# Shannon's Noisy Channel Theorem over a Binary Symmetric Channel

Ramesh Balaji

Rutgers University

August 23, 2023

# Context

- Data transfer is unreliable
- Eg. sending data over a network, eg. using TCP or UDP
- Have to find a way to correct data
- **Error-correcting codes (ECCs):** method to correct data after transmission

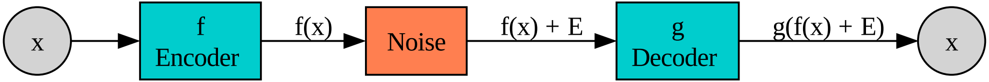
# Context: Representing Data

- Data transmitted can be represented as an array of bits.
- Array of bits as a column vector of 3 bits: $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.
- The set of all bitstrings with 3 bits is denoted as $\{0, 1\}^3$. Similarly, for $n$ bits, this is given as $\{0, 1\}^n$.

# Context: Encoder and Decoder Function

- ECCs have a **encoder** and **decoder**
- Encoder adds *additional data* to original data.
  - This extra data is used after transmission to recover the original data
  - Given as a function $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$.
  - Since there are more bits in the result, $m > n$.
- Decoder converts the *transmitted data* to the original message.
  - Given as a function $g : \{0, 1\}^m \rightarrow \{0, 1\}^n$.
- *Noise* from transmitting $f(\vec{x})$ over the channel.
  - Given as a vector $E \in \{0, 1\}^m$
  - Mathematically, added to the result $f(\vec{x})$ where addition is mod 2 (example will be provided later).

# Context: Encoder and Decoder



# Context: Binary Symmetric Channel

- How is the error vector $E \in \{0, 1\}^n$ generated?
- Different kinds of channels generate different types of noise.
- **Binary Symmetric Channel (BSC):** the probability of a bit flip in the input is $p$.
    - More mathematically, if $E_i$ represents the $i$th bit in $E$, then $E_i = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p } 1 - p \end{cases}$
    - Then, when $E$ is added to the input vector $f(\vec{x})$, it represents the output data *after* transmission over the channel.

# Example: Basic Error-Correction Code over a BSC

- **Encoder** will repeat every bit 3 times. Of every block, **decoder** will choose the bit in the block that occurs the most.
  - $f : \{0, 1\}^n \to \{0, 1\}^{3n}$
  - $g : \{0, 1\}^{3n} \to \{0, 1\}^n$
- Our message is $\vec{x} = \begin{bmatrix} 1 \end{bmatrix}$. Using row vectors to save space.
- $f(\vec{x}) = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$
- Suppose $p = 0.1$ and $E = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$.

| | | | | |
|---|---|---|---|---|
| | f(x) | 1 | 1 | 1 |
| + | E | 0 | 1 | 0 |
| | f(x) + E | 1 | 0 | 1 |

# Example: Basic Error-Correction Code over a BSC (Cont.)

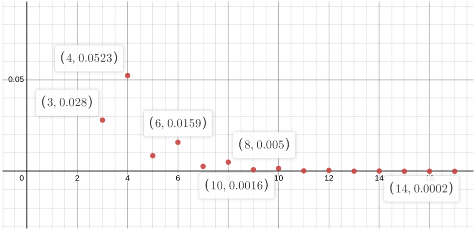
- Decoding: $f(\vec{x}) + E = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$
  - Most common bit is **1**, so the output is $\begin{bmatrix} 1 \end{bmatrix}$.
- Output $g(f(\vec{x}) + E) = \begin{bmatrix} 1 \end{bmatrix} = \vec{x}$.
  - Despite errors in the transmission, we still could decode the original message.

# Statistics on Example Transmission Scheme

- The encoder function is defined as $f : \{0,1\}^m \to \{0,1\}^n$
- The **rate of transmission** is defined as $\frac{m}{n}$.
  - For the example code, the rate of transmission $R = \frac{1}{3}$.
- **Probability of failure** of our sample code:
  - We need to find the probability that either $E$ has two 1s or three 1s.
  - $\binom{3}{2}p^2(1-p) + \binom{3}{3}p^3 = 0.028$

# Tradeoff Between Rate of Transmission and Probability of Failure

- What if we copy the bit more times?
  - If repeated $n$ times, then $R = \frac{1}{n}$
  - Probability of failure? Must be at least $\lceil \frac{n}{2} \rceil$ 1s in $E$ for failure.
  - $P[g(f(x) + E) \neq x] = \sum_{i=\lceil \frac{n}{2} \rceil}^{n} \binom{n}{i} p^i (1-p)^{n-i}$
  - Probability of failure for $n$ repeated bits




  - Observation: worse rate of transmission $(\frac{1}{n})$, but lower probability of failure.

# Shannon's Noisy Channel Coding Theorem over BSC

- Shannon's Noisy Channel Coding Theorem proves the existence ECC scheme with *theoretical* rate of transmission and failure probability.
- For a BSC with bit-flip probabliity $p$, for some arbitrarily small $\epsilon > 0$, there exists some ECC scheme with rate of transmission $1 - H(p) - \epsilon$, and probability of failure less than $\epsilon$.
  - Note that $H(p) = -p \log_2 p - (1 - p) \log_2 1 - p$, which is the binary entropy function.
- Does not tell us *what* that ECC scheme is, but states there exists one.

# “Proving” the Noisy Channel Coding Theorem

- Not a formal proof.
- **Two steps:**
  1. Define a coding scheme with the appropriate rate of transmission
  2. Prove that its probability of failure is less than $\epsilon$.

