Low-Complexity error correction in LDPC Codes with Constituent RS Codes

Victor Zyablov,Vladimir Potapov, Fedor Groshev
Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 101447, Russia
E-mail: zyablov@iitp.ru, potapov@iitp.ru, groshev@iitp.ru

***Abstract*—Reed-Solomon code-based LDPC (RS-LDPC) block codes are obtained by replacing single parity-check codes in Gallager's LDPC codes with Reed-Solomon constituent codes. This paper investigates asymptotic error correcting capabilities of ensembles of random RS-LDPC codes, used over the binary symmetric channel and decoded with a low-complexity hard-decision iterative decoding algorithm. The number of required decoding iterations is a logarithmic function of the code length. It is shown that there exist RS-LDPC codes for which such iterative decoding corrects any error pattern with a number of errors that grows linearly with the code length. The results are supported by numerical examples, for various choices of code parameters.**

I. Introduction

Long block codes can be obtained by combining one or more simpler codes in various types of concatenated structures. Such constructions are of interest since they can yield powerful codes with good error-correcting capabilities, which are decodable with low complexity, using simple constituent decoders as separate modules.

A method for constructing long codes from short constituent codes, based on bipartite graphs, was introduced by Tanner in [1]. In this method, one of the two sets of nodes in a bipartite graph is associated with code symbols, while the other set is associated with constituent block codes of length equal to the node degree. These two sets of nodes are hereinafter referred to as variable nodes and constraint nodes, respectively. Tanner's general code construction unifies many known code families that can be obtained by choosing different underlying bipartite graphs and associating different constituent codes with its constraint nodes. For example, Gallager's Low-Density Parity-Check (LDPC) codes [2], and woven graph codes [3] can all be described using a bipartite graph-based approach.

For Gallager's LDPC codes [2], each constraint node in the corresponding bipartite graph represents a single parity-check (SPC) code over the variable nodes connected to it. In this case, the parity-check matrix of the code coincides with the adjacency matrix[1] of the corresponding bipartite graph. If the degree of each node is very small compared to the number of variable nodes (code length) the parity-check matrix is sparse. When the bipartite graph is regular, all variable nodes have degree j and all constraint nodes have degree k. Then the parity-check matrix contains j ones in each column and k ones in each row, and it specifies a (j, k)-regular LDPC code.

The error-correcting capabilities of LDPC codes for the binary symmetric channel (BSC) were studied in [4], where it was shown that there exist LDPC codes capable of correcting a portion of errors that grows linearly with the code length n, with decoding complexity $\mathcal{O}(n \log n)$. A similar result for expander codes was proven in [5].

The SPC codes associated with constraint nodes in the Tanner graph of an LDPC code can be replaced with other constituent block codes (*e.g.* Reed-Solomon codes [6]), which yields alternative constructions of LDPC codes, often referred to as generalized LDPC codes. The parity-check matrix of such an LDPC code is obtained by replacing every 1 in the graph's adjacency matrix with a column of the constituent code's parity-check matrix, and every 0 with an all-zero column.

In this paper, we consider the asymptotic performance of random RS-LDPC codes, when the code length n grows to infinity. We will prove that there exist RS-LDPC codes which, when decoded with a simple iterative decoder of complexity $\mathcal{O}(n \log n)$, can correct any error pattern with a number of errors growing linearly with the code length. Our approach builds upon the work of [4] where such a result was proved for LDPC codes with constituent SPC codes which have minimum distance $d_0 = 2$. A similar result holds for expander codes if the constituent codes have large enough minimum distance, *cf.* [5]. The work presented here, with constituent Reed-Solomon codes of minimum distance $d_0 = 3$, is a step towards 'closing the gap' between these two results.

