A Note on the Poor–Verdú Upper Bound for the Channel Reliability Function

Fady Alajaji, *Senior Member, IEEE*,
Po-Ning Chen, *Senior Member, IEEE*, and
Ziad Rached, *Student Member, IEEE*

***Abstract*—In an earlier work, Poor and Verdú established an upper bound for the reliability function of arbitrary single-user discrete-time channels with memory. They also conjectured that their bound is tight for all coding rates. In this note, we demonstrate via a counterexample involving memoryless binary erasure channels (BECs) that the Poor–Verdú upper bound is not tight at low rates. We conclude by examining possible improvements to this bound.**

***Index Terms*—Arbitrary channels with memory, binary erasure channels (BECs), channel coding, channel reliability function, information spectrum, probability of error.**

I. Introduction

Consider an arbitrary input $\boldsymbol{X}$ defined by a sequence of finite-dimensional distributions [9]

$$\boldsymbol{X} \triangleq \left\{X^n = \left(X_1^{(n)}, \ldots, X_n^{(n)}\right)\right\}_{n=1}^{\infty}.$$

Denote by

$$\boldsymbol{Y} \triangleq \left\{Y^n = \left(Y_1^{(n)}, \ldots, Y_n^{(n)}\right)\right\}_{n=1}^{\infty}$$

the corresponding output induced by $\boldsymbol{X}$ via the channel

$$\boldsymbol{W} \triangleq \{W^n = P_{Y^n|X^n}: \mathcal{X}^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$$

which is an arbitrary sequence of n-dimensional conditional distributions from $\mathcal{X}^n$ to $\mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively. We assume throughout that $\mathcal{X}$ is finite and that $\mathcal{Y}$ is arbitrary.

In [8], Poor and Verdú established an upper bound for the reliability function $E^*(R)$ of $\boldsymbol{W}$. They then conjectured that this bound is tight for all code rates. However, no known proof could substantiate this conjecture. In this work, we demonstrate via a counterexample that their original upper bound formula is not necessarily tight at low rates. A possible improvement to this bound is then addressed.

Previous related work mainly involved the establishment of upper and lower bounds for $E^*(R)$. In [1], [4], [6], and [10] (cf. also the references therein), the authors examined $E^*(R)$ for discrete memoryless channels (DMCs). More specifically, they presented three upper bounds (sphere packing, space partitioning, and straight line) and two lower bounds (random coding and expurgating) for $E^*(R)$ and employed them to show that $E^*(R)$ is convex and can be exactly determined via a simple expression at high rates (for R beyond some critical rate). The determination of $E^*(R)$ at low rates, which is conjectured to be convex, is still an unsolved problem, even for the simple memoryless binary symmetric channel (BSC). In [5], Egarmin extended the expressions of the random coding lower bound and the space partitioning upper bound for $E^*(R)$ for discrete finite-alphabet channels with modulo-additive irreducible Markov noise. He also proved that the two bounds coincide asymptotically (with the block length n) at high rates. In [7], Han derived an information-spectrum-based lower bound for $E^*(R)$ for arbitrary (not necessarily, stationary, ergodic, etc.) channels with memory. In addition to the general upper bound provided by Poor and Verdú, Chen *et al.* [3] derived another information-spectrum upper bound for $E^*(R)$ for arbitrary channels as a consequence to their result providing a general expression for the asymptotic largest minimum distance of block codes.

Manuscript received August 30, 2000; revised May 28, 2001. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the National Science Council of Taiwan, R.O.C., under Grant NSC 89-2213-E-009-237. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Washington DC, June 2001.

F. Alajaji and Z. Rached are with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: fady@mast.queensu.ca; rached@mast.queensu.ca).

P.-N. Chen is with the Department of Communication Engineering, National Chiao-Tung University, Hsin Chu, Taiwan, R.O.C. (e-mail: ponning@cc.nctu.edu.tw).

Communicated by P. Narayan, Associate Editor for Shannon Theory.
Publisher Item Identifier S 0018-9448(02)00036-6.

II. Preliminaries

Definition 1 (Channel Block Code): An (n, M) code for channel W^n with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is a pair of mappings

$$f: \{1, 2, \ldots, M\} \to \mathcal{X}^n \quad \text{and} \quad g: \mathcal{Y}^n \to \{1, 2, \ldots, M\}.$$

Its average error probability is given by

$$P_e(n, M) \triangleq \frac{1}{M}\sum_{m=1}^{M} \sum_{\{y^n:\, g(y^n)\neq m\}} W^n(y^n|f(m)).$$

Definition 2 (Channel Reliability Function [8]): For any $R > 0$, define the channel reliability function $E^*(R)$ for a channel $\boldsymbol{W}$ as the largest scalar $\beta > 0$ such that there exists a sequence of (n, M_n) codes with

$$\beta \le \liminf_{n\to\infty} -\frac{1}{N}\log_2 P_e(n, M_n)$$

and

$$R < \liminf_{n\to\infty} \frac{1}{n}\log_2 M_n. \tag{1}$$

With this definition, we next derive a slightly different but equivalent expression of the Poor–Verdú upper bound.

