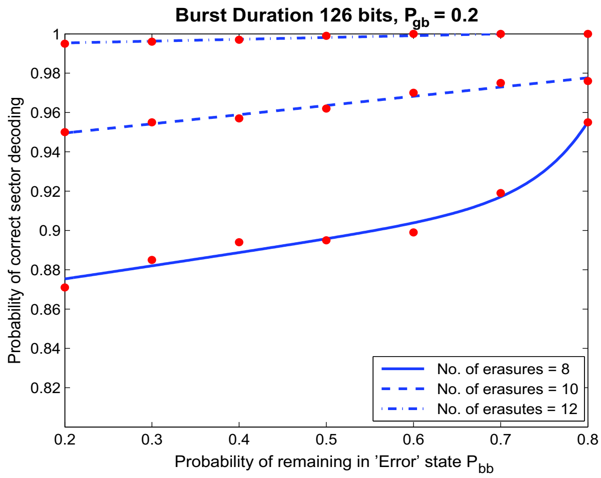


(a) Probability of correct sector decoding for various realizations of a burst with duration 117 bits.

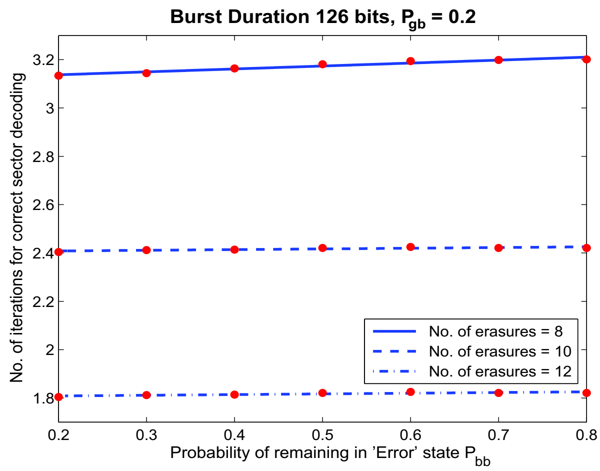


(b) Probability of correct sector decoding for various realizations of a burst with probability of entering in 'Error' State $P_{gb} = 0.2$ and different burst durations.

Figure 7. Performance of the erasure decoding algorithm when a system with $N = 16$ storage fields, a sector size of 2048 bytes and RS(151,129) code is affected by burst errors.



(a) Probability of correct sector decoding for different initial numbers of erasure symbols.



(b) Number of iterations required for correct sector decoding for different initial numbers of erasure symbols.

Figure 8. The effect of the initial number of erasure symbols on the performance of the erasure decoding algorithm when a system with $N = 16$ storage fields, a sector size of 2048 bytes and RS(151,129) code is affected by various realizations of a burst with probability of entering the 'Error' state $P_{gb} = 0.2$ and burst duration of 126 bits.

Fig. 5 and 6 show the probability distributions regarding the mean number of errors that appear in a codeword and the mean number of codewords that cannot be decoded for various scenarios, as they are derived from the statistics produced by the aforementioned multiple-channels burst error model. We assume that errors appear in the channels only when $S[t] = 1$, which means that $P_{gb}(S[0]) = 0$ and no random errors are introduced. This assumption is validated by measurements. Different realizations of the disturbance give different values for $P_{gb}(S[1]) = P_{gb}$ and $P_{bb}(S[1]) = P_{bb}$. According to these results, in these scenarios there is a great chance that the proposed algorithm can be applied, since for these burst durations there is a large probability that at least one codeword suffers from more that t errors, but the total number of errors in the codewords are less that $2t$. For small values of the probability of entering the 'Error' state P_{gb}, we expect a small number of codewords with more than t errors, while for larger values, almost all codewords are affected by more that t errors.

