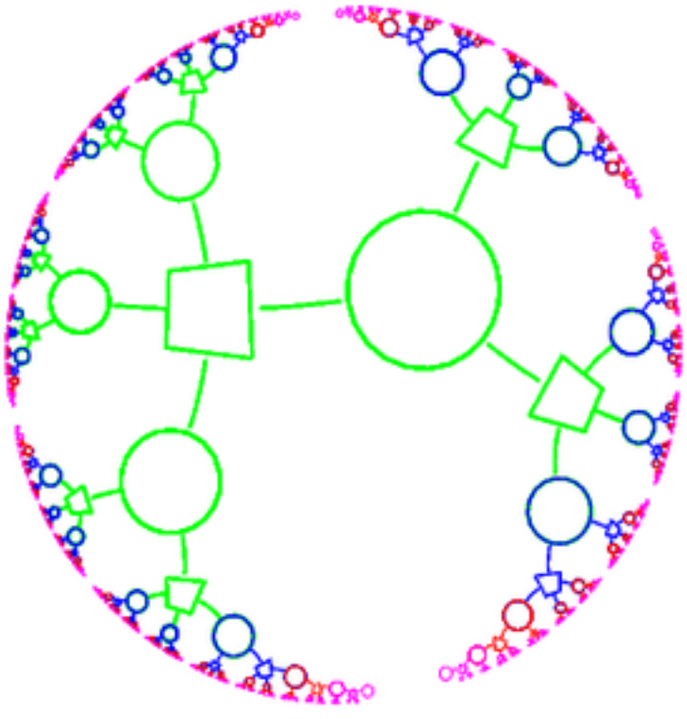
Proof of The Noisy-Channel Coding Theorem



Information theory, pattern recognition, and neural networks

- 1 Noisy-channel coding

Source coding (Data compression)

- 2 Information content, entropy
- 3 Typicality and the source coding theorem
- 4 Symbol codes
- 5 Symbol codes and Arithmetic coding

Noisy-channel coding

- 6 Inference and

  Information measures for noisy channels
- 7 Capacity of a noisy channel
- 8 The noisy-channel coding theorem

Definition of Capacity

The Capacity of a channel

is the maximum, over all input distributions $P(x)$, of the mutual information:

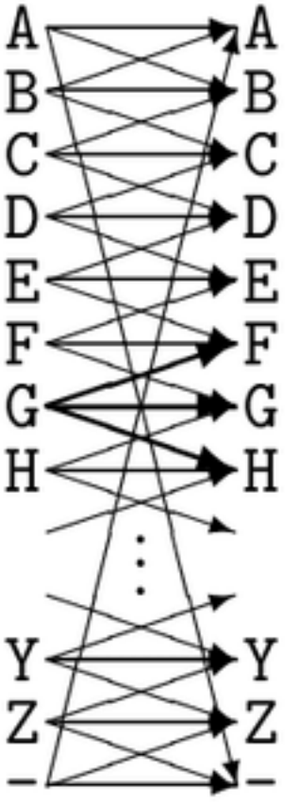
$$C \equiv \max_{\mathcal{P}_X} I(X;Y)$$

The distribution $\mathcal{P}_X^*$ that achieves the maximum is called the optimal input distribution.

- Shannon's noisy channel coding theorem:

