On Communication Over Channels With Varying Sampling Rate

Lara Dolecek
EECS Department
University of California
Berkeley, CA 94720, USA
Email: dolecek@eecs.berkeley.edu

Venkat Anantharam
EECS Department
University of California
Berkeley, CA 94720, USA
Email: ananth@eecs.berkeley.edu

***Abstract*— We survey some of our recent results on communication over channels with varying sampling rate. Varying sampling rate is caused by inadequate timing recovery and can result in a synchronization error, exhibited as a repetition or a deletion of a symbol. We review number-theoretic constructions and the upper bounds of sets of strings capable of overcoming certain number of synchronization errors. We also review results on how to use some practical substitution error correcting codes when the varying sampling rate causes a synchronization error. In particular we present the results for the Reed-Muller(1,m) code and the array-based LDPC codes.**

I. Introduction

Substitution error correcting codes are traditionally used in communication systems to encode a binary input message into a coded sequence. The modulated version of this coded sequence is typically sent over a noisy channel, which corrupts it by additive noise. At the receiver, the received waveform is sampled at certain sampling points that are determined by the timing recovery block. The resulting sampled sequence is passed to the decoder whose task is to determine the original binary message used in transmission. In the analysis of substitution error correcting codes and their decoding algorithms it is traditionally assumed that the decoder receives a sequence which is a properly sampled version of the received waveform. This follows from the assumption that the timing recovery block produces near optimal estimates of the sampling instances. To achieve this requirement in modern communication chips, the timing recovery block incurs a significant cost in terms of power dissipation and occupied chip area, which is typically around 20%, and potentially even more, [2]. Some advanced solutions for improved timing recovery (which may come at a complexity and delay cost) include: iterative timing recovery which incorporates the timing block into the turbo equalization loop [18], MAP detection based on the Markov chain modelling of timing offsets [24], and a multiple-hypothesis testing approach [13]. To reduce some of the cost (desirable in systems employing advanced signal processing algorithms for other components, for example for iterative decoding) while maintaining the required throughput, one may try to use a cheaper version of the timing recovery block, and instead exploit the substitution error correcting code and its decoder to overcome both inadequate synchronization as well as additive errors. We investigate this approach.

We adopt a set-theoretic model for the synchronization errors such that each codeword gives rise to a whole set of possible received sampled sequences, where the set depends on how accurate the timing recovery is. Even if we assume for the moment that the effect of the noise is negligible, under inadequate synchronization, a transmitted codeword can still result in many different possible sampled sequences. When two different codewords result in the same sampled sequence we say that they have an identification problem since it is no longer possible to uniquely determine the transmitted codeword. We associate with each codeword a set of strings obtained by repeating (deleting) an arbitrary bit in the codeword. We refer to the minimum Hamming distance between such sets as post-repetition (post-deletion) distance. It is desirable to have post-repetition (post-deletion) distance as large as possible. Since the error-correction codes are traditionally designed to overcome additive errors, the challenge lies in determining their performance when both synchronization and substitution errors are present and in finding a way to modify codes such that the modified code possesses improved immunity to synchronization errors without sacrificing rate and the decoding complexity.

A problem related to studying channels that introduce synchronization errors is the investigation of collections of strings that are immune to a certain number of synchronization errors. Specifically, one may be interested in finding the largest such collections immune to a prescribed number of synchronization errors, and in finding explicit constructions of such sets.

In the remainder we review some of our recent results and the existing relevant works on constructions of sets of binary strings capable of overcoming a certain number of synchronization errors and some results on how codes of practical interest and their decoding algorithms can be modified to be able to overcome synchronization as well as additive errors.

II. Number-theoretic constructions of a set of strings immune to synchronization errors

Binary single insertion/deletion codes were studied by Levenshtein [14] who established the asymptotic upper bound on the cardinality of a set of strings immune to a single insertion or a deletion to be $2^n/n$ for the codeword length

n. It was also shown in [14] that the so-called Varshamov-Tenengolts codes [22] originally introduced for the correction of an asymmetric error asymptotically reach this upper bound, though their optimality (or of any other explicit construction for that matter) for a specific finite blocklength remains an open problem. The exception are blocklengths of some powers of 2 for which it was shown exhaustively that the size of the largest set correcting single insertion/deletion is indeed that of the largest Varshamov-Tenengolts code for the given codeword length [1]. Using number-theoretic techniques, Sloane [19] derived exact cardinalities of these codes. Additional work on Levenshtein/Varshamov-Tenengolts-type codes includes extensions to the non-binary case [21]; correction of single peak shift [12]; connection with codes possesing spectral nulls [10]; and a code concatenation technique with the inner code being a small Varshamov-Tenengolts code [3].

