

Error Exponents of Erasure/List Decoding Revisited via Analysis of Distance Enumerators

Neri Merhav

Department of Electrical Engineering
Technion—Israel Institute of Technology
Haifa 32000, Israel

Partly joint work with **Anelia Somekh–Baruch**, EE Department, Princeton University

The 2008 Information Theory & Applications (ITA) Workshop
UCSD, San Diego, CA, January–February 2008

Background

In 1968 Forney studied the problem generalized decoding rules where the decision regions may have:

overlaps – list decoding and/or **holes** – erasure decoding

Consider a DMC $P(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and an n -block code $\mathcal{C} = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$, $M = e^{nR}$. In the case of erasure decoding, a decision rule is a partition of \mathcal{Y}^n into $(M + 1)$ regions where

$\mathbf{y} \in \mathcal{R}_0$ means: erase

and

$\mathbf{y} \in \mathcal{R}_m$ ($m \geq 1$) means: decide that \mathbf{x}_m was transmitted.

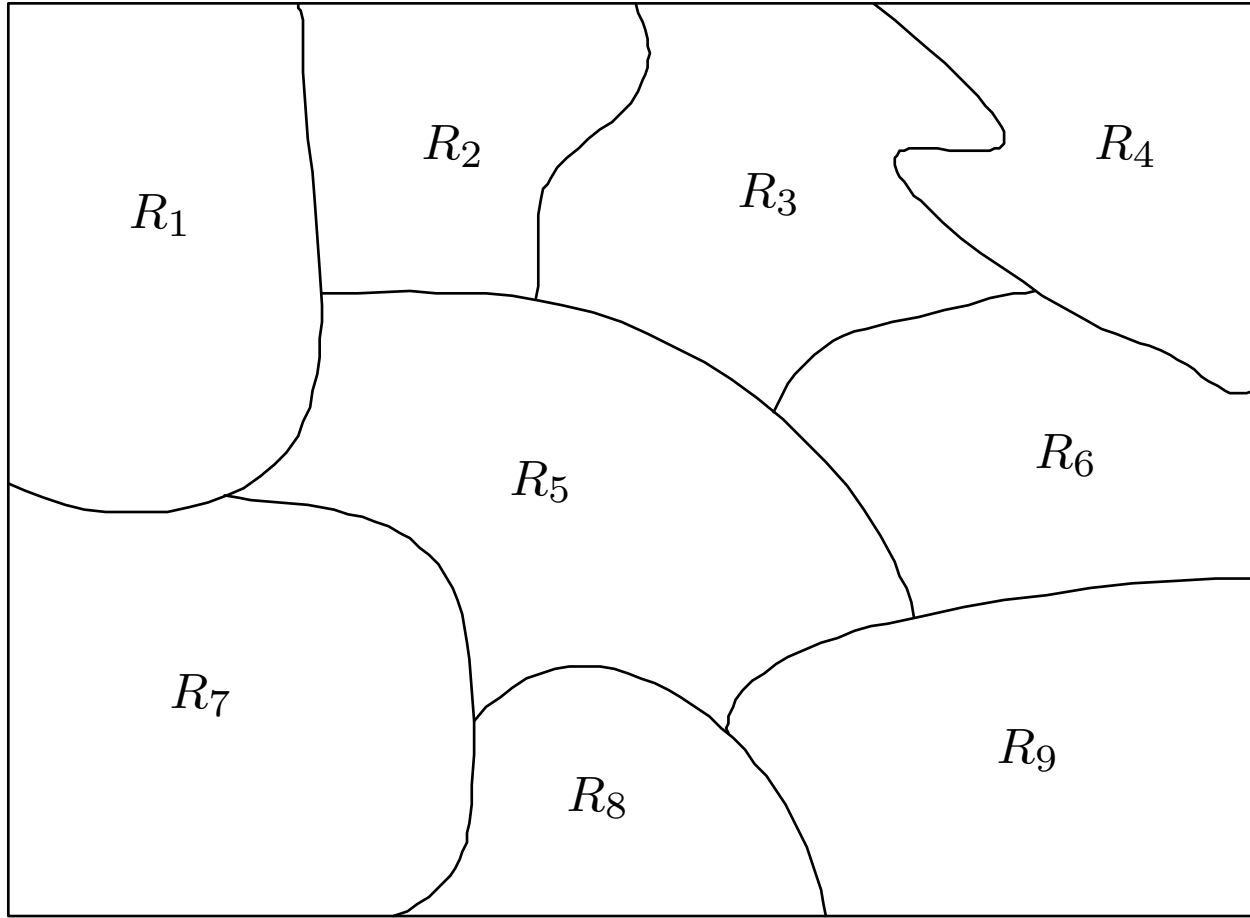


Figure 1: Standard decoding

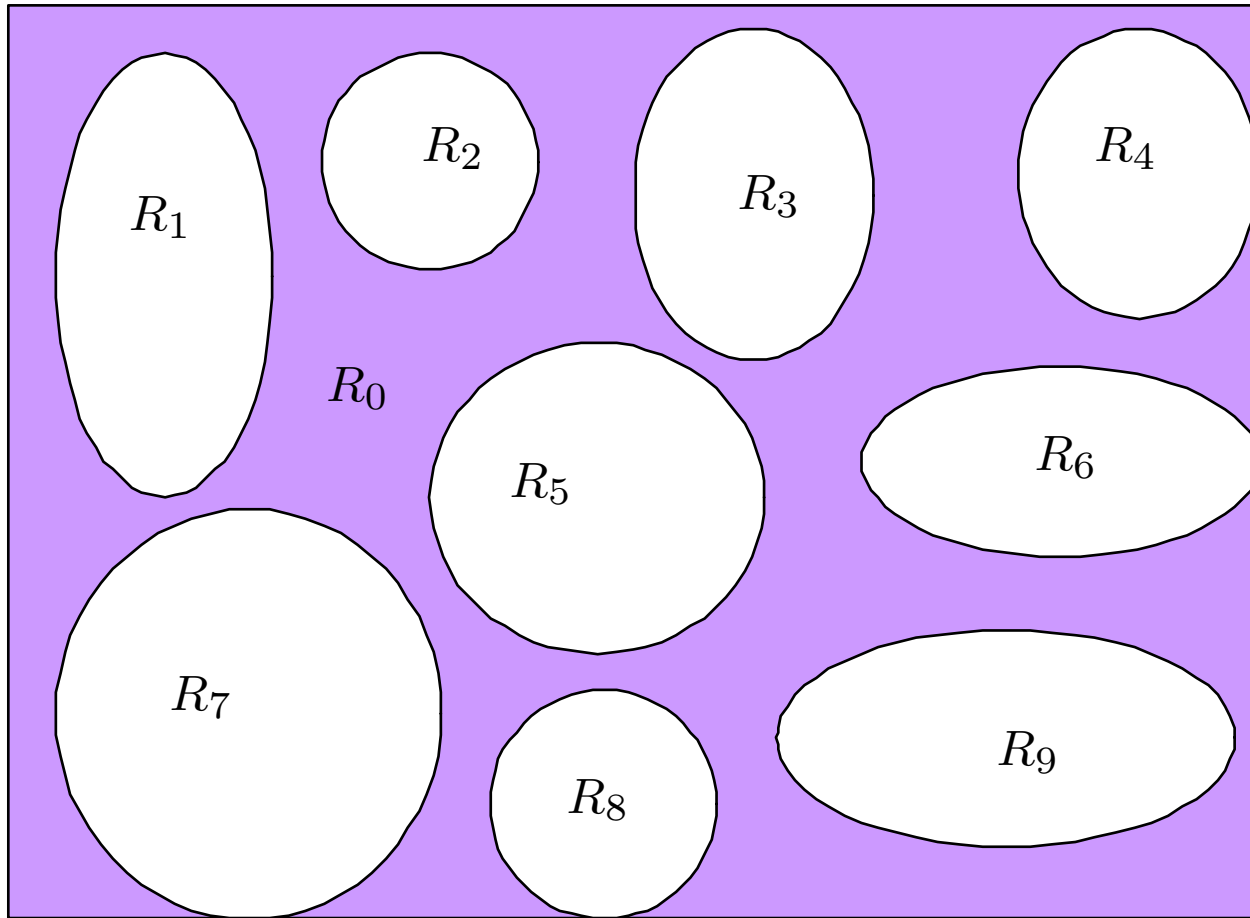


Figure 2: Decoding with an erasure option

Background (Cont'd)