Reliable (virtually error-free) communication is possible at rates up to C

Noisy typewriter

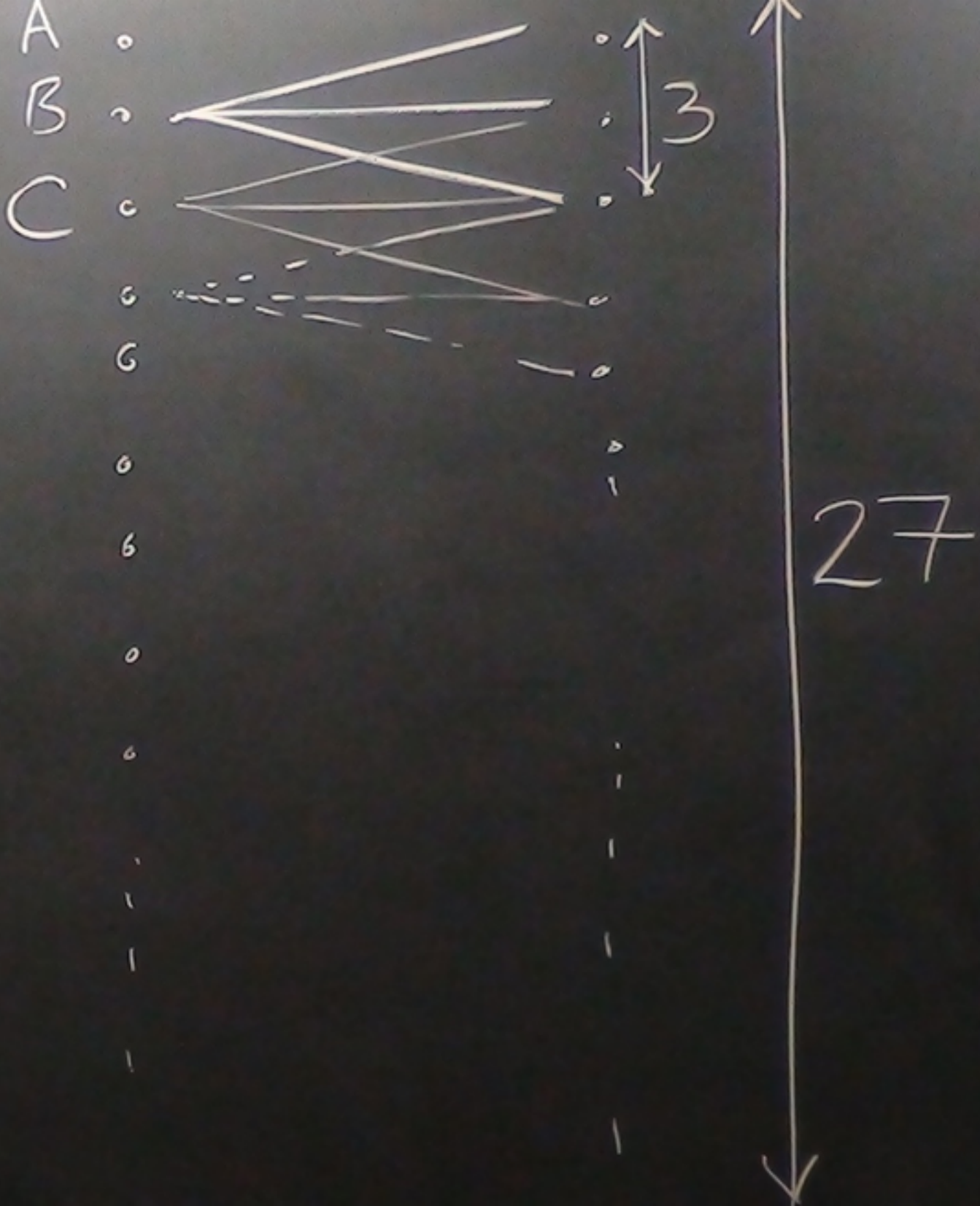


$$
\begin{aligned}
&\vdots \\
P(y=\mathtt{F}\,|\,x=\mathtt{G}) &= 1/3; \\
P(y=\mathtt{G}\,|\,x=\mathtt{G}) &= 1/3; \\
P(y=\mathtt{H}\,|\,x=\mathtt{G}) &= 1/3; \\
&\vdots
\end{aligned}
$$

- What is the optimal input distribution?
- What is the capacity?

NEL

EM



