Error Correcting Codes: Combinatorics, Algorithms and Applications **(Fall 2007)**

Lecture 10: Shannon's Theorem

September 19, 2007

Lecturer: Atri Rudra *Scribe: Atri Rudra & Michael Pfetsch*

In the last lecture, we proved part (2) of Shannon's capacity theorem for the binary symmetric channel (BSC), which we restate here (throughout these notes we will use "$\mathbf{e} \sim BSC_p$" as a shorthand for "noise $\mathbf{e}$ from BSC_p"):

Theorem 0.1. *Let $0 \leq p < \frac{1}{2}$ be a real number. For every $0 < \varepsilon \leq \frac{1}{2} - p$, the following statements are true for large enough integer n:*

1. *There exists a real $\delta > 0$, and encoding function $E : \{0,1\}^k \rightarrow \{0,1\}^n$, and a decoding function $D : \{0,1\}^n \rightarrow \{0,1\}^k$, where $k \leq \lfloor (1 - H(p+\varepsilon))\, n \rfloor$ such that the following holds for every $\mathbf{m} \in \{0,1\}^k$:*

$$\Pr_{\mathbf{e} \sim BSC_p} [D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \leq 2^{-\delta n}$$

2. *If $k \geq \lceil (1 - H(p) + \varepsilon)\, n \rceil$ then for every encoding function $E : \{0,1\} \rightarrow \{0,1\}^n$ and decoding function $D : \{0,1\}^n \rightarrow \{0,1\}^k$ the following is true for some $\mathbf{m} \in \{0,1\}^k$:*

$$\Pr_{\mathbf{e} \sim BSC_p} [D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \geq \frac{1}{2}$$

In today's lecture, we will prove part (1) of Theorem 0.1

1 Proof overview

The proof of part (1) of Theorem 0.1 will be accomplished by randomly selecting an encoding function $E : \{0,1\}^k \rightarrow \{0,1\}^n$. That is, for every $\mathbf{m} \in \{0,1\}^k$ pick $E(\mathbf{m})$ uniformly and independently at random from $\{0,1\}^n$. D will be the maximum likelihood decoding (MLD) function. The proof will have the following two steps:

1. Step 1: For any arbitrary $\mathbf{m} \in \{0,1\}^k$, we will show that for a random choice of E, the probability of failure, over BSC_p noise, is small. This implies the existence of a good encoding function for any arbitrary message.

2. Step 2: We will show a similar result for all $\mathbf{m}$. This involves dropping half of the code words.

The proof method above has its own name– "random coding with expurgation."
Note that there are two sources of randomness in the proof:

1. Randomness in the choice of encoding function E and
2. Randomness in the noise.

We stress that the first kind of randomness is for the probabilistic method while the second kind of randomness will contribute to the decoding error probability.

2 "Proof by picture" of Step 1

Before proving part (1) of Theorem 0.1, we will provide a pictorial proof of Step 1. We begin by fixing $\mathbf{m} \in \{0,1\}^k$. In Step 1, we need to estimate the following quantity:

$$\mathbb{E}_E \left[\underset{\mathbf{e} \sim BSC_p}{Pr} \left[D\left(E\left(\mathbf{m}\right) + \mathbf{e}\right) \neq \mathbf{m} \right] \right].$$

By the Chernoff bound, with all but an exponentially small probability, the received word will be contained in a Hamming ball of radius $(p+\varepsilon')\,n$ (for some $\varepsilon' > 0$ that we will choose appropriately). So one can assume that the received word $\mathbf{y}$ whp satisfies $\Delta(E(\mathbf{m}), \mathbf{y}) \leq (p + \varepsilon')n$. Given this, pretty much the only thing to do is to estimate the decoding error probability for such a $\mathbf{y}$. Note that by the fact that D is MLD, an error can happen only if there exists another message $\mathbf{m}'$ such that $\Delta(E(\mathbf{m}'), \mathbf{y}) \leq \Delta(E(\mathbf{m}), \mathbf{y})$. The latter event is implied by the event that $\Delta(E(\mathbf{m}'), \mathbf{y}) \leq (p + \varepsilon')n$ (see Figure 2). Thus, the decoding error probability is upper bounded by

$$\underset{\mathbf{e} \sim BSC_p}{Pr} \left[E\left(\mathbf{m}'\right) \in B_2\left(\mathbf{y}, \left(p+\varepsilon'\right)n\right) \right] = \frac{\left| B_2\left(E\left(\mathbf{m}'\right), \left(p+\varepsilon'\right)n\right) \right|}{2^n} \approx \frac{2^{H(p)n}}{2^n}.$$

Finally, by the union bound, the existence of such a "bad" $\mathbf{m}'$ is upper bounded by $\approx \frac{2^k 2^{nH(p)}}{2^n}$, which by our choice of k is $2^{-\Omega(n)}$, as desired.