II. Construction and Properties of RS-LDPC Codes

An (n_0, k_0, d_0) extended Reed-Solomon code has length $n_0 = 2^q$, dimension $k_0 = n_0 - d_0$-1, code rate $R_0 = 1$-$(d_0$-$1)/n_0$. We will consider single-error correcting extended RS code with minimum distance $d_0 = 3$,

A parity-check matrix H_0 of a Reed-Solomon code is an $(d_0$-$1) \times n_0$ matrix whose columns are all nonzero q-nary $(d_0$-$1)$-tuples. We will consider RS-LDPC codes with identical constituent codes. Let $\boldsymbol{H}_\mathrm{b}$ denote a block-diagonal matrix with the b constituent parity-check matrices $\boldsymbol{H}_0$ on the main

[1]Here it is assumed that the adjacency matrix $\boldsymbol{A}$ of a bipartite graph with two vertex sets $\mathcal{V}_1$ and $\mathcal{V}_2$ is a $|\mathcal{V}_1| \times |\mathcal{V}_2|$ binary matrix specifying connections among vertices, that is, $(\boldsymbol{A})_{ij} = 1$ iff nodes $v_i \in \mathcal{V}_1$ and $v_j \in \mathcal{V}_2$ are connected with a branch.

diagonal, that is,

$$\boldsymbol{H}_{\mathrm{b}} = \begin{pmatrix} \boldsymbol{H}_0 & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_0 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{H}_0 \end{pmatrix} \tag{1}$$

where b is very large. The matrix $\boldsymbol{H}_{\mathrm{b}}$ is of size $bm \times bn_0$. Let $\pi(\boldsymbol{H}_{\mathrm{b}})$ denote a random column permutation of $\boldsymbol{H}_{\mathrm{b}}$. Then the matrix constructed using $\ell \geq 2$ such permutations as *layers*,

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \\ \vdots \\ \boldsymbol{H}_\ell \end{pmatrix} = \begin{pmatrix} \pi_1(\boldsymbol{H}_{\mathrm{b}}) \\ \pi_2(\boldsymbol{H}_{\mathrm{b}}) \\ \vdots \\ \pi_\ell(\boldsymbol{H}_{\mathrm{b}}) \end{pmatrix} \tag{2}$$

is a sparse $\ell bm \times bn_0$ parity-check matrix which characterizes the ensemble of Reed-Solomon code-based LDPC codes of length $n = bn_0$, where $n \gg n_0$. Let $\mathscr{C}(n_0, \ell, b)$ denote this ensemble. For a given constituent Reed-Solomon code with parity-check matrix $\boldsymbol{H}_0$, the elements of the ensemble $\mathscr{C}(n_0, \ell, b)$ are obtained by sampling independently the permutations π_l, $l = 1, 2, ..., \ell$, which are all equiprobable. The rate of a code $\mathcal{C} \in \mathscr{C}(n_0, \ell, b)$ is lower-bounded by [1]

$$R \geq 1 - \frac{\ell b(n_0 - k_0)}{n} = 1 - \ell(1 - R_0) \tag{3}$$

with equality iff the matrix $\boldsymbol{H}$ has full rank. This imposes a restriction on the rate of the constituent codes, namely,

$$R_0 > 1 - \frac{1}{\ell}$$

that is, the more layers there are, the higher the rate of the constituent codes must be.

The RS-LDPC codes from the ensemble $\mathscr{C}(n_0, \ell, b)$ contain ℓb constituent Reed-Solomon codes; b in each layer. Such RS-LDPC codes can be represented by a Tanner graph [1] with $n = bn_0$ variable nodes, and ℓb constraint nodes, as illustrated in Figure 1. Each constraint node comprises $n_0 - k_0$ parity-check constraints specified by the rows of the corresponding constituent parity-check matrix. If a codesymbol is checked by a constituent code (that is, by at least one row of its parity-check matrix), there is a branch connecting the corresponding variable node and the constraint node. Each codesymbol is checked by exactly one Reed-Solomon code in each layer. The graph is regular, with the variable-node degree equal to ℓ, and the constraint-node degree equal to n_0. Such a graph is a special type of expander [7], where it is required that the ℓ constraint nodes adjacent to each variable node all belong to different layers.

Consider communication over a binary symmetric channel (BSC) using RS-LDPC codes with hard-decision decoding. Let $\boldsymbol{v}$ be the transmitted codeword and $\boldsymbol{e}$ be the error pattern. Then the received sequence is given by $\boldsymbol{r} = \boldsymbol{v} + \boldsymbol{e}$. The weight of the error sequence is $W = |\boldsymbol{e}|$ and the fraction of erroneous symbols is $\omega = W/n$. For code length $n \to \infty$, the fraction of erroneous symbols ω converges in probability to the crossover probability of the BSC.