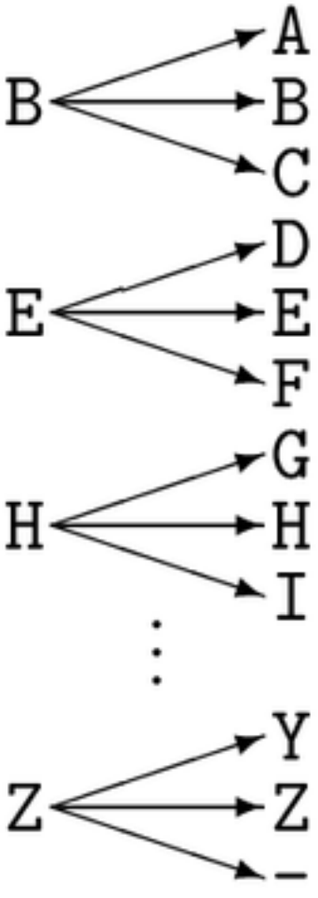
$$\text{Capacity} = \max_{P_X} I(X;Y)$$

$$= \max_{P_X} H(Y) - H(Y|X)$$

$$= \log 27 - \log 3$$

$$= \log 9 \text{ bits}$$

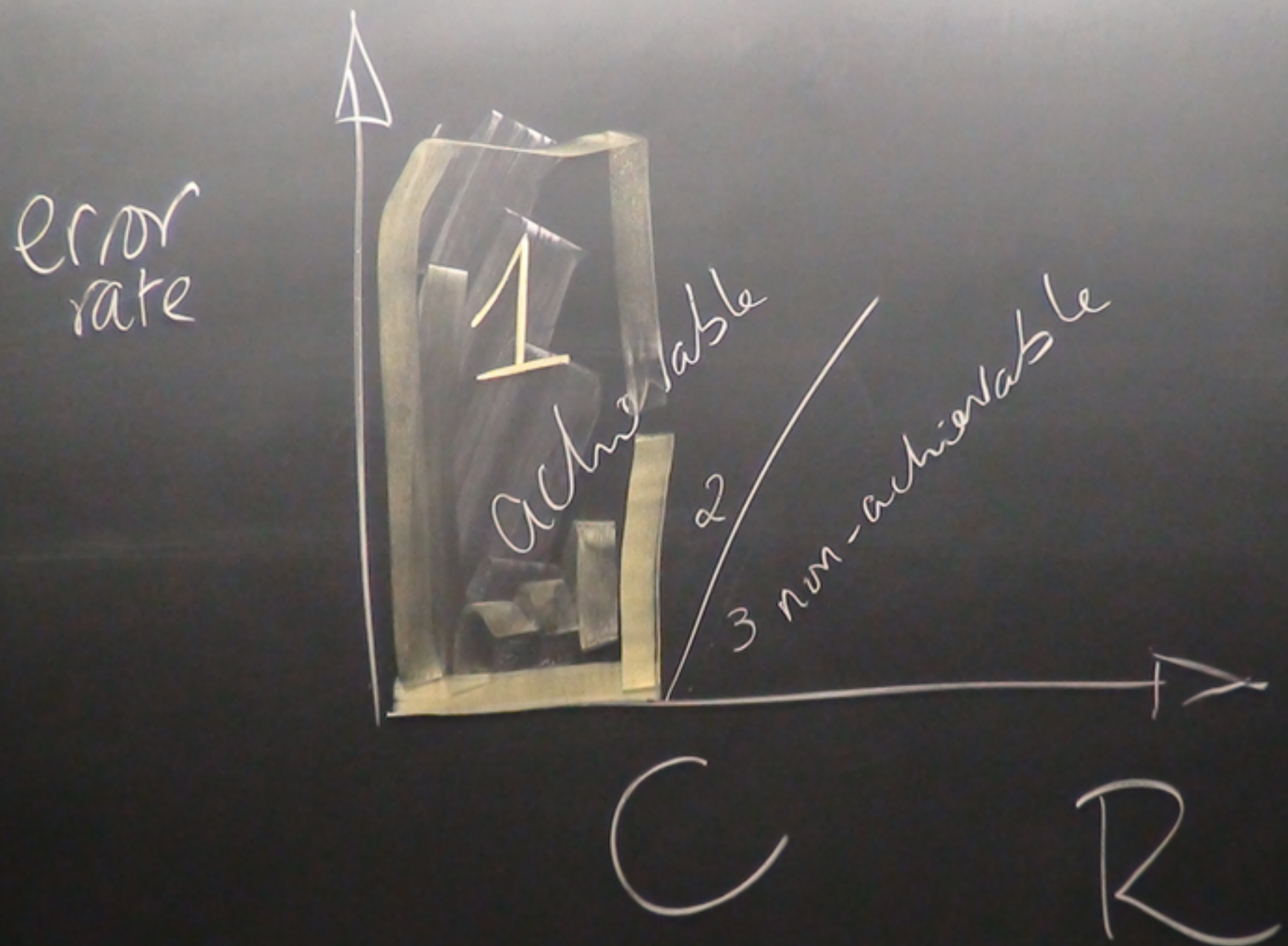
Non-confusable subset of inputs



reliable communication is possible at C = log 9 bits

$$P_X^* = \left\{ \frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \ldots\ldots, \frac{1}{27} \right\}$$