Definition 3: Fix $R > 0$. For an input $\boldsymbol{X}$ and a channel $\boldsymbol{W}$, the large-deviation spectrum of the channel is defined as

$$\pi_{\boldsymbol{X}}(R) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log_2 \Pr\left[\frac{1}{n} i_{X^nW^n}(X^n; Y^n) \le R\right]$$

where

$$i_{X^nW^n}(X^n; Y^n) = \log_2 \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)}$$

is the channel information density.

Theorem 1 (Poor–Verdú Upper Bound to $E^*(R)$ [8, eq. (14)]): The channel reliability function satisfies

$$E^*(R) \le \liminf_{n\to\infty} \sup_{X^n} -\frac{1}{n}\log_2 \Pr\left[\frac{1}{n} i_{X^nW^n}(X^n; Y^n) \le R\right] \tag{2}$$

$$= \sup_{\boldsymbol{X}} \pi_{\boldsymbol{X}}(R) \tag{3}$$

for $R > 0$.

Proof: Inequality (2) is actually given by [8, eq. (14)]; so we only need to prove equality (3). For any X^n

$$-\frac{1}{n}\log_2 \Pr\left[\frac{1}{n}\, i_{X^nW^n}(X^n;\, Y^n) \le R\right] \le \sup_{X^n} -\frac{1}{n}\log_2 \Pr\left[\frac{1}{n}\, i_{X^nW^n}(X^n;\, Y^n) \le R\right]$$

which implies that for any $\boldsymbol{X}$

$$\pi_{\boldsymbol{X}}(R) \le \liminf_{n\to\infty} \sup_{X^n} -\frac{1}{n}\log_2 \Pr\left[\frac{1}{n}\, i_{X^nW^n}(X^n; Y^n) \le R\right].$$

Accordingly, we have

$$\sup_{\boldsymbol{X}} \pi_{\boldsymbol{X}}(R) \le \liminf_{n\to\infty} \sup_{X^n} -\frac{1}{n}\log_2 \Pr\left[\frac{1}{n}\, i_{X^nW^n}(X^n; Y^n) \le R\right].$$

On the other hand, the finite alphabet assumption ensures the existence of $\hat{X}^n$ such that

$$-\frac{1}{n}\log_2 \Pr\left[\frac{1}{n}\, i_{\hat{X}^nW^n}\left(\hat{X}^n;\, \hat{Y}^n\right) \le R\right] = \sup_{X^n} -\frac{1}{n}\log_2 \Pr\left[\frac{1}{n}\, i_{X^nW^n}(X^n;\, Y^n) \le R\right]$$

where $\hat{Y}^n$ is the channel output due to channel input $\hat{X}^n$. Let $\hat{\boldsymbol{X}}$ be the triangular-array process having $\hat{X}^n$ as its n-dimensional marginal (for each n). Then

$$\sup_{\boldsymbol{X}} \pi_{\boldsymbol{X}}(R) \ge \pi_{\hat{\boldsymbol{X}}}(R) = \liminf_{n\to\infty} \sup_{X^n} -\frac{1}{n}\log_2 \Pr\left[\frac{1}{n}\, i_{X^nW^n}(X^n;\, Y^n) \le R\right].$$

□

In the previous theorem, the range of the supremum operation includes *all possible inputs*. However, it is straightforward from the proof of the Poor–Verdú upper bound (e.g., [8, eq. (14)]) that one can place both a *uniformity* restriction and the asymptotic condition of (1) on the input to yield a (possibly) better bound. This is illustrated in the next corollary.

Corollary 1: The channel reliability function satisfies

$$E^*(R) \le E_{\rm PV}(R) \triangleq \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \pi_{\boldsymbol{X}}(R)$$

for any $R > 0$, where

$$\mathcal{Q}(R) \triangleq \left\{\boldsymbol{X}\text{: Each } X^n \text{ in } \boldsymbol{X} \text{ is uniformly distributed over its support } \mathcal{S}(X^n), \text{ and } R < \liminf_{n\to\infty} \frac{1}{n}\log_2 |\mathcal{S}(X^n)|\right\}.$$