# Defining an ECC

- Define $\delta$ such that $p + \delta < 0.5$, and $H(p + \delta) < H(p) + \frac{\epsilon}{2}$.
- **Encoder:** $f : \{0, 1\}^{n(1-H(p)-\epsilon)} \to \{0, 1\}^n$. Thus the rate of transmission is correct.
  - Given an input, $f$ will output a random vector in $\{0, 1\}^n$ (there are some problems with this, namely that $f$ could end up not being a function, but I think the probability is low)
- **Decoder:** $g : \{0, 1\}^n \to \{0, 1\}^{n(1-H(p)-\epsilon)}$.
  - Given transmitted data $f(\vec{x}) + E$, choose the value $\vec{y} \in f(\{0, 1\}^n)$ such that the number of differing bits (called Hamming Distance) between $\vec{y}$ and $f(\vec{x}) + E$ is less than $n(p + \delta)$

# Probability of Failure

- **Two ways for failure to occur:**
  1. There is no vector in the range of $f$ that is within $n(p+\delta)$ from $f(\vec{x})+E$.
  2. There is a vector $\vec{z} \in f(\{0,1\}^n)$, where $\vec{z}$ is closer to $f(\vec{x})+E$ than $\vec{x}$ itself.
     - Mathematically, $\exists \vec{z} \in f(\{0,1\}^n)$ such that $\Delta(\vec{z}, f(\vec{x})+E) < \Delta(\vec{x}, f(\vec{x})+E)$ (note that $\Delta(\vec{a}, \vec{b})$ represents the Hamming Distance between $\vec{a}$ and $\vec{b}$)

# Case 1: Vector not Within $n(p+\delta)$

- In this case, the random variable $\Delta(E, f(\vec{x})+E)$ represents the number of 1s in $E$. This must be greater than $n(p+\epsilon)$
- Chernoff bound is decreasing. Note $np + np\epsilon < np + n\epsilon$, so $Pr[\Delta(E, f(\vec{x})+E) > np + n\epsilon] < Pr[\Delta(E, f(\vec{x})+E) > np + np\epsilon]$.
- We can use the Chernoff bound to know $Pr[\Delta(E, f(\vec{x})+E) > np(1+\epsilon)] < e^{-np\epsilon^2}$
- Thus $Pr[\Delta(E, f(\vec{x})+E) > n(p+\epsilon)] < e^{-np\epsilon^2}$
- For $n$ arbitrarily large, this probability exponentially decreases, and the probability will be less than epsilon.

# Case 2: Vector Closer to Output than $\vec{x}$

Take an arbitrary $f(\vec{x}) \in f(\{0,1\}^n)$

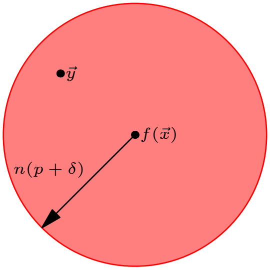



Figure: Hamming ball of volume $n(p+\delta)$

- The probability that a vector $\vec{y}$ exists within the Hamming ball is $\frac{Vol(n(p+\delta), f(\vec{x}))}{2^n}$, where $Vol(r, \vec{x})$ is the volume of the Hamming ball of radius $r$ centered at $\vec{x}$.
- Note there are $2^n$ vectors in $\{0,1\}^n$.

# Case 2: Vector Closer to Output than $\vec{x}$ (cont.)

Let $V_i$ represent the event that for $\vec{x}_i$, the $i$th vector in $\{0,1\}^{1-H(p)-\epsilon}$, there exists a $\vec{y}$ such that $\Delta(\vec{y}, f(\vec{x}_i) + E) < \Delta(\vec{x}_i, f(\vec{x}_i) + E)$. Already done on previous slide: $V_i = Vol(r, \vec{x})2^{-n}$

The probability of the union of these events (there are exactly $2^{n(1-H(p)-\epsilon)}$ events) can be bounded with the union bound.

$$Pr\left[\bigcup_{i=0}^{n(1-H(p)-\epsilon)} V_i\right] \leq \sum_{i=0}^{n(1-H(p)-\epsilon)} V_i = Vol(r, \vec{x})2^{-n}2^{n(1-H(p)-\epsilon)}$$

# Case 2: Vector Closer to Output than $\vec{x}$ (cont.)

Volume of a Hamming Ball

- Found through summing each "ring" of the ball
- Each "ring" has $\binom{n}{i}$ vectors in it (for a vector of size $n$)
- Total is $\sum_{i=0}^{n(p+\delta)} \binom{n}{i}$, where $n(p+\delta)$ is the radius

Approximation of Hamming Ball Volume

- Entropy function $H(p)$ is involved here
- Can use Stirling's approximation to expand and exponent properties to expand $2^{nH(p)}$ and find $\binom{n}{pn} \approx 2^{nH(p)}$.

# Case 2: Vector Closer to Output than $\vec{x}$ (cont.)

First simplify bounds for approximation of Hamming ball volume:

$$\begin{aligned}\sum_{i=0}^{n(p+\delta)} \binom{n}{i} &\leq \binom{n}{p(n+\delta)} \\ &\leq 2^{nH(p+\delta)} \\ &\leq 2^{n(H(p)+\frac{\epsilon}{2})}\end{aligned}$$

We need to expand $Vol(r, \vec{x})2^{-n}2^{n(1-H(p)-\epsilon)}$:

$$\begin{aligned}Vol(r, \vec{x})2^{-n}2^{n(1-H(p)-\epsilon)} &\leq 2^{n(H(p)+\epsilon)-n+n(1-H(p)-\epsilon)} \\ &\leq 2^{nH(p)+\frac{n\epsilon}{2}-n+n-nH(p)-n\epsilon} \\ &\leq 2^{-\frac{n\epsilon}{2}}\end{aligned}$$

Evidently, for $n$ large, the probability of the vector being within the Hamming ball is exponentially small and thus less than $\epsilon$.