On Guided Scrambling with Guaranteed Maximum Run-Length Constraints

Adriaan J. van Wijngaarden and Kees A. Schouhamer Immink

Abstract

Guided scrambling generates for each possible source word a unique set of candidate codewords and selects the "best" word subject to certain given channel constraints. This is an effective technique to generate long and highly efficient codes that satisfy the given channel constraints with high probability. These codes are referred to as weak constrained codes, because the guided scrambling method cannot guarantee that all constraints are satisfied. One of the common constraints, the maximum run-length constraint, is of particular importance, because sequences that violate this constraint are likely to cause loss of timing. For this reason, methods are developed in conjunction with guided scrambling to impose a guaranteed maximum run-length constraint. The performance of combined guided scrambling and maximum run-length limited codes is analyzed. It will be demonstrated that the combination of guided scrambling and a well-chosen maximum run-length limited code may offer a sound trade-off between overall code rate and performance in terms of the probability of violating the channel constraints.

Index Terms

Recording systems, run-length constraints, guided scrambling, weakly constrained codes.

I. Introduction

Constrained codes, such as run-length limited codes and DC-free codes, have found widespread application in data storage and communication systems, see [1] and references therein. Runlength-limited sequences, also known as (d, k) sequences, are binary sequences with at least d and at most k 'zeros' between consecutive 'ones'. Of particular interest is the subclass of $(0, k)$ sequences, used extensively in recording and digital communication systems for clock recovery and synchronization. These sequences are generally constructed by concatenating words of a binary block code of length n, each fulfilling $(0, k, k_l, k_r)$ constraints, where k_l and k_r denote the maximum number of leading and trailing zeros ($k_l + k_r \leq k$). These codes are referred to as maximum run-length limited (MRL) codes.

Codes that do not strictly guarantee the fulfillment of specified channel constraints are often called *weakly constrained codes*. Guided scrambling [1] is a typical example of an embodiment of a weakly constrained code. In guided scrambling, each source word of length m, $m \gg 1$, is supplemented by

A. J. (de Lind) van Wijngaarden is with Bell Laboratories, Lucent Technologies, Room MH 2C-360, 600 Mountain Ave, Murray Hill, NJ 07974, U.S.A. (e-mail: alw@research.bell-labs.com).

K. A. Schouhamer Immink is with the Institute for Experimental Mathematics, Ellernstr. 29, Essen, Germany (e-mail: immink@exp-math.uni-essen.de).

r_s bits. The r_s supplement bits make it possible to generate a selection set of $L = 2^{r_s}$ pseudo-random sequences of length $n_s = r_s + m$. The guided scrambling method typically generates a selection set of L pseudo-random sequences that comprise sufficiently distinct and random words.

Constrained codes can generally not correct and at best detect transmission errors. Therefore, short byte-oriented block codes have typically been used in systems that are susceptible to errors, in conjunction with symbol error correcting codes. With guided scrambling, one can easily generate long and efficient sequences. To make these sequences resilient against errors, effective error control can be provided using techniques described in [5] and references therein.

Evidently, the drawback of weakly constrained codes is their sensitivity for specific worst case source data. In this article, we will study the combination of guided scrambling and high-rate maximum run-length constrained $(0, k_0)$ codes, that serves as a 'safety net' to be able to guarantee a prescribed maximum run-length, k_0. To illustrate the effectiveness of the proposed methods, we will calculate the probability that a transmitted codeword will violate a maximum run-length k, where $k < k_0$.

II. Combined Guided Scrambling and MRL Coding

Let $u^{(m)}$ denote the source word of length m. The guided scrambling scheme adds r_s supplement bits and generates a selection set of $L = 2^{r_s}$ pseudo-random sequences $x_1^{(n_s)}, \cdots, x_L^{(n_s)}$ of length $n_s = r_s + m$. There are now several possibilities to impose the maximum run-length constraint k_0, two of which are depicted in Figure 1 and Figure 2. In the first case, the L sequences $x_1^{(n_s)}, \cdots, x_L^{(n_s)}$ are first converted

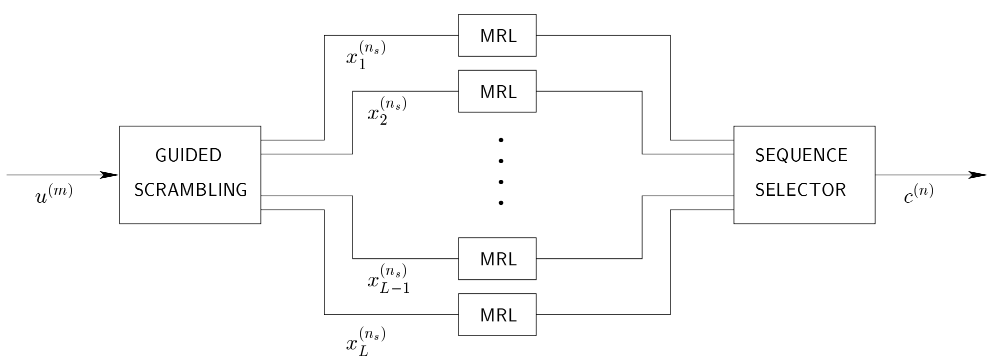

Fig. 1. Possible realization of combined guided scrambling and MRL coding, where one of the L scrambled $(0, k_0)$ sequences is selected.