Fig. 7(a) and 7(b) show the improvement in the reliability of the device with the new decoding algorithm, for different realizations of the external disturbance, in terms of statistical characteristics and burst duration. According to these results, for a certain burst duration, the erasure flagging process leads to more reliable estimations when the external disturbance is more aggressive, thus resulting in higher values of the tips displacement and subsequently in higher values for P_{gb} and P_{bb} . This can be explained by the fact that high bit error rates mean that in the same bit location, there will be a large number of bits in error in all storage fields, which results to high erasure probabilities. Note that even one codeword decoding failure leads to a sector decoding error. This makes clearer the improvement that the proposed method achieves, since even $M-1$ codewords can be decoded correctly in some cases, by knowing only the error locations in one codeword. The performance deteriorates as the duration of the burst increases, but it still leads to a significant improvement of over 50% in the device reliability comparing to the conventional decoding process.

Finally, Figures 8(a) and 8(b) show how the number of erasure symbols that will be used by the errors-and-erasures decoding procedures affects the performance of the proposed method. The correct determination of the number of erasures can increase significantly not only the performance of the proposed method, but also the speed of the decoding process, since it affects the number of iterations that are needed to retrieve the sector correctly. Since the number of erasure symbols does not change seriously the complexity of the decoding circuits, it should be specified such that it can correct a small number of potential false erasure estimations. Measurements on an actual system can assist this definition.

4. CONCLUSIONS

We proposed a heuristic approach for erasure estimation and error correction in storage devices, that use multiple, simultaneously accessed parallel fields, when they are affected by burst errors caused by external disturbances. The presented algorithm exploits the parallelism of the multiple fields and the error locations revealed by the initial errors-only decoding attempt and identifies symbols as erasures. Simulation results show that the approach improves significantly the reliability of the coding scheme without increasing seriously the complexity of the decoding procedure.

5. REFERENCES

[1] A. Pantazi, A.Sebastian, *et al.*, "Probe-based ultrahigh-density storage technology," *IBM J. Res. and Dev.*, vol. 52, no. 4/5, pp. 493–511, 2008.

[2] P. Vettiger, T. Albrecht, M. Despont, *et al.*, "Thousands of Micro-Cantilevers for Highly Parallel and Ultra-Dense Data Storage," in *Proc. IEDM 2003 - IEEE Int'l Electron Devices Meeting 2003*, Washington, DC, Dec. 2003, pp. pp. 32.1.1 – 32.1.4.

[3] M. Varsamou and T. Antonakopoulos, "A new data allocation method for parallel probe-based storage devices," *IEEE Transactions on Magnetics*, vol. 44, no. 4, pp. 547–554, Apr. 2008.

[4] R. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1983.

[5] R. McEliece and L. Swanson, "Reed-Solomon codes and the exploration of the solar system," in *Reed-Solomon Codes and Their Applications*, S. B. Wicker and V. K. Bhargava, Eds. Piscataway, NJ: IEEE Press, 1994, pp. 25–40.

[6] K. Leung and L. Welch, "Erasure Decoding in Burst-Error Channels," *IEEE Transactions on Information Theory*, vol. 27, no. 2, pp. 160–167, Mar. 1981.

[7] E. Eleftheriou, T. Antonakopoulos, *et al.*, "The Millipede, a MEMS-based scanning-probe data-storage system," *IEEE Transactions on Magnetics*, vol. 39, no. 2, pp. 938–945, Mar. 2003.

[8] G. Binnig, C. F. Quate, and C. Gerber, "Atomic force microscope," *Phys. Rev. Lett*, vol. 56, no. 9, pp. 930–933, 1986.

[9] P. Vettiger, G. Cross, *et al.*, "The Millipede - Nanotechnology entering data storage," *IEEE Transactions on Nanotechnology*, vol. 1, no. 1, pp. 39–55, Mar. 2002.

[10] A.Sebastian, A. Pantazi, and H. Pozidis, "Jitter Investigation and Performance Evaluation of a Small-Scale Probe Storage Device Prototype," in *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, Washington, DC, USA, Nov. 2007, pp. 288–293.

[11] A.Sebastian, A. Pantazi, H. Pozidis, and E. Eleftheriou, "Nanopositioning for Probe-Based Storage Device," *IEEE Control Systems Magazine*, pp. 26–35, August 2008.