2.1 A Digression: An Additive Chernoff Bound

We have seen a "multiplicative" form of the Chernoff bound. Today we will use the following "additive" form of the Chernoff bound. Let $X_1, \ldots, X_n$ be independent $0-1$ random variables which take a value of 1 with probability p. Then for small enough $\gamma > 0$, $Pr[\sum_i X_i \geq (p+\gamma)n] \leq e^{-\gamma^2 n/2}$.

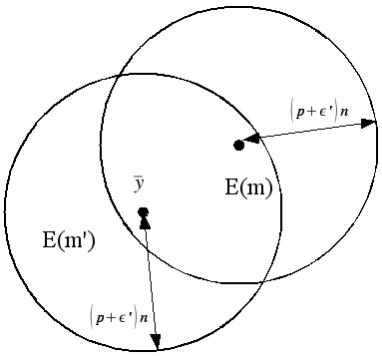

Figure 1: Hamming balls of radius $(p+\varepsilon')\,n$ and centers $E\,(\mathbf{m})$ and $\mathbf{y}$ illustrates Step 1 in the proof of part (2) of Shannon's capacity theorem for the BSC.

3 Proof of first part of Shannon's Theorem

For notational convenience, we will use $\mathbf{y}$ and $E\,(\mathbf{m})+\mathbf{e}$ interchangeably:

$$\mathbf{y} = E\,(\mathbf{m})+\mathbf{e}.$$

That is, $\mathbf{y}$ is the received word when $E\,(\mathbf{m})$ is transmitted and $\mathbf{e}$ is the error pattern.

We start the proof by restating the decoding error probability in part (1) of Shannon's capacity theorem for the BSC (Theorem 0.1) by breaking up the quantity in two two sums:

$$\begin{aligned}\underset{\mathbf{e}\sim BSC_p}{Pr}\left[D\left(E\left(\mathbf{m}\right)+\mathbf{e}\right)\neq\mathbf{m}\right] = &\sum_{\mathbf{y}\in B(E(\mathbf{m}),(p+\varepsilon')n)} Pr[\mathbf{y}|E(\mathbf{m})]\cdot\mathbf{l}_{D(\mathbf{y})\neq\mathbf{m}}\\ &+\sum_{\mathbf{y}\notin B(E(\mathbf{m}),(p+\varepsilon')n)} Pr[\mathbf{y}|E(\mathbf{m})]\cdot\mathbf{l}_{D(\mathbf{y})\neq\mathbf{m}},\end{aligned}$$

where $\mathbf{1}_{D(\mathbf{y})\neq\mathbf{m}}$ is the indicator function for the event that $D(\mathbf{y})\neq\mathbf{m}$ *given* that $E(\mathbf{m})$ was the transmitted codeword and we use $\mathbf{y}|E(\mathbf{m})$ as a shorthand for "$\mathbf{y}$ is the received word given that $E(\mathbf{m})$ was the transmitted codeword." As $\mathbf{1}_{D(\mathbf{y})\neq\mathbf{m}}\leq 1$ (since it takes a value in $\{0,1\}$) and by the (additive) Chernoff bound we have

$$\underset{\mathbf{e}\sim BSC_p}{Pr}\left[D\left(E\left(\mathbf{m}\right)+\mathbf{e}\right)\neq\mathbf{m}\right]\leq\sum_{\mathbf{y}\in B(E(\mathbf{m}),(p+\varepsilon')n)} Pr[\mathbf{y}|E(\mathbf{m})]\cdot\mathbf{l}_{D(\mathbf{y})\neq\mathbf{m}}+e^{-(\varepsilon')^2n/2}.$$

In order to apply the probabilistic method, we will analyze the expectation (over the random choice of E) of the decoding error probability, which by the upper bound above satisfies

$$\mathbb{E}_E\left[\underset{\mathbf{e}\sim BSC_p}{Pr}\left[D\left(E\left(\mathbf{m}\right)+\mathbf{e}\right)\neq\mathbf{m}\right]\right]\leq e^{-\varepsilon'^2n/2}+\sum_{\mathbf{y}\in B(E(\mathbf{m}),(p+\varepsilon')n)}\underset{\mathbf{e}\sim BSC_p}{Pr}[\mathbf{y}|E(\mathbf{m})]\cdot\mathbb{E}_E\left[\mathbf{1}_{D(\mathbf{y})\neq\mathbf{m}}\right]. \tag{1}$$

Next, for a fixed received word $\mathbf{y}$ and the transmitted codeword $E(\mathbf{m})$ such that $\Delta(\mathbf{y}, E(\mathbf{m})) \leq (p+\varepsilon')n$ we estimate $\mathbb{E}_E[\mathbf{1}_{D(\mathbf{y})\neq\mathbf{m}}]$. Since D is MLD, we have

$$\mathbb{E}_E\left[\mathbf{1}_{D(\mathbf{y})\neq\mathbf{m}}\right] = Pr_E\left[\mathbf{1}_{D(\mathbf{y})\neq\mathbf{m}}|E(\mathbf{m})\right] \leq \sum_{\mathbf{m}'\neq\mathbf{m}} Pr\left[\Delta\left(E\left(\mathbf{m}'\right),\mathbf{y}\right) \leq \Delta\left(E\left(\mathbf{m}\right),\mathbf{y}\right)|E(\mathbf{m})\right], \tag{2}$$

where in the above "$|E(\mathbf{m})$" is short for "being conditioned on $E(\mathbf{m})$ being transmitted" and the inequality follows from the union bound and the fact that D is MLD.