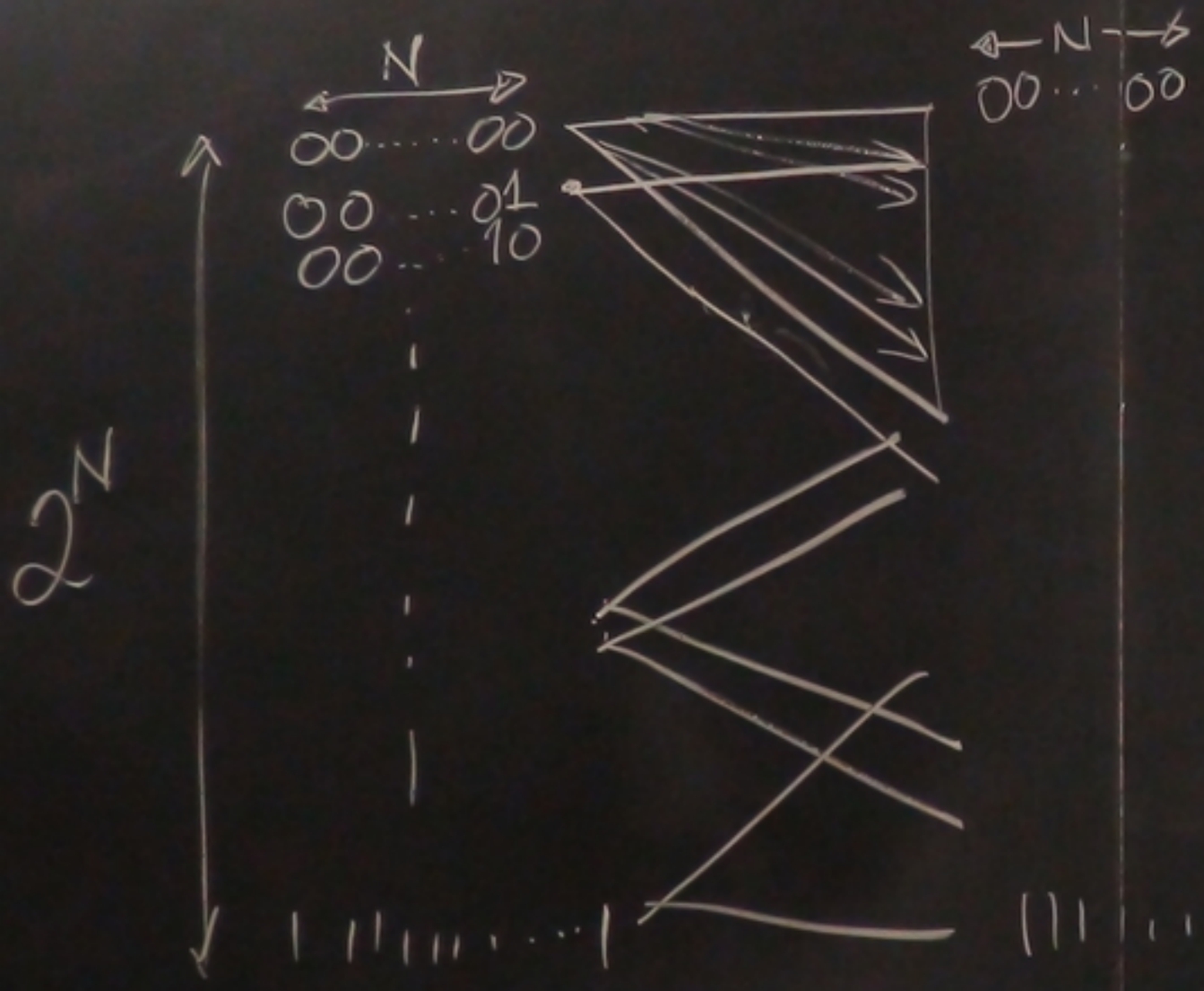
$$P_X^* = \left\{ 0, \frac{1}{9}, 0, 0, \frac{1}{9}, 0, 0 \ldots\ldots \right\}$$