For a given error pattern with W errors, we introduce the ℓ-tuple $\boldsymbol{a} = (a_1\ a_2\ ...\ a_\ell)$, where a_l, $l = 1, 2, ..., \ell$, denotes the number of constituent codes at the lth layer whose codewords are affected by errors. Note that $\boldsymbol{a}$ contains realizations of ℓ independent random variables that are integer-valued in the range $0 \leq a_l \leq b$, $l = 1, 2, ..., \ell$. Furthermore, let a denote the total number of constituent codes affected by errors, that is,

$$a = |\boldsymbol{a}| = \sum_{l=1}^{\ell} a_l.$$

In other words, a is the number of constraint nodes in the Tanner graph that are connected to at least one variable node with an erroneously received value.

III. Decoding Algorithm

Consider an iterative hard-decision decoding algorithm $\mathscr{A}$, whose decoding iterations i, $i = 1, 2, ..., i_{\max}$, consist of the following two steps:

1) For the tentative sequence $\boldsymbol{r}^{(i)}$, where $\boldsymbol{r}^{(1)}$ is the received sequence $\boldsymbol{r}$, decode independently ℓb constituent Reed-Solomon codes (that is, compute their syndromes $\boldsymbol{s}_{j,l}$, $j = 1, 2, ..., b$, $l = 1, 2, ..., \ell$, and if the value is nonzero, output the n_0-tuple where the position indicated by the syndrome is flipped). This yields ℓ independent decisions for each of the n symbols.

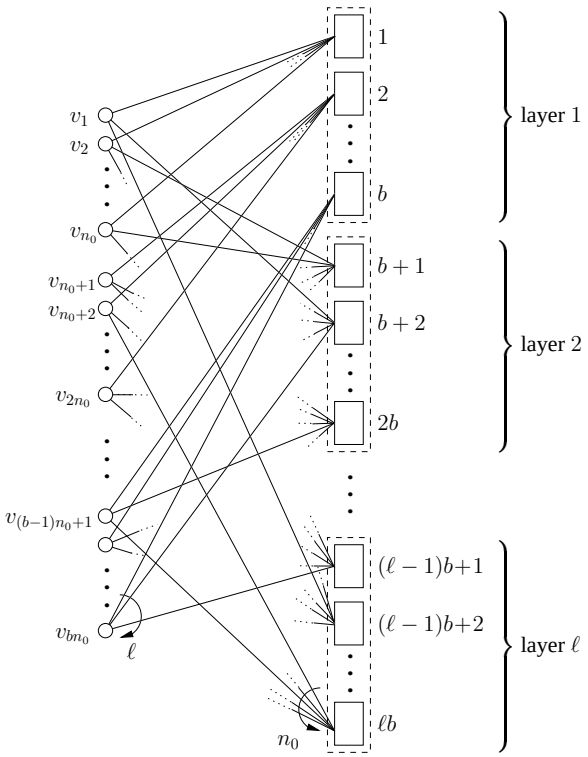

Fig. 1. Tanner graph of an RS-LDPC code defined by the parity-check matrix $\boldsymbol{H}$ given in (2). The graph illustrates the case when the first layer of $\boldsymbol{H}$ is the matrix $\boldsymbol{H}_{\mathrm{b}}$ itself, *i.e.*, $\pi_1(\boldsymbol{H}_{\mathrm{b}}) = \boldsymbol{H}_{\mathrm{b}}$ (then the b constraint nodes in layer 1 are connected to the consecutive blocks of n_0 variable nodes). Other layers are obtained with arbitrary permutations.

2) Flip every symbol $r_k^{(i)}$, $k = 1, 2, ..., n$, in the sequence $\boldsymbol{r}^{(i)}$, for which *at least one* of the ℓ decisions requires that. This yields the updated sequence $\boldsymbol{r}^{(i+1)}$.