Noting that $\Delta(E(\mathbf{m}',\mathbf{y}) \leq \Delta(E(\mathbf{m}),\mathbf{y}) \leq (p+\varepsilon')n$ (see Figure 2), by (2) we have

$$\begin{aligned}
\mathbb{E}_E\left[\mathbf{1}_{D(\mathbf{y})\neq\mathbf{m}}\right] &\leq \sum_{\mathbf{m}'\neq\mathbf{m}} Pr\left[E\left(\mathbf{m}'\right) \in B\left(\mathbf{y},\left(p+\varepsilon'\right)n\right)|E(\mathbf{m})\right] \\
&= \sum_{\mathbf{m}'\neq\mathbf{m}} \frac{\left|B\left(\mathbf{y},\left(p+\varepsilon'\right)n\right)\right|}{2^n} &(3)\\
&\leq \sum_{\mathbf{m}'\neq\mathbf{m}} \frac{2^{H(p+\varepsilon')n}}{2^n} &(4)\\
&\leq 2^k \cdot 2^{-n(1-H(p+\varepsilon'))} \\
&\leq 2^{n(1-H(p+\varepsilon))-n(1-H(p+\varepsilon')n)} &(5)\\
&= 2^{-n(H(p+\varepsilon)-H(p+\varepsilon'))} &(6)
\end{aligned}$$

In the above (3) follows from the fact that the choice for $E(\mathbf{m}')$ is independent of $E(\mathbf{m})$. (4) follows from the upper bound on the volume of a Hamming ball that we have seen before while (5) follows from our choice of k.

Using (6) in (1), we get

$$\begin{aligned}
\mathbb{E}_E\left[\Pr_{\mathbf{e}\sim BSC_p}[D(E(\mathbf{m})+\mathbf{e})\neq\mathbf{m})]\right] &\leq e^{-\varepsilon'^2 n/2} + 2^{-n(H(p+\varepsilon)-H(p+\varepsilon'))} \sum_{\mathbf{y}\in B(E(\mathbf{m}),(p+\varepsilon')n} Pr[\mathbf{y}|E(\mathbf{m})] \\
&\leq e^{-\varepsilon'^2 n/2} + 2^{-n(H(p+\varepsilon)-H(p+\varepsilon'))} \leq 2^{-\delta' n},
\end{aligned}$$

where the second inequality follows from the fact that $\sum_{\mathbf{y}\in\{0,1\}^n} Pr[\mathbf{y}|E(\mathbf{m})] = 1$ while the last inequality follows for large enough n by picking $\delta' > 0$ small enough (and say $\varepsilon' = \varepsilon/2$).

Thus, we have show that for any arbitrary $\mathbf{m}$ the average (over the choices of E) decoding error probability is small. However, we still need to show that the decoding error probability is exponentially small for *all* messages *simultaneously*. Further, as the bound holds for each $\mathbf{m}$ we have

$$\mathbb{E}_{\mathbf{m}}\left[\mathbb{E}_E\left[\Pr_{\mathbf{e}\sim BSC_p}\left[D\left(E\left(\mathbf{m}\right)+\mathbf{e}\right)\neq\mathbf{m}\right]\right]\right] \leq 2^{-\delta' n}$$

The order of the summation in the expectation with respect to $\mathbf{m}$ and the summation in the expectation with respect to the choice of E can be switched (as the probability distributions are defined over different domains), resulting in the following expression:

$$\mathbb{E}_E\left[\mathbb{E}_{\mathbf{m}}\left[\underset{\mathbf{e}\sim BSC_p}{Pr}\left[D\left(E\left(\mathbf{m}\right)+\mathbf{e}\right)\neq\mathbf{m}\right]\right]\right]\leq 2^{-\delta' n}$$

By the probabilistic method, there exists an encoding function E^* (and a corresponding decoding function D^*) such that

$$\mathbb{E}_{\mathbf{m}}\left[\underset{\mathbf{e}\sim BSC_p}{Pr}\left[D^*\left(E^*\left(\mathbf{m}\right)+\mathbf{e}\right)\neq\mathbf{m}\right]\right]\leq 2^{-\delta' n} \tag{7}$$

(7) implies that the *average* decoding error probability is exponentially small. However, recall we need to show that the *maximum* decoding error probability is small. To achieve such a result, we will throw away half of the messages, i.e. "expurgate" the code. In particular, we will order the messages in decreasing order of their decoding error probability and then drop the top half. We claim that the maximum decoding error probability for the remaining messages is $2 \cdot 2^{\delta' n}$ (this essentially is Markov's inequality and is generally called the "averaging argument"). We will flesh out this latter argument in the next lecture.