error rate
1
achievable
2
3 non-achievable
C
R

Extended channel

eg BSC

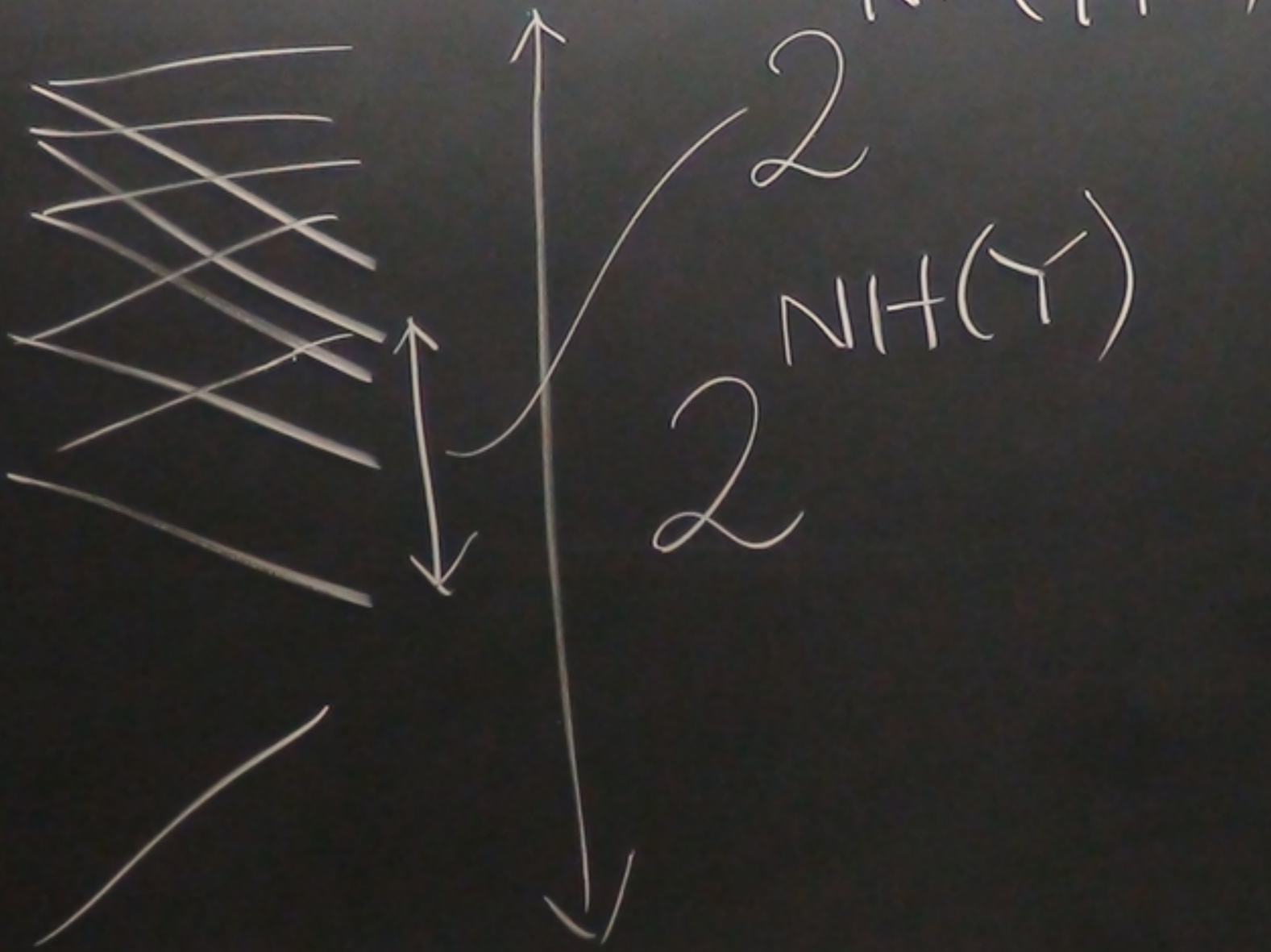
N uses



f^N flips

$\pm\sqrt{N}$

ρ_X



almost certainly

$$\#\left|\text{non-confusable inputs}\right| = \frac{2^{NH(Y)}}{2^{NH(Y|X)}} = 2^{NI(X;Y)}$$

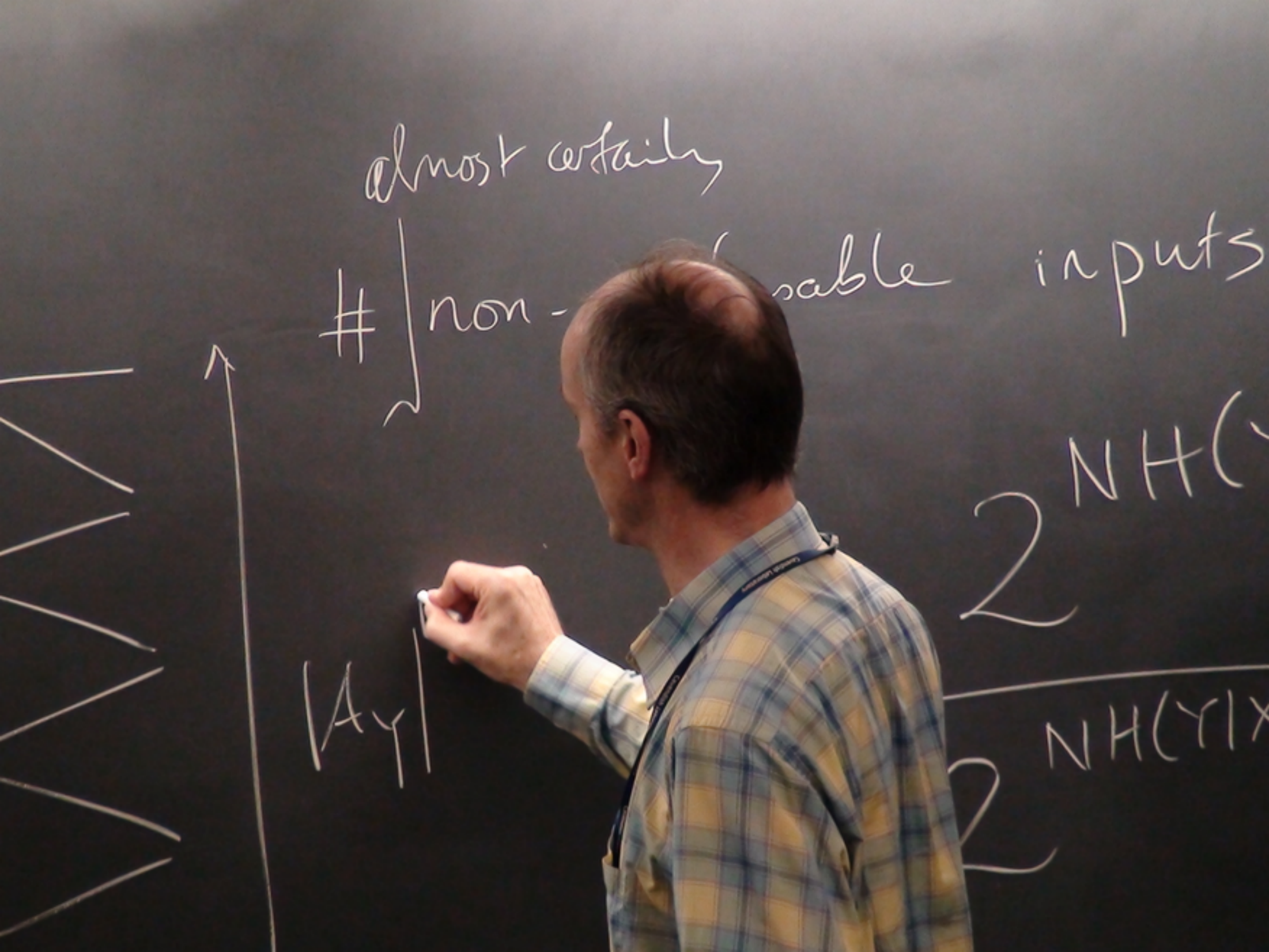
ity

nfusable inputs

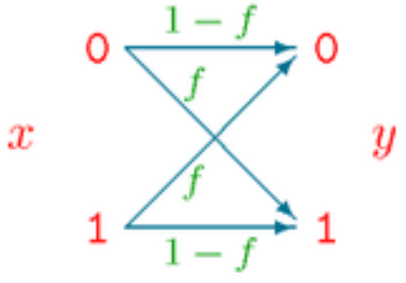
$$= \frac{2^{NH(Y)}}{2^{NH(Y|X)}} = 2^{NI(X;Y)}$$