Construction of codes capable of overcoming s insertions/deletions that reach their asymptotic upper bound, which is $s!2^n/n^s$ [14] for the codeword length n, or even come within a factor of that bound proves to be a difficult problem. A construction proposed by Helberg and Ferreira [11] has asymptotic cardinality $2^n/M(n)$ where $M(n)$ grows exponentially with n, indeed quite far from the upper bound. A greedy construction for overcoming two errors that appears to work only for very short lengths was proposed in [20].

A related problem is that of studying optimal codes capable of overcoming multiple repetitions. When studying such codes one may use a transformation that maps a binary string of length n into a string of length $n-1$ by taking mod 2 sum of each pair of consecutive bits. As a result, a repetition in the original domain translates into an insertion of a zero in the new domain. It is thus sufficient to study binary codes immune to insertions of only zeros.

These codes were also studied by Levenshtein, [15], where the upper bound on a set of strings immune to a single insertion of a zero was established to be $2^{n+1}/n$. In [6] we proposed a construction that asymptotically reaches this upper bound. Inspired by [19], we also provided exact cardinality results for this construction. In [6] we also presented another number-theoretic construction of a set of strings immune to s, for $s \geq 1$, insertions of zeros (and consequently a construction immune to s repetitions by inverting the aforementioned transformation). The lower asymptotic bound of the proposed construction is within a factor of the asymptotic upper bound derived in [15] which is of the type $c(s)2^{n+s}/n^s$, where the constant $c(s)$ depends on the number of inserted ones, s, and not on n. Our construction improves by at least a factor on the lower asymptotic bound of the construction in [15].

We now summarize the results on the performance of two specific additive error correcting codes in the presence of synchronization errors.

III. Reed-Muller codes

We evaluated the performance of Reed-Muller(1,m) codes over channels which in addition to a certain number of substitution errors permits a deletion or a repetition of a bit [7]. Reed-Muller(1,m) codes have minimum distance 2^{m-1} for codeword length $n = 2^m$ and dimension $m + 1$. When used over a channel that introduces additive noise, they are typically decoded using a bounded distance decoding algorithm based on the fast Hadamard transform of complexity $O(n \log n)$. We studied the run-length structure of these codes in detail and established several structural properties of these codes. These results appear to be new and thus may also be of independent interest. From this analysis we systematically determined the codeword pairs causing the identification problem under single deletion. We also showed that no codewords cause the identification problem under single repetition. Interestingly, the number of identification problem causing pairs is constant (the constant being 11) and not a function of the codeword length. By dropping only one information bit we were able to eliminate all identification problem causing codewords, and more importantly, we increased post deletion distance from 0 to 2^{m-3} as well as the post repetition distance from 2 to $2^{m-3} + 1$.

In addition, we proposed a bounded distance decoding algorithm, also based on the fast Hadamard transform, which recovers the transmitted codeword in the presence of a bit deletion (repetition) as well as substitution errors whose number is up to the half the minimum post-deletion (post-repetition) minimum distance. The proposed algorithm has the same complexity as the bounded distance decoding algorithm that deals with substitution errors only, namely $O(n \log n)$.

Additional structural properties and the study of Reed-Muller(1,m) codes over channels which in addition to a certain number of substitution errors simultaneously permit both a deletion and a repetition of a bit were reported in [8]. Again the number of identification problem causing codeword pairs was constant (the constant being 35) and by dropping two information bits all such codewords were eliminated.

We also briefly review the findings on the performance of array-based LDPC codes.

IV. Array-based LDPC codes

Array based LDPC codes are regular LDPC codes parameterized by integers j and p, where $j \leq p$, p is an odd prime, and have the parity check matrix $H_{p,j}$ consisting of a two dimensional array of circulants, [9]. They feature good performance, low encoding complexity and have been proposed for a variety of applications, including digital subscriber lines and magnetic recording applications [23].

However, as we have shown in [5], this code suffers from the identification problem under a single repetition. We proposed a technique for thinning array based LDPC codes which results in enhanced repetition error correction properties. The technique suffers only a small loss in the rate.

Furthermore in [5] we developed a modified message passing algorithm appropriate for channels in which in addition to additive errors, a repetition error also occurs. Even though the algorithm handles a repetition error as well as additive errors, with organized calculations, the complexity of this algorithm is the same as that of the traditional message passing, namely $O(n)$ for the codeword length n.

Extensions to multiple synchronization error correction (along with additive errors) based on embedding a codeword in a larger string such that the enlarged string is immune to multiple repetitions while keeping the rate loss minimal is in preparation.

V. Concluding Remarks

We surveyed some of our recent results as well as related literature relevant for the study of communication over channels when the sampling rate varies. These include modifications of some practical codes as well as constructions and the asymptotics of sets of strings immune to multiple synchronization errors.