Remark: An observation that upholds the result of the above corollary is that for a channel $\boldsymbol{W}$ and any input $\boldsymbol{X}$ uniformly distributed over its support and satisfying

$$\liminf_{n\to\infty} \frac{1}{n}\log_2 |\mathcal{S}(X^n)| < R \tag{4}$$

the channel large deviation spectrum satisfies

$$\pi_{\boldsymbol{X}}(R) = 0.$$

This is justified as follows. Let $M_n \triangleq |\mathcal{S}(X^n)|$. We then observe that

$$\begin{aligned} i_{X^nW^n}(x^n;\, y^n) &= \log_2 \frac{P_{X^n,\,Y^n}(x^n,\, y^n)}{P_{X^n}(x^n)P_{Y^n}(y^n)} \\ &= \log_2 \frac{P_{X^n,\,Y^n}(x^n,\, y^n)}{\frac{1}{M_n}\sum_{x^n\in\mathcal{S}(X^n)} P_{X^n,\,Y^n}(x^n,\, y^n)} \\ &= \log_2 M_n + \log_2 \frac{P_{X^n,Y^n}(x^n, y^n)}{\sum_{x^n\in\mathcal{S}(X^n)} P_{X^n,Y^n}(x^n, y^n)} \\ &\le \log_2 M_n. \end{aligned}$$

Hence by (4)

$$\Pr\left\{\frac{1}{n}\, i_{X^nW^n}(X^n;\, Y^n) \le R\right\} \ge \Pr\left\{\frac{1}{n}\, i_{X^nW^n}(X^n;\, Y^n) \le \frac{1}{n}\log_2 M_n\right\} = 1$$

for n infinitely often, which immediately gives that $\pi_{\boldsymbol{X}}(R) = 0$. Consequently, when maximizing $\pi_{\boldsymbol{X}}(R)$ over all $\boldsymbol{X}$ that are uniformly distributed over their support, one only needs to consider those $\boldsymbol{X}$ violating (4); this justifies the upper bound formula in Corollary 1.

III. Looseness of $E_{\rm PV}(R)$ at Low Rates

In this section, we provide a counterexample in terms of a binary erasure channel (BEC) with crossover probability ε $(0 < \varepsilon < 1)$, which proves the looseness of $E_{\rm PV}(R)$ at low rates.

Denote by $\tilde{\boldsymbol{X}}$ the input to the BEC, where $\tilde{X}^n$ is uniformly distributed over $\{0,\, 1\}^n$. Then for any $s > 0$

$$\begin{aligned} &\Pr\left\{\frac{1}{n}\, i_{\tilde{X}^nW^n}\left(\tilde{X}^n;\, \tilde{Y}^n\right) \le R\right\} \\ &= \Pr\left\{2^{-s\cdot i_{\tilde{X}^nW^n}(\tilde{X}^n;\, \tilde{Y}^n)} \ge 2^{-n\cdot s\cdot R}\right\} \\ &\le 2^{n\cdot s\cdot R} E\left[2^{-s\cdot i_{\tilde{X}^nW^n}(\tilde{X}^n;\, \tilde{Y}^n)}\right] \quad \text{(Markov's inequality)} \\ &= 2^{n\cdot s\cdot R} E^n\left[2^{-s\cdot i_{\tilde{X}W}(\tilde{X};\, \tilde{Y})}\right] \end{aligned}$$

where $\tilde{Y}^n$ is the channel output due to the input $\tilde{X}^n$. This implies

$$\begin{aligned} E_{\rm PV}(R) &\triangleq \sup_{\boldsymbol{X}\in\mathcal{Q}(R)} \pi_{\boldsymbol{X}}(R) \\ &\ge \pi_{\tilde{\boldsymbol{X}}}(R) \\ &\ge \sup_{s>0}\left(-sR - \log_2 E\left[2^{-s\cdot i_{\tilde{X}W}(\tilde{X};\, \tilde{Y})}\right]\right) \\ &= \sup_{s>0}\left(-sR - \log_2\left[\varepsilon + (1-\varepsilon)2^{-s}\right]\right) \\ &= \begin{cases} R\log_2 \frac{R}{1-\varepsilon} + (1-R)\log_2 \frac{1-R}{\varepsilon}, & \text{for } 0 < R < 1-\varepsilon \\ 0, & \text{for } R \ge 1-\varepsilon. \end{cases} \end{aligned} \tag{5}$$

Observe that (5) is *exactly* the space partitioning upper bound $E_{\rm par}(R)$ for the BEC which is given by [1]

$$\begin{aligned} E_{\rm par}(R) &\triangleq \sup_{X}\sup_{s>0} \\ &\quad \times \left[-sR - \log_2 \sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)P_{Y|X}^{1/(1+s)}(y|x)\right)^{1+s}\right] \\ &= \sup_{s>0}\left[-sR - \log_2\left(\varepsilon + 2^{-s}(1-\varepsilon)\right)\right]. \end{aligned}$$

Hence, the conjectured tightness of $E_{\rm PV}(R)$ can be disproved via the looseness of the space partitioning upper bound. We then conclude from [1, Theorem 10.7.3] that

$$E_{\rm PV}(R) > E^*(R)$$

for $0 < R < 1 - \sqrt{\varepsilon}$ (see the Appendix).