Assume that the error pattern $\boldsymbol{e}$ is such that the number of errors that can be corrected by the constituent codes is larger than the number of uncorrectable errors. Then, during the first iteration of the algorithm $\mathscr{A}$, all correctable errors will be corrected, while the uncorrectable ones will result in erroneous decodings. Since Reed-Solomon codes are perfect single-error correcting codes with covering radius $\rho = 1$, each erroneous decoding will introduce at most one new error. Hence, the new error pattern, resulting from one decoding iteration has fewer errors than the initial error pattern. Clearly, if in each of the following iterations, the number of correctable errors is larger than the number of uncorrectable ones, then the total number of errors in $\boldsymbol{r}^{(i)}$ will decrease with the iteration number i and the algorithm yields the correct decision, *i.e.*, $\boldsymbol{r}^{(i_{\max})} = \boldsymbol{v}$. Then, we can state the following

Lemma 1: For any RS-LDPC code from the ensemble $\mathscr{C}(n_0, \ell, b)$, if an error pattern is such that in each iteration of algorithm $\mathscr{A}$ the number of errors correctable by the constituent codes is larger than the number of uncorrectable errors, then algorithm $\mathscr{A}$ yields a correct decision after $\mathcal{O}(\log n)$ iterations, where $n = bn_0$ is the code length.

Proof: Let $W = \omega n$ be the weight of the error pattern, and let ε denote a lower bound on the fraction of errors that are corrected in each iteration, $0 < \varepsilon < 1$. Then, after x iterations, the number of remaining errors is at most $\omega n(1-\varepsilon)^x$. The final decoding iteration $i_{\max}$ is reached when

$$\omega n(1-\varepsilon)^{i_{\max}} \leq 1$$

that is,

$$\log(\omega n) + i_{\max}\log(1-\varepsilon) \leq 0$$

which yields

$$i_{\max} \leq \frac{1}{\log\left(\frac{1}{1-\varepsilon}\right)}\log(\omega n). \tag{4}$$

Thus, the number of iterations is a logarithmic function of the code length. ∎

The complexity of each decoding iteration of the algorithm $\mathscr{A}$ is proportional to the code length n. Thus, according to Theorem 1, the overall decoding complexity is $\mathcal{O}(n\log n)$, given that the number of correctable errors in the error pattern is larger than the number of the uncorrectable ones. The following lemma formulates a condition under which this holds.

Lemma 2: If for any error pattern with $w \leq W$ errors, the number of constituent Reed-Solomon codes of an RS-LDPC code from the ensemble $\mathscr{C}(n_0, \ell, b)$ that are affected by errors is $a = \alpha w\ell$ with $\alpha > 2/3$, then the number of correctable errors in any such error pattern is always larger than the number of uncorrectable errors.

In other words, $\alpha > 2/3$ specifies the necessary expansion of the Tanner (expander) graph of the code [7], which ensures that the number of errors decreases in each iteration of algorithm $\mathscr{A}$.

Proof: The W variable nodes with erroneously received values are connected via $W\ell$ branches to $a \leq W\ell$ constraint nodes. If $a > (2/3)W\ell$, then more than $(2/3)W\ell$ branches reach distinct constraint nodes, while the remaining less than $(1/3)W\ell$ branches arrive to constraint nodes that are already connected with one branch to a variable node with an erroneously received value. Thus, out of the a constraint nodes, more than $1/2$ is connected to only one variable node with an erroneously received value, which is a correctable error pattern for a Reed-Solomon code, while less than $1/2$ have uncorrectable errors. ∎

Note that there is an important difference between the algorithm $\mathscr{A}$ and the decoding algorithm considered in [4]: in [4], a *majority rule* is applied for each symbol, that is, a symbol is flipped only if more than $\ell/2$ constituent SPC codes requires that. In the algorithm $\mathscr{A}$, however, a symbol is flipped as soon as *at least one* constituent Reed-Solomon code requires that.

IV. Asymptotic Performance

As shown in the previous section, the iterative algorithm $\mathscr{A}$ corrects any error pattern with W or fewer errors, if the code's Tanner graph has the expansion coefficient $\alpha > 2/3$. The question that arises, however, is whether such a code exists in the ensemble $\mathscr{C}(n_0, \ell, b)$. The following theorem allows us to receive the positive answer.