Performance is judged according to the tradeoff between two criteria:

$$\Pr\{\mathcal{E}_1\} = \frac{1}{M} \sum_m \sum_{\mathbf{y} \in \mathcal{R}_m^c} P(\mathbf{y}|\mathbf{x}_m) \quad \text{erasure + undetected error}$$

and

$$\Pr\{\mathcal{E}_2\} = \frac{1}{M} \sum_m \sum_{\mathbf{y} \in \mathcal{R}_m} \sum_{m' \neq m} P(\mathbf{y}|\mathbf{x}_{m'}) \quad \text{undetected error}$$

The optimum decoder decides in favor of message m ($\mathbf{y} \in \mathcal{R}_m$) iff

$$\frac{P(\mathbf{y}|\mathbf{x}_m)}{\sum_{m' \neq m} P(\mathbf{y}|\mathbf{x}_{m'})} \geq e^{nT} \quad (T \geq 0 \text{ for the erasure case}).$$

Erasure is declared if this holds for no message m .

Background (Cont'd)

Forney's **lower bounds** to the random coding error exponents of \mathcal{E}_1 and \mathcal{E}_2 :

$$E_1(R, T) = \max_{0 \leq s \leq \rho \leq 1} [E_0(s, \rho) - \rho R - sT]$$

where

$$E_0(s, \rho) = -\ln \left[\sum_y \left(\sum_x P(x) P^{1-s}(y|x) \right) \cdot \left(\sum_{x'} P(x') P^{s/\rho}(y|x') \right)^\rho \right],$$

$$E_2(R, T) = E_1(R, T) + T.$$

and $\{P(x), x \in \mathcal{X}\}$ is the i.i.d. random coding distribution.

Main Result

Our main result is in the following **exact** single-letter expressions for the random coding exponents of the probabilities of \mathcal{E}_1 and \mathcal{E}_2 :

$$E_1(R, T) = \min_{P'_{XY}} \left[D(P'_{XY} \| P_{XY}) + \min_{Q_{XY}} \left\{ \mathbf{E}_Q \ln \frac{1}{P(X)} - H_Q(X|Y) \right\} - R \right]$$

where the inner minimization is subject to the constraints:

$$R + H_Q(X|Y) + \mathbf{E}_Q \ln P(X) \leq 0$$

$$\mathbf{E}_{P'} \ln P(Y|X) - \mathbf{E}_Q \ln P(Y|X) \leq T$$

$$Q_Y = P'_Y.$$

Main Result (Cont'd)

$$E_2(R, T) = \min_{P'_Y} \left[D(P'_Y \| P_Y) + \min_{\theta \leq \theta_0(P'_Y)} \{E_A(P'_Y, \theta) + E_B(P'_Y, \theta)\} - R \right]$$

where

$$E_A(P'_Y, \theta) = \min_{Q_{XY}: \mathbf{E}_Q \ln P(Y|X) = T - \theta, Q_Y = P'_Y} \left[\mathbf{E}_Q \ln \frac{1}{P(X)} - H_Q(X|Y) \right]$$

$$E_B(P'_Y, \theta) = \min_{Q_{XY}: \mathbf{E}_Q \ln P(Y|X) \leq -\theta, Q_Y = P'_Y} \left[\mathbf{E}_Q \ln \frac{1}{P(X|Y)} - H_Q(X|Y) \right]$$

and

$$\theta_0(P'_Y) = \min_Q \left[\mathbf{E}_Q \ln \frac{1}{P(X, Y)} - H_Q(X|Y) \right] - R$$

subject to

$$R + H_Q(X|Y) + \mathbf{E}_Q \ln P(X) \geq 0.$$

Main Ideas of the Analysis Technique – BSC Case

Fix \mathbf{x}_0 and \mathbf{y} . Then, defining $\beta = \ln \frac{1-p}{p}$:

$$\begin{aligned}\Pr\{\mathcal{E}_1\} &= \Pr \left\{ \sum_{m>0} P(\mathbf{y}|\mathbf{X}_m) > e^{-nT} P(\mathbf{y}|\mathbf{x}_0) \right\} \\ &= \Pr \left\{ \sum_{\delta} N_{\mathbf{y}}(n\delta) \cdot e^{-\beta n\delta} > e^{-nT} \cdot e^{-\beta n\delta_0} \right\} \\ &\doteq \Pr \left\{ \max_{\delta} \left[N_{\mathbf{y}}(n\delta) \cdot e^{-\beta n\delta} \right] > e^{-nT} \cdot e^{-\beta n\delta_0} \right\} \\ &= \Pr \bigcup_{\delta} \left\{ N_{\mathbf{y}}(n\delta) > e^{n[\beta(\delta-\delta_0)-T]} \right\} \\ &\doteq \max_{\delta} \Pr \left\{ N_{\mathbf{y}}(n\delta) > e^{n[\beta(\delta-\delta_0)-T]} \right\}\end{aligned}$$

So, it is all about the large deviations behavior of $\{N_{\mathbf{y}}(n\delta)\}$.

Main Ideas of the Analysis (Cont'd)

Now, $N_{\mathbf{y}}(n\delta) = \sum_{m=1}^{M-1} 1 \{d(\mathbf{X}_m, \mathbf{y}) = n\delta\}$. Thus, there are $M - 1 \approx e^{nR}$ trials, each with probability of 'success' $q \doteq e^{-n[\ln 2 - h(\delta)]}$.

- If $M \gg 1/q$, i.e., $R > \ln 2 - h(\delta)$, then $N_{\mathbf{y}}(n\delta) \doteq e^{n[R+h(\delta)-\ln 2]}$ with **extremely** high probability – **double–exponentially** close to 1.
- If $M \ll 1/q$, i.e., $R < \ln 2 - h(\delta)$, then typically, $N_{\mathbf{y}}(n\delta) = 0$, and $N_{\mathbf{y}}(n\delta) \geq 1$ with probability $e^{-n[\ln 2 - h(\delta)]}$.

$$\Pr \{N_{\mathbf{y}}(n\delta) \text{ is exponential in } n\}$$

decays **double–exponentially**.

This observation is the basis of understanding phase transition phenomena in analogous probabilistic models of spin glasses, like the REM [Mézard & Montanari '07].

The Case $R > \ln 2 - h(\delta)$

Or, equivalently, $\delta_{GV}(R) \leq \delta \leq 1 - \delta_{GV}(R)$.

In this range, $N_{\mathbf{y}}(n\delta) \doteq e^{n[R+h(\delta)-\ln 2]}$ w.h.p., so

$$\max_{\delta \in [\delta_{GV}(R), 1 - \delta_{GV}(R)]} \Pr \left\{ N_{\mathbf{y}}(n\delta) > e^{n[\beta(\delta - \delta_0) - T]} \right\}$$

is **very** close to 1 for all δ_0 such that

$$R + h(\delta) - \ln 2 > \beta(\delta - \delta_0) - T \text{ for some } \delta \in [\delta_{GV}(R), 1 - \delta_{GV}(R)],$$

or equivalently,

$$\max_{\delta \in [\delta_{GV}(R), 1 - \delta_{GV}(R)]} [h(\delta) - \beta\delta] + R - \ln 2 > -\beta\delta_0 - T.$$

But the maximizer is $\delta^* = \delta_{GV}(R)$ and so,

$$-\beta\delta_{GV}(R) > -\beta\delta_0 - T \implies \delta_0 > \delta_{GV}(R) - T/\beta.$$

The probability of this event is $\doteq e^{-nD(\delta_{GV}(R) - T/\beta \| p)}$.

The Case $R < \ln 2 - h(\delta)$

Or, equivalently, $\delta < \delta_{GV}(R)$ or $\delta > 1 - \delta_{GV}(R)$.

Given δ_0 , every δ such that $\beta(\delta - \delta_0) - T > 0$ yields a doubly-exponentially small contribution to

$$\Pr \left\{ N_{\mathbf{y}}(n\delta) > e^{n[\beta(\delta - \delta_0) - T]} \right\}$$

and for every δ such that $\beta(\delta - \delta_0) - T \leq 0$ (or, equivalently, $\delta < \delta_0 + T/\beta$) contributes

$$\Pr \left\{ N_{\mathbf{y}}(n\delta) > e^{n[\beta(\delta - \delta_0) - T]} \right\} = \Pr \left\{ N_{\mathbf{y}}(n\delta) \geq 1 \right\} \doteq e^{-n[\ln 2 - h(\delta) - R]}.$$

Thus, the dominant contribution is given by the largest δ , that is

$$\exp\{-n[\ln 2 - h(\min\{\delta_{GV}(R), \delta_0 + T/\beta\}) - R]\}.$$

This should be weighed by the probability of δ_0 , which is $\doteq e^{-nD(\delta_0||p)}$ and summed over all δ_0 .