Remarks:

- It can be shown by Cramer's theorem [2] that for DMCs

$$\pi_{\tilde{\boldsymbol{X}}}(R) = \sup_{s>0}\left(-sR - \log_2 E\left[2^{-s\cdot i_{\tilde{X}W}(\tilde{X};\,\tilde{Y})}\right]\right) \qquad (6)$$

  for a channel input $\tilde{\boldsymbol{X}}$ uniformly distributed over its entire space. The memoryless BEC, however, is indeed a peculiar channel for which the space-partitioning upper bound $E_{\rm par}(R)$ is actually equal[1] to (6). For example, in the case of the memoryless BSC, (6) is numerically observed to be strictly less than $E_{\rm par}(R)$ (and the straight-line bound); hence, the above simple technique that is used to disprove the tightness of $E_{\rm PV}(R)$ at low rates certainly does not apply for the BSC. Furthermore, since (6) is equal to $E_{\rm par}(R)$ for the BEC, we also cannot use this technique to disprove the tightness of $E_{\rm PV}(R)$ at high rates [since $E_{\rm par}(R)$ is tight at high rates].

- One may ask that the looseness of $E_{\rm PV}(R)$ at low rates may be due to the fact that in its formula, the range of the supremum operation includes all the inputs in $\mathcal{Q}(R)$, which may not be necessary. From the proofs of the Poor–Verdú upper bound in [8] and Theorem 1, we can further restrict the condition on the input to yield that for any $\rho > 0$

$$E^*(R) \le E_{PV}^{(\rho)}(R) \triangleq \sup_{\boldsymbol{X}\in\mathcal{P}(R,\,\rho)} \pi_{\boldsymbol{X}}(R)$$

  where

$$\mathcal{P}(R,\,\rho) \triangleq \left\{\boldsymbol{X}\colon X^n \text{ is uniform over its support } \mathcal{S}(X^n) \text{ and } R < \liminf_{n\to\infty}\frac{1}{n}\log_2|\mathcal{S}(X^n)| < R+\rho\right\}.$$

  Clearly, $E_{\rm PV}^{(\rho)}(R)$ is no greater than $E_{\rm PV}(R)$ since the former involves an additional restriction on the choice of $\boldsymbol{X}$. Also, note that the uniform input over $\{0,\,1\}^n$, which is used to disprove the tightness of $E_{\rm PV}(R)$, does not belong to $\mathcal{P}(R,\,\rho)$ for $0 < R \le 1-\rho$; hence, a possible improvement on $E_{\rm PV}(R)$ may be rendered from $E_{\rm PV}^{(\rho)}(R)$. We, however, can create another counterexample to show that $E_{\rm PV}^{(\rho)}(R)$ is still not tight at rates close to zero.

Claim: Consider a BEC with crossover probability ε, and fix $\rho > 0$. Then for $0 < R < 1/k$

$$E_{\rm PV}^{(\rho)}(R) \ge \sup_{s>0}\left\{-sR - \frac{1}{k}\log_2\left(\varepsilon^k + 2^{-s}\left[1-\varepsilon^k\right]\right)\right\}$$

where $k \triangleq \lceil 1/\rho\rceil$.

Proof: Let $\hat{\boldsymbol{X}}$ be block-wise independent with block size k; i.e.,

$$P_{\hat{X}^n}(x_1^n) = \left(\prod_{i=1}^{\omega} P_{\hat{X}^k}\left(x_{(i-1)k+1}^{ik}\right)\right)\times P_{\hat{X}^j}(x_{\omega k+1}^n)$$

where $\omega \triangleq \lfloor n/k\rfloor$, $j \triangleq n-\omega k$, $P_{\hat{X}^j}$ is the j-dimensional marginal of $P_{\hat{X}^k}$, and $P_{\hat{X}^k}$ is equally distributed over a set consisting of the all-zero sequence $\underline{0}$ and the all-one sequence $\underline{1}$ of dimension k ($P_{\hat{X}^k}(\underline{0}) = P_{\hat{X}^k}(\underline{1}) = 1/2$). Then

$$\liminf_{n\to\infty}\frac{1}{n}\log_2\left|\mathcal{S}\left(\hat{X}^n\right)\right| = \liminf_{n\to\infty}\frac{1}{n}\log_2 2^{\lceil n/k\rceil} = \frac{1}{k}$$

which implies that $\hat{\boldsymbol{X}} \in \mathcal{P}(R,\,\rho)$ for $0 < R < 1/k$, and

$$E_{\rm PV}^{(\rho)} \triangleq \sup_{\boldsymbol{X}\in\mathcal{P}(R,\rho)} \pi_{\boldsymbol{X}}(R) \ge \pi_{\hat{\boldsymbol{X}}}(R).$$