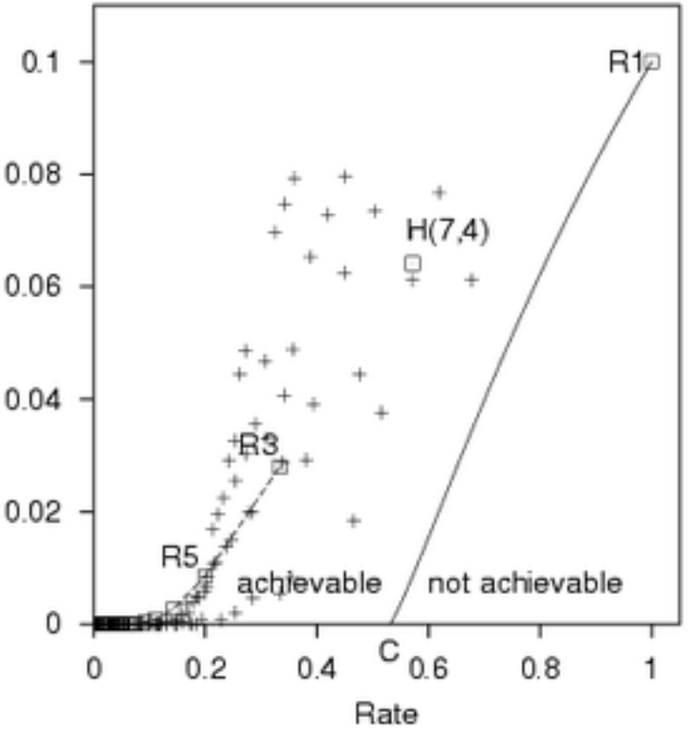
$$\leq 2^{NC}$$

$\Rightarrow$ NC bits per N


almost certainly
#
non-
sable
inputs
|Ay|
NH(
NH(Y|X

Shannon's noisy channel coding theorem





For any channel:

Reliable (virtually error-free) communication is possible

at rates up to C

For the BSC with flip prob f

(whose capacity is $C = 1 - H_2(f)$)

for any $\varepsilon > 0$ & $R < C$

for large enough N

$\exists$ a code of length N & rate $\geq R$

and a decoder s.t. the probability of block error is $< \varepsilon$

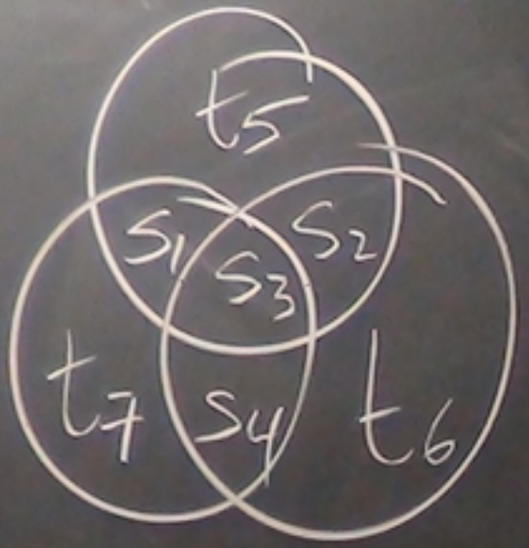
(7,4) Hamming Code

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \uparrow \\ M = 3 \\ \downarrow \end{matrix}$$

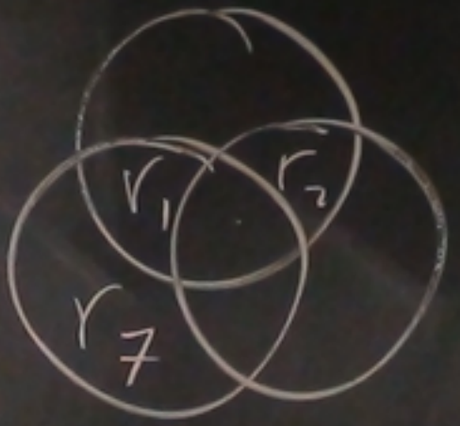
$$\longleftarrow \quad N = 7 \quad \longrightarrow$$

Valid transmissions $\mathbf{t}$ satisfy

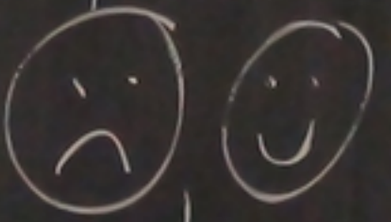
$$\mathbf{H}\,\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \bmod 2$$



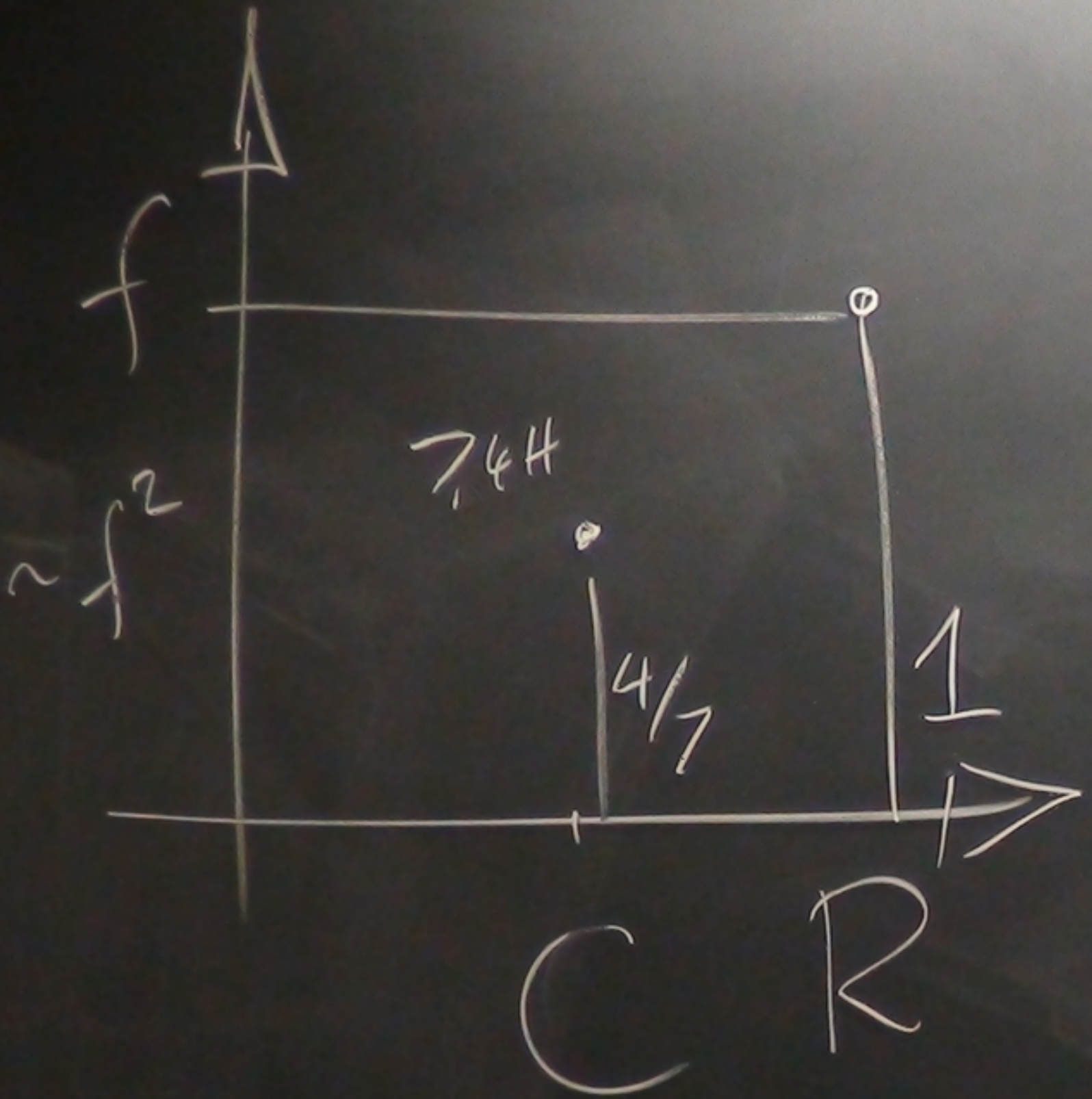
(7, 4)