into $(0, k_0)$ sequences of length n, after which one, $c^{(n)}$, is selected. In the second case, the coding process is performed sequentially. In this case, the sequence selector may anticipate the presence of the MRL coding block.

There are several more variations, in particular when the schemes are also combined with the modulation and error control schemes presented in [5].

The high rate $(0, k, k_l, k_r)$ codes presented in [5], [6] are particularly suitable for the construction of higher rate codes because of their tight constraints. This is achieved by interspersing the codewords with

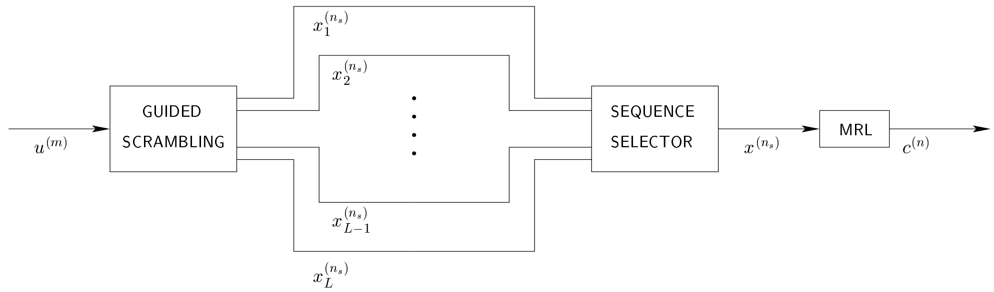

Fig. 2. Possible realization of combined guided scrambling and MRL coding, where the sequence selector anticipates post-processing with an MRL code.

uncoded source symbols as described in [4], [5]. One can get an additional degree of freedom in the selection of the scrambled sequences by using a strategy where the interspersed bits are observed and the source symbols of the underlying constrained code are transformed only if the overall constraints are violated, or, interestingly, the encoder may select one of the two possible sequences based on the overall constraints.

III. Performance Analysis

In this study, we assume that the $L = 2^{r_s}$ length-n_s sequences are translated by a well-chosen constrained encoder into a selection set of L $(0, k_0)$ constrained length-n sequences. To exemplify the performance of guided scrambling with guaranteed maximum run-length constraints, we consider the situation where the sequence selection criterion is the maximum run-length. In this situation the encoder selects and transmits the sequence with the shortest maximum run-length, k, where $k < k_0$. The above procedure guarantees that the maximum run-length of transmitted sequences never exceeds k_0, but is usually less than k_0. Let $R = m/(r_s + m)$ and let R_0 denote the rate of the $(0, k_0)$ constrained code. Then the overall rate of the coding scheme is $R_0 m/(r_s + m) = R_0 R$.

We will compute the probability that a length-n $(0, k_0)$ sequence generated by guided scrambling violates the k constraint, where $k < k_0$. According to Shannon [2], the number $N(n)$ of $(0, k)$ constrained codewords of length n can be approximated by

$$N(n) \approx A \cdot 2^{nC(0,k)}, \tag{1}$$

where A is a constant and $C(0, k)$ denotes the capacity of the $(0, k)$ channel. Note that $C(0, k) = \log_2(\lambda)$, where λ is the largest real root of the characteristic equation [1]

$$z^{k+2} - 2z^{k+1} + 1 = 0. \tag{2}$$

The probability that in L drawings from randomly generated $(0, k_0)$ sequences we will not find any $(0, k)$ sequence, $k_0 > k$, is

$$p = (1 - p_0)^L, \tag{3}$$

where

$$p_0 \approx B \cdot 2^{(C(0,k_0)-C(0,k))n}, \tag{4}$$

and B is a quantity, which depends on k and k_0. As $L = 2^{r_s}$ and $r_s = (1-R)n$, we have

$$p = \left(1 - B \cdot 2^{(C(0,k_0)-C(0,k))n}\right)^{2^{(1-R)n}}. \tag{5}$$

If, for simplicity, it is assumed that $B \cdot 2^{(C(0,k_0)-C(0,k))n} \ll 1$, we have

$$\ln(p) = -B \cdot 2^{(C(0,k_0)-C(0,k)+1-R)n}. \tag{6}$$

In the (patent) literature several excellent high rate $(0,k)$ constrained codes can be found [3]–[6]. Patapoutian *et al.* [3] presented a rate 32/33, (0,6) code. This code can be interleaved to generate a rate 64/65 (0,10) code, which looks attractive for our purpose. Figure 3 shows the violation probability, p, that no sequence taken from a selection set of size L of (0,10) sequences obeys the $(0,k)$ constraint with the codeword length n as a parameter. The overall code redundancy is $1-(64/65)\cdot(99/100) = 2.5\%$.