Observe that under $\hat{\boldsymbol{X}}$, the BEC (when the very last term is excluded) is transformed into a DMC with transition probability described by

$$P_{\hat{Y}^k|\hat{X}^k}\left(y^k\middle|\underline{0}\right) = (1-\varepsilon)^{v_0(y^k)}\varepsilon^{k-v_0(y^k)}\cdot\mathbf{1}\left\{v_1\left(y^k\right)=0\right\} \qquad (7)$$

and

$$P_{\hat{Y}^k|\hat{X}^k}\left(y^k\middle|\underline{1}\right) = (1-\varepsilon)^{v_1(y^k)}\varepsilon^{k-v_1(y^k)}\cdot\mathbf{1}\left\{v_0\left(y^k\right)=0\right\} \qquad (8)$$

where $v_0(y^k)$ and $v_1(y^k)$, respectively, represent the number of 0's and 1's in y^k, and $\mathbf{1}\{\cdot\}$ is the set indicator function. Since $P_{\hat{Y}^k|\hat{X}^k}$ only depends on v_0 and v_1, we can rewrite (7) and (8) as

$$P_{V^2|\hat{X}^k}(v_0,\,v_1|\underline{0}) = \binom{k}{v_0}\mu^{v_0}\varepsilon^k\cdot\mathbf{1}\{v_1=0\}$$

and

$$P_{V^2|\hat{X}^k}(v_0,\,v_1|\underline{1}) = \binom{k}{v_1}\mu^{v_1}\varepsilon^k\cdot\mathbf{1}\{v_0=0\}$$

where $\mu \triangleq (1-\varepsilon)/\varepsilon$. Therefore,

$$\begin{aligned}
&Pr\left\{\frac{1}{n}i_{\hat{X}^nW^n}\left(\hat{X}^n;\,\hat{Y}^n\right)\le R\right\}\\
&= \Pr\left\{\frac{1}{n}\sum_{i=1}^{\omega} i_{\hat{X}^kW^k}(\hat{X}_i^k;\,\hat{Y}_i^k) + \frac{1}{n}i_{\hat{X}^jW^j}\left(\hat{X}_{\omega+1}^j;\,\hat{Y}_{\omega+1}^j\right)\le R\right\}\\
&\le \Pr\left\{\frac{1}{(\omega+1)k}\sum_{i=1}^{\omega} i_{\hat{X}^kW^k}\left(\hat{X}_i^k;\,\hat{Y}_i^k\right) + \frac{1}{(\omega+1)k}i_{\hat{X}^jW^j}\left(\hat{X}_{\omega+1}^j;\,\hat{Y}_{\omega+1}^j\right)\le R\right\} \qquad (9)\\
&\le \Pr\left\{\frac{1}{(\omega+1)k}\sum_{i=1}^{\omega} i_{\hat{X}^kW^k}\left(\hat{X}_i^k;\,\hat{Y}_i^k\right)\le R\right\} \qquad (10)
\end{aligned}$$

where (9) holds since $(1/n)\ge 1/[(\omega+1)k]$, and (10) follows because

$$i_{\hat{X}^jW^j}\left(\hat{X}_{\omega+1}^j;\,\hat{Y}_{\omega+1}^j\right)\ge 0 \quad \text{with probability one.}$$

Accordingly,

$$\begin{aligned}
&-\frac{1}{n}\log_2\Pr\left\{\frac{1}{n}i_{\hat{X}^nW^n}\left(\hat{X}^n;\,\hat{Y}^n\right)\le R\right\}\\
&\ge \frac{1}{k}\left[-\frac{1}{\omega}\log_2\Pr\left\{\frac{1}{\omega+1}\sum_{i=1}^{\omega} i_{\hat{X}^kW^k}\left(\hat{X}_i^k;\,\hat{Y}_i^k\right)\le kR\right\}\right].
\end{aligned}$$

[1]This property actually holds for all memoryless q-ary ($q\ge 2$) erasure channels with input alphabet $\{0,1,\ldots,q-1\}$, output alphabet $\{0,1,\ldots,q-1,e\}$, and crossover probability ε. So the Poor–Verdú bound is also loose at low rates for this entire family of channels.