Finally, the dominant contribution of the two ranges dictates the exponent.

A Simple Lower Bound \geq Forney's Bound

Forney's bound is based on bounding the error indicator function by the likelihood ratio, where the main obstacle is in handling the expression

$$\mathbf{E} \left\{ \left(\sum_{m' \neq m} P(\mathbf{y} | \mathbf{X}_{m'}) \right)^s \right\}$$

which is upper bounded by

$$\mathbf{E} \left\{ \left(\sum_{m' \neq m} P(\mathbf{y} | \mathbf{X}_{m'})^{s/\rho} \right)^\rho \right\}, \quad \rho \geq s,$$

and then Jensen's inequality is applied.

Here we compute the former expression exponentially tightly and avoid the need for the additional parameter ρ .

A Simple Lower Bound (Cont'd)

For the case of the BSC:

$$\begin{aligned} \mathbf{E} \left\{ \left(\sum_{m' \neq m} P(\mathbf{y} | \mathbf{X}_{m'}) \right)^s \right\} &= \mathbf{E} \left\{ \left[(1-p)^n \sum_{d=0}^n N_{\mathbf{y}}(d) e^{-\beta d} \right]^s \right\} \\ &\doteq (1-p)^{ns} \sum_{d=0}^n \mathbf{E}\{N_{\mathbf{y}}^s(d)\} e^{-\beta sd}. \end{aligned}$$

Now, the moments $\mathbf{E}\{N_{\mathbf{y}}^s(d)\}$ behave as follows:

$$\mathbf{E}\{N_{\mathbf{y}}^s(n\delta)\} \doteq \begin{cases} e^{ns[R+h(\delta)-\ln 2]} & \delta_{GV}(R) < \delta < 1 - \delta_{GV}(R) \\ e^{n[R+h(\delta)-\ln 2]} & \delta \leq \delta_{GV}(R) \text{ or } \delta \geq 1 - \delta_{GV}(R) \end{cases}$$

The More General Lower Bound

Assume that the random coding distribution $\{P(x), x \in \mathcal{X}\}$ and the channel transition matrix $\{P(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ are such that for every real s ,

$$\gamma_y(s) \triangleq -\ln \left[\sum_{x \in \mathcal{X}} P(x) P^s(y|x) \right]$$

is **independent** of y , in which case, it will be denoted by $\gamma(s)$. Let s_R be the solution to the equation $\gamma(s) - s\gamma'(s) = R$, where $\gamma'(s)$ is the derivative of $\gamma(s)$.

Then,

$$E_1^*(R, T) = \max_{s \geq 0} [\Lambda(R, s) + \gamma(1-s) - sT] - \ln |\mathcal{Y}|$$

where

$$\Lambda(R, s) = \begin{cases} \gamma(s) - R & s \geq s_R \\ s\gamma'(s_R) & s < s_R \end{cases}$$

Concluding Discussion

- We offer:
 - an **exact** but complicated expression for a general DMC and random coding distribution, and
 - A **simple** lower bound that holds under the symmetry condition $\forall s, \sum_{x \in \mathcal{X}} P(x) P^s(y|x)$ is independent of y .
- This condition holds when the columns of the matrix $\{P(y|x)\}$ are permutations of each other and $\{P(x)\}$ is uniform. E.g., modulo-additive channels. More generally, different columns of $\{P(y|x)\}$ are formed under the rule: $P(y|x)$ can be exchanged with $P(y|x')$ if $P(x) = P(x')$.
- Without this condition, we are losing the simplicity – better to adopt the exact expression.
- For certain channels, like the BSC, the optimum s can be found in closed form.
- We have **not** found any numerical example that strictly beats Forney's bound. In all cases we examined, the two bounds and the exact expression all coincide. This further supports the conjecture that Forney's bound is tight for the average code.