We may notice that for longer codewords, and thus larger selection sets, the violation decreases more rapidly.

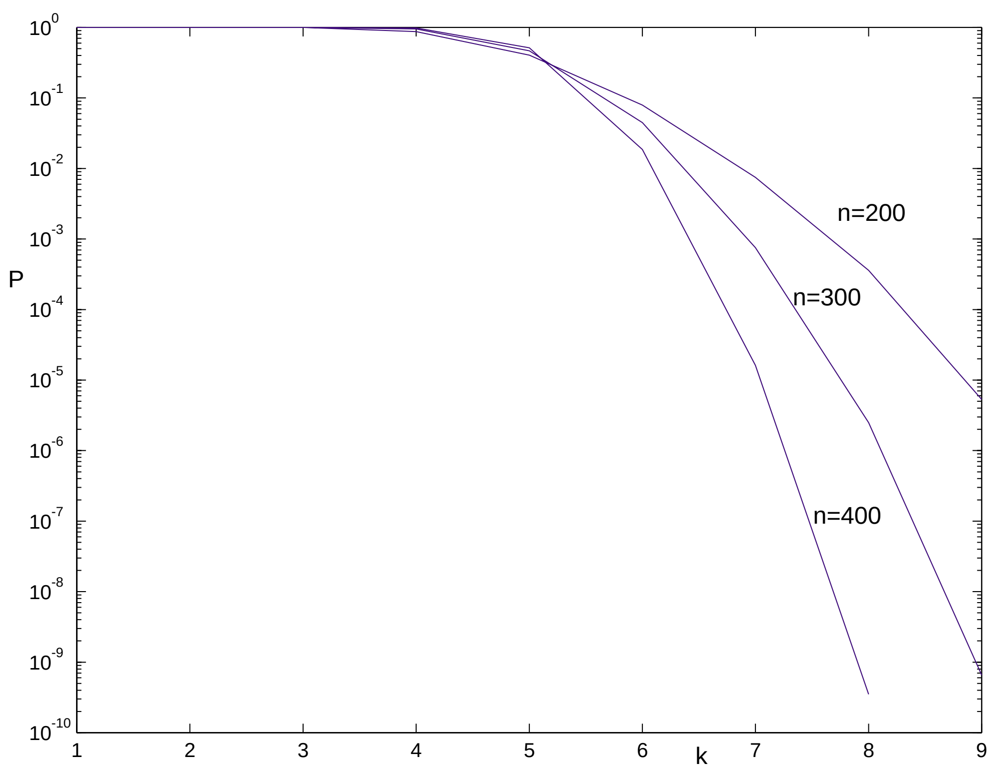

Fig. 3. Probability, p, that no sequence of a sample of size $L = 2^{n/100}$ of random $(0,10)$ sequences of length n satisfies the $(0,k)$ constraint versus k with the codeword length n as a parameter. The overall code redundancy is 2.5%.

IV. Conclusion

We have demonstrated that the combination of guided scrambling and a well-chosen $(0, k_0)$ constrained code may offer a sound trade-off between overall code rate and performance. This is exemplified for combined guided scrambling and maximum run-length limited coding where the sequence selection criterion is the minimization of the probability of violating the maximum run-length k constraint, where $k < k_0$.

References

[1] K.A.S. Immink, *Codes for Mass Data Storage Systems*, ISBN 90-74249-23-X, Shannon Foundation Publishers, Geldrop, The Netherlands, 1999.

[2] C.E. Shannon, "A Mathematical Theory of Communication", *Bell Syst. Tech. J.*, vol. 27, pp. 379-423, July 1948.

[3] A. Patapoutian, J. Stander, P. McEwen, B. Zafer, and J. Fitzpatrick, "Rate 32/33, $(d = 0, k = 6)$ runlength limited modulation code having optimized error propagation", US Patent 6,184,806, Feb. 2001.

[4] A. J. van Wijngaarden and K. A. S. Immink, 'Combinatorial construction of high rate runlength-limited codes," *IEEE Global Telecommun. Conf. GLOBECOM'96*, pp. 343–347, 1996.

[5] A.J. van Wijngaarden and K.A.S. Immink, "Maximum run-length limited codes with error control capabilities", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 4, Apr. 2001, pp. 602–611.

[6] A.J. van Wijngaarden and E. Soljanin, "A combinatorial technique for constructing high-rate MTR-RLL codes", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 4, Apr. 2001, pp. 582–588.