The proof is completed by noting that $\{i_{\hat{X}^k W^k}(\hat{X}_i^k; \hat{Y}_i^k)\}_{i=1}^{\omega}$ is independent and identically distributed (i.i.d.), and hence we can apply Cramer's theorem [2] to obtain

$$\begin{aligned}\pi_{\hat{X}}(R) \geq \sup_{s>0}\Bigg\{&-sR - \frac{1}{k}\log_2\Bigg(\sum_{x^k\in\mathcal{X}^k}\sum_{y^k\in\mathcal{Y}^k} P_{\hat{X}^k}\left(x^k\right) \\ &\times P^{1-s}_{\hat{Y}^k|\hat{X}^k}\left(y^k \middle| x^k\right) P^s_{\hat{Y}^k}\left(y^k\right)\Bigg)\Bigg\} \\
= \sup_{s>0}\Bigg\{&-sR - \frac{1}{k}\log_2\Bigg(\frac{1}{2}\sum_{y^k\in\mathcal{Y}^k} P^s_{\hat{Y}^k}\left(y^k\right) \\ &\times \left[P^{1-s}_{\hat{Y}^k|\hat{X}^k}\left(y^k \middle| \underline{0}\right) + P^{1-s}_{\hat{Y}^k|\hat{X}^k}\left(y^k \middle| \underline{1}\right)\right]\Bigg)\Bigg\} \\
= \sup_{s>0}\Bigg\{&-sR - \frac{1}{k}\log_2\Bigg(\frac{1}{2}\sum_{v_0=0}^{k}\sum_{v_1=0}^{k-v_0} \\ &\times \left[\frac{1}{2}\binom{k}{v_0}\mu^{v_0}\varepsilon^k \mathbf{1}(v_1=0) \right. \\ &\left. + \frac{1}{2}\binom{k}{v_1}\mu^{v_1}\varepsilon^k \mathbf{1}(v_0=0)\right]^s \\ &\times \left[\left(\binom{k}{v_0}\mu^{v_0}\varepsilon^k\mathbf{1}(v_1=0)\right)^{1-s} \right. \\ &\left. + \left(\binom{k}{v_1}\mu^{v_1}\varepsilon^k\mathbf{1}(v_0=0)\right)^{1-s}\right]\Bigg)\Bigg\} \\
= \sup_{s>0}\Bigg\{&-sR - \frac{1}{k}\log_2 \\ &\times \left(\frac{1}{2}\left[2\varepsilon^k + \sum_{v_1=1}^{k}\frac{1}{2^s}\binom{k}{v_1}\mu^{v_1}\varepsilon^k \right.\right. \\ &\left.\left. + \sum_{v_0=1}^{k}\frac{1}{2^s}\binom{k}{v_0}\mu^{v_0}\varepsilon^k\right]\right)\Bigg\} \\
= \sup_{s>0}\Bigg\{&-sR - \frac{1}{k}\log_2\left(\varepsilon^k + \frac{1}{2^s}\left[1-\varepsilon^k\right]\right)\Bigg\}. \qquad \square\end{aligned}$$

Based on the above claim, we can take $R \downarrow 0$ to obtain

$$\begin{aligned}\lim_{R\downarrow 0} E^{(\rho)}_{PV}(R) &\geq \lim_{R\downarrow 0}\sup_{s>0}\left\{-sR - \frac{1}{k}\log_2\left(\varepsilon^k + 2^{-s}[1-\varepsilon^k]\right)\right\} \\ &= -\log_2(\varepsilon),\end{aligned}$$

which is strictly greater than $\lim_{R\downarrow 0} E^*(R) = -\log_2(\varepsilon)/2$. Consequently, $E^{(\rho)}_{PV}(R)$ is not tight at rates close to zero.

- The previous remarks, together with the remark following Corollary 1, indicate that when bounding the reliability function of a channel by its large deviation spectrum, one should always consider the input whose normalized support size ultimately achieves the considered code rate. Any small deviation of the asymptotic normalized support size from the code rate could lead to a loose upper bound (at low rates). As a consequence, the best upper bound that can be readily obtained from the proofs of the Poor–Verdú upper bound in [8] and Theorem 1 is

$$E^*(R) \leq \inf_{\rho>0} E^{(\rho)}_{\mathrm{PV}}(R). \qquad (11)$$

Further investigation of the tightness of (11) at low rates for the BEC is an interesting future work.

Appendix

Lemma 1: For a BEC with crossover probability ε

$$E_{\mathrm{par}}(R) > E_{\mathrm{sl}}(R)$$

for $0 < R < 1 - \sqrt{\varepsilon}$, where $E_{\mathrm{sl}}(R)$ represents a straight-line upper bound for the channel reliability function.