Theorem 1: In the ensemble $\mathscr{C}(n_0, \ell, b)$ of RS-LDPC codes, there exist codes (with probability p, where $\lim_{n\to\infty} p = 1$), which can correct any error pattern of weight up to $\omega_\alpha n$, with decoding complexity $\mathcal{O}(n\log n)$. The value ω_α is the largest root of the equation

$$h(\omega) - \ell F(\alpha, \omega, n_0) = 0 \tag{5}$$

where $h(\omega) = -\omega\log_2\omega - (1-\omega)\log_2(1-\omega)$ and the function $F(\alpha, \omega, n_0)$ is given by

$$F(\alpha, \omega, n_0) \triangleq h(\omega) + \omega\log_2(q-1) - \frac{1}{n_0}h(\alpha\omega n_0) + \max\left\{\omega\log_2 s - \alpha\omega\log_2\left((1+s(q-1))^{n_0} - 1\right)\right\} \tag{6}$$

where $\alpha > 2/3$ and the maximization is performed over all s such that

$$s > \frac{1}{(1-\alpha\omega n_0)^{1/n_0}} - 1.$$

Proof: For a *fixed* combination of $W = \omega n$ errors, the probability that the number of constituent Reed-Solomon codes of an RS-LDPC code from the ensemble $\mathscr{C}(n_0, \ell, b)$ that

are affected by errors, will not exceed a certain value $\alpha W \ell$ is upper-bounded by:

$$P(a \leq \alpha W \ell) \leq 2^{-n\ell F(\alpha,\omega,n_0)} \tag{7}$$

where the function $F(\alpha, \omega, n_0)$ is given by (6). The proof of this statement follows Appendix 1 in [4] and is omitted here for brevity.

Now consider the probability that the number of constituent codes affected by errors is at most $\alpha W \ell$ for *any* error pattern of a given weight W. If this probability is smaller than 1, then there exist codes in the ensemble $\mathscr{C}(n_0, \ell, b)$ for which $a > \alpha W \ell$ for any weight-W error pattern. Thus, the existence of such codes is ensured if

$$\binom{n}{W} P(a \leq \alpha W \ell) < 1.$$

Taking the logarithm and using the inequalities (7) and

$$\binom{n}{\omega n} \lesssim 2^{nh(\omega)}$$

where the asymptotic equality holds for $n \to \infty$, we readily obtain

$$h(\omega) - \ell F(\alpha, \omega, n_0) < 0. \tag{8}$$

The largest value of ω which satisfies (8) for a given α is ω_α. Finally, we have from Lemmas 1 and 2 that for $\alpha > 2/3$, the algorithm $\mathscr{A}$ corrects $\omega_\alpha n$ errors with complexity $\mathcal{O}(n \log n)$, which completes the proof. ■

Theorem 1 allows us to compute ω_α numerically for several choices of code parameters. The computations confirm the existence of codes with a nonvanishing ω_α. We use $\alpha = 0.67$, which is slightly above the limit value of $2/3$. First, we consider code ensembles of a rate close to $1/2$. Figure 2 illustrates the values of ω_α computed for several code ensembles $\mathscr{C}(n_0, \ell, b)$ of rates approximately $1/2$. With increasing n_0 (and, in order to keep the rate fixed, also with increasing ℓ) the value of ω_α increases only up to a certain point, $n_0 = 128$, where it reaches its maximum. With further increase of n_0 and ℓ, ω_α decays quickly.

Next we consider code ensembles of different rates, but with a fixed constituent code. Figure 3 illustrates the values ω_α for RS-LDPC codes with the constituent $(128, 126, 3)$ Reed-Solomon code and with different code rates R, obtained by varying the choice of ℓ. We have found a nonvanishing ω_α for a wide range of code rates, and its value decreases with increasing code rate.

It is interesting to note that our construction compares favourably to the best known expander given in [7]. For example, for an RS-LDPC code ensemble of rate $R \approx 1/2$ with the constituent code length $n_0 = 128$ and $\ell = 32$ layers (*cf.* Figure 2), we obtain $\omega = 0.00105$, while the expander in [7] yields $\omega = 0.00014$. This improvement increases with increasing n_0 and ℓ.