syndrome



deduce most probable explanation


f
~f²
7,4H
4/7
1
C
R

(7,4) Hamming Code

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \uparrow \\ M = 3 \\ \downarrow \end{matrix}$$

$$\longleftarrow \quad N = 7 \quad \longrightarrow$$

Valid transmissions $\mathbf{t}$ satisfy

$$\mathbf{H}\,\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \bmod 2$$

Parity check matrix

$$H = \left[\begin{array}{cccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$K = 4$

$N = 7$

(7,4) Hamming Code

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \uparrow \\ M = 3 \\ \downarrow \end{matrix}$$

$$\longleftarrow \quad N = 7 \quad \longrightarrow$$

Valid transmissions $\mathbf{t}$ satisfy

$$\mathbf{H}\,\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \bmod 2$$

Received signal $\mathbf{r} = \mathbf{t} + \mathbf{n}$
Syndrome $\mathbf{z} = \mathbf{H}\mathbf{r} = \mathbf{H}\mathbf{n}$.
Syndrome decoder $\mathbf{z} \longrightarrow \hat{\mathbf{n}}$.

flip bit 2

t = 0000 000

r = 0100 000

Parity check matrix

$$H = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$K = 4$

$N = 7$

$r = t + n \pmod 2$

$z = Hr$

$= H(t + n)$

$= Ht + Hn$

$= Hn$

flip bit 2

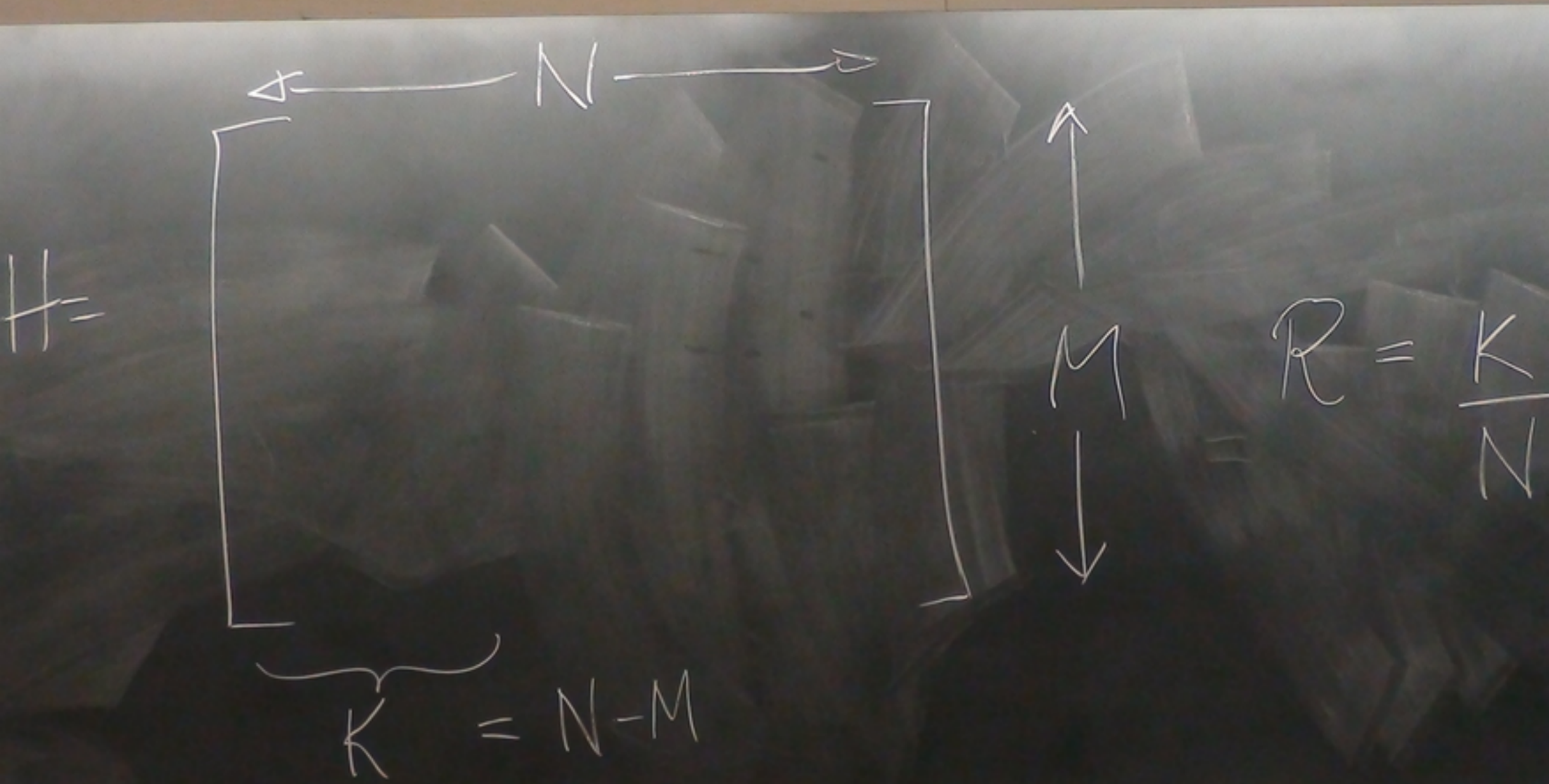
$t = 0000\ 0000$

$r = 0100\ 0000$

$z = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

$$z = Hr$$

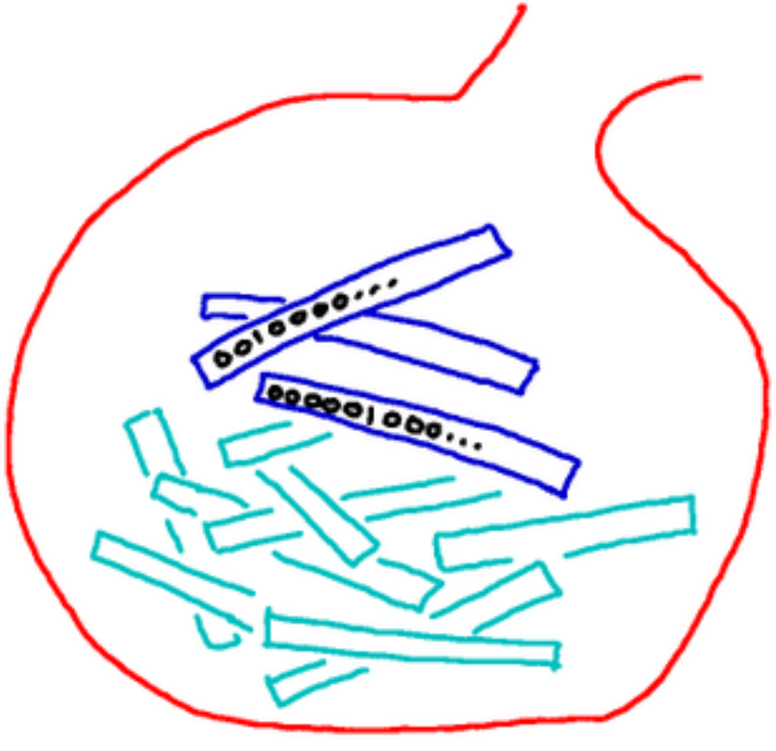
$$z = H\hat{n}$$


N
H=
M
R = K/N
K = N-M

How we won the bent coin lottery

Probability of '1' = f

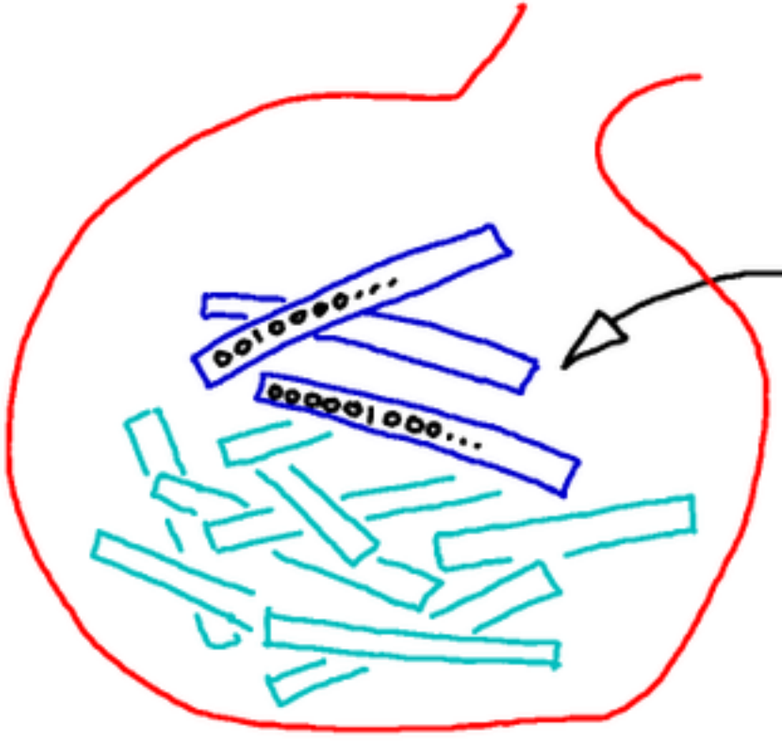
To have a 99.99% chance of winning,
we bought all the typical tickets



How we won the bent coin lottery

Probability of '1' = f

To have a 99.99% chance of winning,
we bought all the typical tickets



Number of tickets in 'typical set'