Proof: First recall that for a BEC, the low-rate reliability function can be written as

$$\begin{aligned}E_L(R) &= \sup_{s\geq 0}\max_{P_X}\Bigg\{-sR - s\log_2\Bigg[\sum_{i,j\in\{0,1\}} P_X(i)P_X(j) \\ &\qquad \times \left(\sum_{k\in\{0,e,1\}}\sqrt{W(k|i)W(k|j)}\right)^{\frac{1}{s}}\Bigg]\Bigg\} \\ &= \sup_{s\geq 0}\left[-sR - s\log_2\left(\frac{1+\varepsilon^{1/s}}{2}\right)\right].\end{aligned}$$

In the limit as $R \to 0$, the sphere-packing upper bound $E_U(R)$ coincides with $E_L(R)$ [1, p. 410]. Hence,

$$E_U(0) = E_L(0) = \sup_{s\geq 0}\left[-s\log_2\left(\frac{1+\varepsilon^{1/s}}{2}\right)\right].$$

It is easy to check that $-s\log_2([1+\varepsilon^{1/s}]/2)$ is increasing in s. Therefore,

$$\begin{aligned}E_U(0) &= \lim_{s\to\infty}\left[-s\log_2\left(\frac{1+\varepsilon^{1/s}}{2}\right)\right] \\ &= \lim_{s\to\infty} -\frac{\log_2\left(\frac{1+\varepsilon^{1/s}}{2}\right)}{1/s} \\ &= -\log_2\sqrt{\varepsilon}\end{aligned}$$

where the last equality follows by l'Hôpital's rule [11].

Now [1, Theorem 10.7.3] indicates that any line segment between a point on the sphere-packing upper bound and a point on the spacing partitioning upper bound is an upper bound for the channel reliability. Construct the straight-line upper bound by taking the point $(0, -\log_2\sqrt{\varepsilon})$ from the sphere-packing upper bound and the point $(1-\sqrt{\varepsilon}, E_{\mathrm{par}}(1-\sqrt{\varepsilon}))$ from the space partitioning upper bound $E_{\mathrm{par}}(R)$. This straight-line upper bound should be of the form

$$\begin{aligned}E_{\mathrm{sl}}(R) &= RE'_{\mathrm{par}}(R_0) - \log_2\sqrt{\varepsilon} \\ &= R\log_2\left(\frac{R_0\varepsilon}{(1-R_0)(1-\varepsilon)}\right) - \log_2\sqrt{\varepsilon}\end{aligned}$$

for $0 < R < R_0$, where R_0 satisfies $E_{\mathrm{par}}(R_0) = E_{\mathrm{sl}}(R_0)$, which is exactly $R_0 = 1-\sqrt{\varepsilon}$. Taking $R_0 = 1-\sqrt{\varepsilon}$ into the above equation, yields

$$E_{\mathrm{sl}}(R) = R\cdot\log_2\frac{\sqrt{\varepsilon}}{1+\sqrt{\varepsilon}} - \log_2\sqrt{\varepsilon}, \qquad \text{for } 0 < R < 1-\sqrt{\varepsilon}.$$

The proof is then completed by noting that $E_{\rm par}(R)$ is strictly convex in its domain, and hence is larger than $E_{\rm sl}(R)$ for $0 < R < 1 - \sqrt{\varepsilon}$. □

References

[1] R. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison Wesley, 1988.

[2] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*. New York: Wiley, 1990.

[3] P.-N. Chen, T.-Y. Lee, and Y. Han, "Distance-spectrum formulas on the largest minimum distance of block codes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 869–885, May 2000.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[5] M. Egarmin, "Upper and lower bounds for the probability of error in coding for discrete channels," *Probl. Pered. Inform.*, vol. 5, no. 1, pp. 23–39, 1969.

[6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[7] T. S. Han, *Information-Spectrum Methods in Information Theory* (in Japanese). Tokyo, Japan: BaifuKan, 1998.

[8] H. V. Poor and S. Verdú, "A lower bound on the probability of error in multi-hypothesis testing," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1992–1994, Nov. 1995.