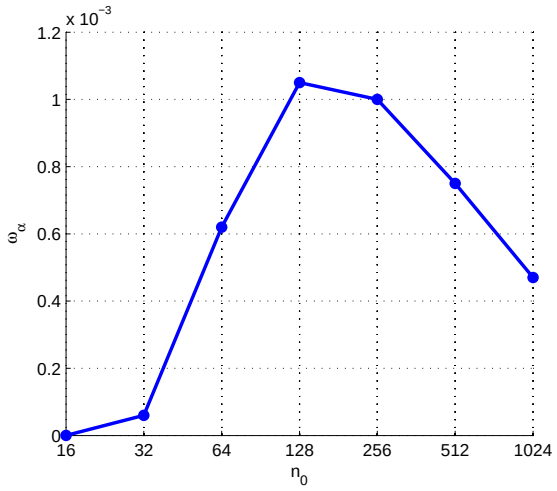

Fig. 2. Values of ω_α computed for $\alpha = 0.67$ according to Theorem 1 for several code ensembles of rates approximately $R \approx 1/2$. The maximum is achieved with the constituent code length $n_0 = 128$.

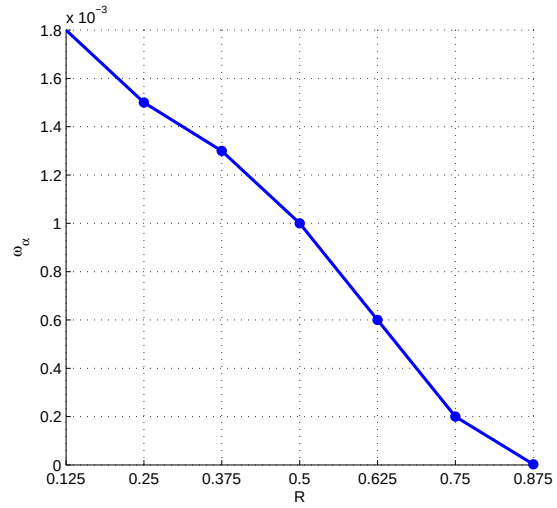

Fig. 3. Values of ω_α computed for $\alpha = 0.67$ according to Theorem 1 for several code ensembles of different rates with the fixed constituent code length $n_0 = 128$.

V. Conclusions

We have studied the performance of ensembles of Reed-Solomon code-based LDPC codes used over the BSC, when the code length n grows to infinity. It was shown that these codes can be decoded with a simple iterative decoding algorithm whose complexity is $\mathcal{O}(n \log n)$, and that there exist Reed-Solomon-LDPC codes which, when decoded with such an algorithm, are asymptotically capable of correcting a number of errors that grows linearly with the code length n. Such a property was previously proven to hold only for Gallager's LDPC codes and for the expander codes. The key property, which determines the code's error-correcting

capability, is how good an expander the underlying bipartite graph is.

The maximum fraction of errors ω, correctable with the iterative decoder, was computed numerically for two types of code ensembles, which are known to have minimum distances that asymptotically almost meet the Gilbert-Varshamov bound: codes of fixed rate $R \approx 1/2$ and codes of variable rates with a fixed constituent Reed-Solomon code.

REFERENCES

[1] M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533–547, Sept. 1981.

[2] R. G. Gallager, *Low-Density Parity-Check Codes*, Ph.D. thesis, MIT Press, Cambridge, MA, USA, 1963.

[3] I. E. Bocharova, B. D. Kudryashov, R. Johannesson, and V. V. Zyablov, "Woven graph codes over hyper graphs," in *Proc. 7th Int. ITG Conf. Source and Channel Coding*, Ulm, Germany, Jan. 2008.

[4] V. V. Zyablov and M. S. Pinsker, "Estimation of the error-correction complexity for Gallager low-density codes," *Problems of Inform. Transmission*, vol. 11, no. 1, pp. 23–26, Jan.–March 1975.

[5] A. Barg and G. Zémor, "Error exponents of expander codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1725–1729, June 2002.

[6] N. Miladinović and M. Fossorier, "Generalized LDPC codes with Reed-Solomon and BCH codes as component codes for binary channels," in *Proc. IEEE Global Comm. Conf. (GLOBECOM)*, St. Louis, MO, USA, Nov. 2005.

[7] L. A. Bassalygo, "Asymptotic optimal switching systems," *Problems of Inform. Transmission*, vol. 17, no. 3, pp. 81–88, Sept. 1981.