Acknowledgment

This research was supported in part by NSF award CCF-0635372, Marvell Semiconductor Inc., University of California MICRO program, Mentored Research Award and the Dissertation Year Fellowship.

References

[1] http://www.research.att.com/˜njas/doc/graphs.html Challenge Problems: Independent Sets in Graphs, maintained by N. J. A. Sloane

[2] J. Ammer, "Low Power Synchronization for Wireless Communication", Ph.D. dissertation, University of California, Berkeley, 2004.

[3] G. Chen, M. Mitzenmacher, C. Ng and N. Varnica, "Concatenated codes for deletion channels," In *Proc. of the IEEE International Symposium on Information Theory* 2003, Yokohama, Japan.

[4] M.C. Davey and D.J.C. MacKay, "Reliable communication over channels with insertions, deletions and substitutions", *IEEE Trans. on Information Theory* vol. 47(2), pp. 687-698, Feb. 2001.

[5] L. Dolecek and V. Anantharam, "A synchonization technique for array-based LDPC codes", *International Symposium on Information Theory*, Seattle, WA, July 9-13, 2006.

[6] L. Dolecek and V. Anantharam, "On subsets of binary strings immune to multiple repetition errors", submitted for publication.

[7] L. Dolecek and V. Anantharam, "Using Reed-Muller(1,m) codes over channels with synchronization and substitution errors", to appear in *IEEE Transactions on Information Theory*.

[8] L. Dolecek, "On structrural properties of Reed-Muller codes and their use in channels with synchronization and substitution errors", Report No. EECS-2006-43, EECS Department, University of California, Berkeley, 2006.

[9] J. L. Fan, "Array-codes as low-density parity-check codes", in *Proc. of the 2nd International Symposium on Turbo Codes and Related Topics*, Brest, France, Sept. 2000.

[10] H.C. Ferreira, W.A. Clarke, A.S.J. Helberg, K.A.S. Abdel-Gaffar and A.J. Han Vinck, "Insertion/deletion correction with spectral nulls", *IEEE Trans. on Information Theory* vol. 43, pp. 722–732, March 1997.

[11] A. S. J. Helberg and H. C. Ferreira, "On multiple insertion/deletion correcting codes", *IEEE Transactions on Information Theory* vol. 48(1), pp. 305–308, Jan. 2002.

[12] T. Kløve "Codes correcting a single insertion/deletion of a zero or a single peak-shift", *IEEE Trans. on Information Theory* vol. 41(1), pp. 279–283, Jan. 1995.

[13] P. Kovintavewat, J. R. Barry, M. F. Erden and E. Kurtas, "Per-survivor timing recovery for uncoded partial response channels,"*Proceedings of the IEEE International Conference on Communications* 2004, Paris, France.

[14] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals", *Sov. Phys.-Dokl.*, vol. 10(8), pp. 707–710, Feb. 1966.

[15] V. I. Levenshtein, "Binary codes capable of correcting spurious insertions and deletions of ones", *Problems of Information Transmission*, vol. 1(1),pp. 8–17, Jan. 1965.

[16] V. I. Levenshtein, "On perfect codes in deletion and insertion metric", *Discrete Math. Appl.*, vol. 2(3), pp. 241–258, 1992.

[17] T. Mittelholzer, "Efficient encoding and minimum distance bounds of Reed-Solomon-type array codes", In *Proc. of the IEEE International Symposium on Information Theory* 2003, p. 282, Lausanne Switzerland.

[18] A. R. Nayak, J. Barry and S. W. McLaughlin, "Joint timing recovery and turbo equalization for coded partial response channels," *IEEE Trans. On Magnetics*, vol. 38(5), pp. 2295–97, Sept. 2002.

[19] N.J.A. Sloane, "On single deletion correcting codes", 2000. Available at http://www.research.att.com/˜njas/doc/dijen.pdf

[20] T. G. Swart and H.C. Ferreira, "A note on double insertion/deletion correcting codes", *IEEE Transactions on Information Theory* vol. 49(1), pp. 269–273, Jan. 2003.

[21] G. Tenengolts, "Nonbinary codes correcting single deletion or insertion", *IEEE Trans. on Information Theory* vol. 30, pp. 766–769, Sept. 1984.

[22] R. R. Varshamov and G.M. Tenengolts, "Codes which correct single asymmetric errors," *Avtomatika i Telemehkanika*, vol. 26, no. 2, pp. 288–292,1965.

[23] B. Vasic and E. Kurtas, "Coding and Signal Processing for Magnetic Recording Systems", CRC press, 2005.

[24] W. Zeng and A. Kavcic, "MAP detection in noisy channels with syncronization errors (including the insertion/deletion channel), In *Proc. of the IEEE International Symposium on Information Theory* 2003, Yokohama, Japan.