[9] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

[10] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[11] W. R. Wade, *An Introduction to Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

Strong Law of Large Numbers and Shannon–McMillan Theorem for Markov Chain Fields on Trees

Weiguo Yang and Wen Liu

***Abstract*—We study the strong law of large numbers and the Shannon–McMillan theorem for Markov chain fields on trees. First, we prove the strong law of large numbers for the frequencies of occurrence of states and ordered couples of states for Markov chain fields on trees. Then, we prove the Shannon–McMillan theorem with almost everywhere (a.e.) convergence for Markov chain fields on trees. We prove the results on a Bethe tree and then just state the analogous results on a rooted Cayley tree. In the proof, a new technique for establishing the strong limit theorem in probability theory is applied**

***Index Terms*—Bethe tree, Markov chain fields, random fields, rooted Cayley tree, Shannon–McMillan theorem, strong law of large numbers.**

I. Introduction

A tree is a graph $G = \{T, E\}$ which is connected and contains no circuits. Given any two vertices $\alpha \neq \beta \in T$, let $\overline{\alpha\beta}$ be the unique path connecting α and β. Define the graph distance $d(\alpha, \beta)$ to be the number of edges contained in the path $\overline{\alpha\beta}$.

Manuscript received Nov. 1, 1999; revised April 21, 2001.

W. Yang is with the Jiangsu University of Science and Technology, Zhenjiang 212013, China (e-mail: yangweiguo1@263.net).

W. Liu is with the Hebei University of Technology, Tianjin 300130, China.

Communicated by I. Csiszár, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(02)00060-3.

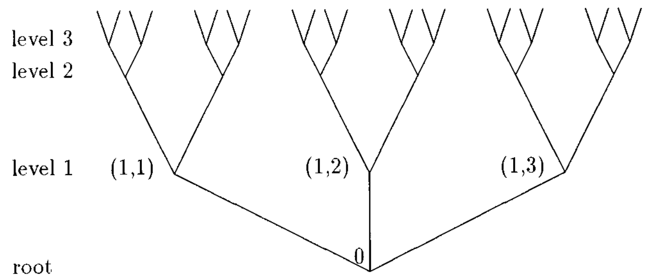

Fig. 1. Bethe tree $T_{B,2}$.

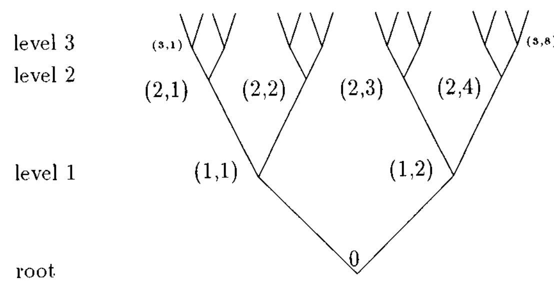

Fig. 2. Cayley tree $T_{C,2}$ (i.e., binary tree).

We discuss mainly a Bethe tree $T_{B,N}$ on which each vertex has $N+1$ neighboring vertices. For simplicity, we investigate only $T_{B,2}$ (see Fig. 1) in this correspondence.

To index the vertices on $T_{B,2}$, we first fix any one vertex as the "root" and label it by 0. A vertex is said to be on the nth level if the path linking it to the root has n edges.

We also discuss a rooted Cayley tree $T_{C,2}$ (i.e., a binary tree, see Fig. 2). In a Cayley tree $T_{C,2}$, the root has only two neighbors and all other vertices have three neighbors just as in $T_{B,2}$. When the context permits, $T_{B,2}$ and $T_{C,2}$ are all denoted simply by T.

We denote by L_n^m the subgraph of T containing the vertices from nth level to the mth level. In particular, $T^{(n)} \triangleq L_0^n$ is the subtree of T containing the vertices from level 0 (the root) to level n.

We use (n, j) to denote the jth vertex at the nth level. Thus, (n, j) has neighbors $(n+1, 2j-1)$, $(n+1, 2j)$ and $(n-1, [j/2])$, where $[c]$ is the smallest integer not less than c.

We denote by $|B|$ the number of vertices in subgraph B. It is easy to see that if T is a Bethe tree $T_{B,2}$

$$\left|T^{(n)}\right| = 3 \cdot 2^n - 2 \tag{1}$$

if T is a Cayley tree $T_{C,2}$

$$\left|T^{(n)}\right| = 2^{n+1} - 1. \tag{2}$$

Let $\Omega = \{0, 1\}^T$, $\mathcal{F}$ be the smallest Borel field containing all cylinder sets in Ω. Let $X = \{X_t, t \in T\}$ be the stochastic process defined on the measurable space $(\Omega, \mathcal{F})$, that is, for any $\omega = \{\omega(t), t \in T\}$, define

$$X_t(\omega) = \omega(t), \qquad t \in T. \tag{3}$$

Let μ be a probability measure on the measurable space $(\Omega, \mathcal{F})$. We will call μ a random field on tree T.

Definition 1 (see [5]): Let μ be a probability measure on the measurable space $(\Omega, \mathcal{F})$. If

$$\mu(\omega(j)|\omega(k), k \in T - \{j\}) = \mu(\omega(j)|\omega(k), k \in N(j)) \tag{4}$$