$|T| \simeq 2^{NH_2(f)}$

How to prove good codes exist

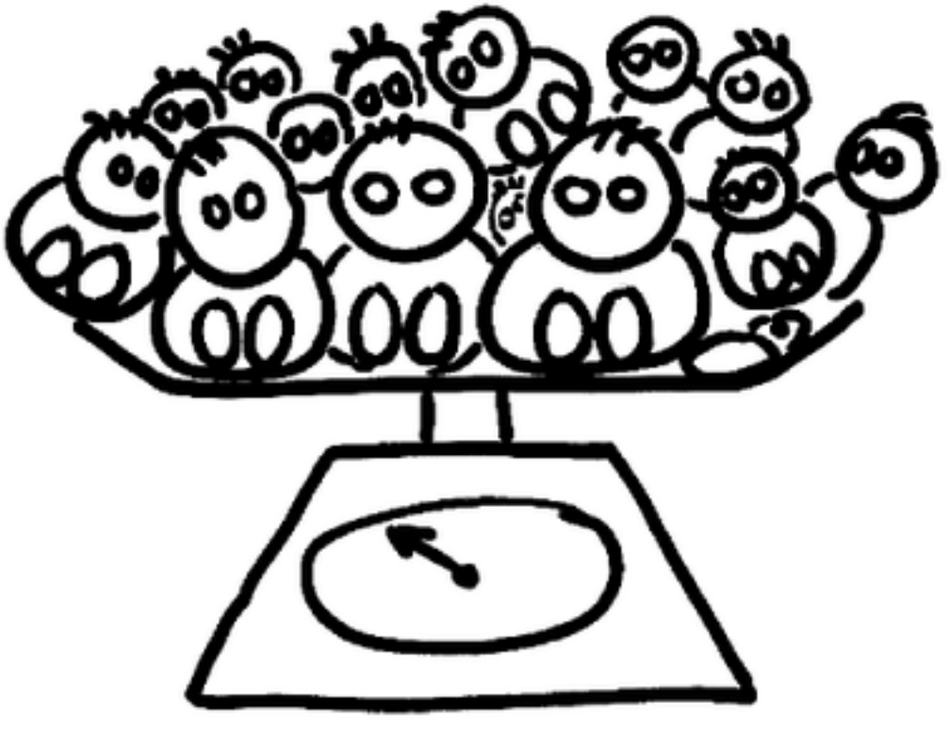
- Constructive proof

Given required $R < C$, and $\epsilon > 0$,

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 & & & \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & \cdots & \cdots & \cdots \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & & & \\ \vdots & & & \vdots & & & \vdots & \ddots & & \\ \vdots & & & \vdots & & & \vdots & & \ddots & \\ \vdots & & & \vdots & & & \vdots & & & \ddots \end{bmatrix}$$

- Non-constructive proof

Shannon's way of proving malnutrition



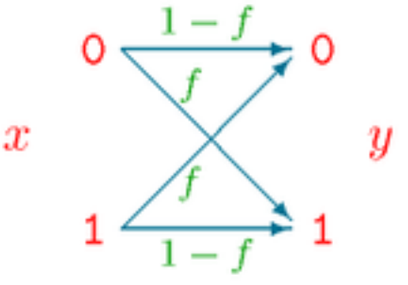
If average weight
of all babies is $< \epsilon$,
there must be (at least!)
one baby with weight $< \epsilon$.

Capacity

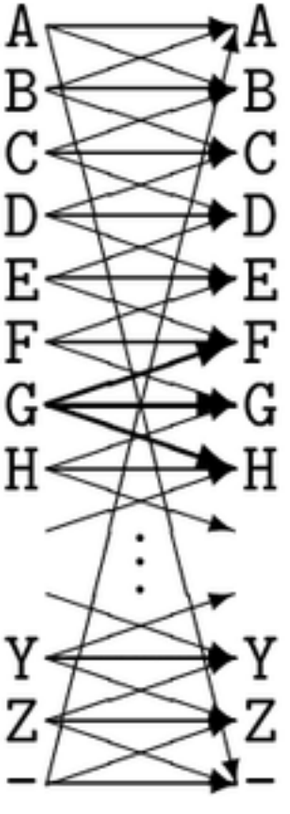
The Capacity of a channel

is the maximum, over all input distributions $P(x)$, of the mutual information:

$$C \equiv \max_{\mathcal{P}_X} I(X;Y)$$



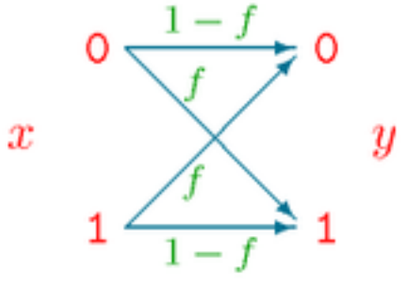
The distribution $\mathcal{P}_X^*$ that achieves the maximum is called the optimal input distribution.

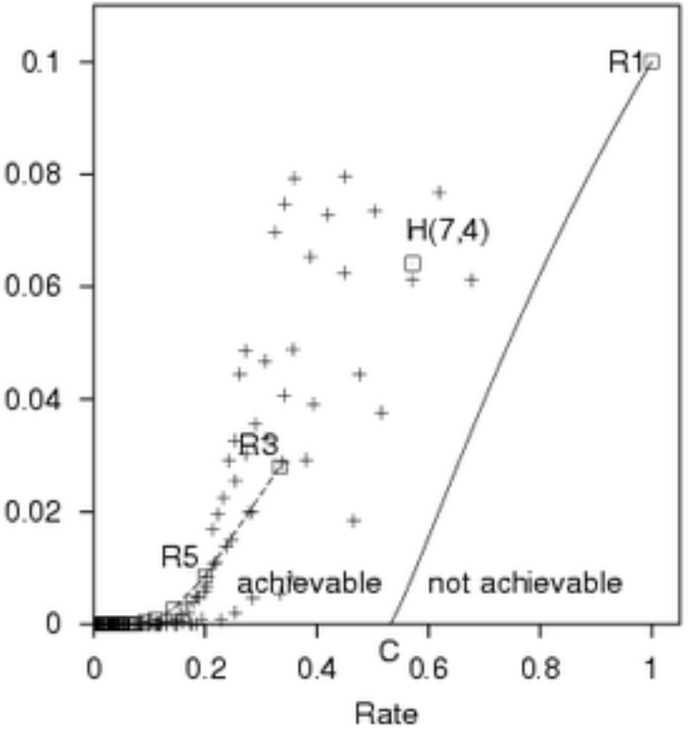


- Shannon's noisy channel coding theorem:

Reliable (virtually error-free) communication is possible at rates up to C

Proof for the Binary Symmetric Channel



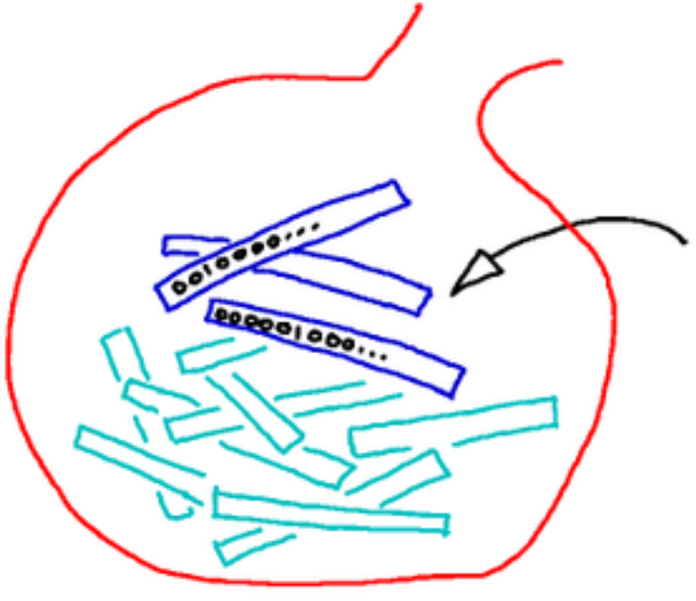


- Linear block code
- Syndrome decoder
- Use typical noise vectors
- Compute average probability of error

How we won the bent coin lottery

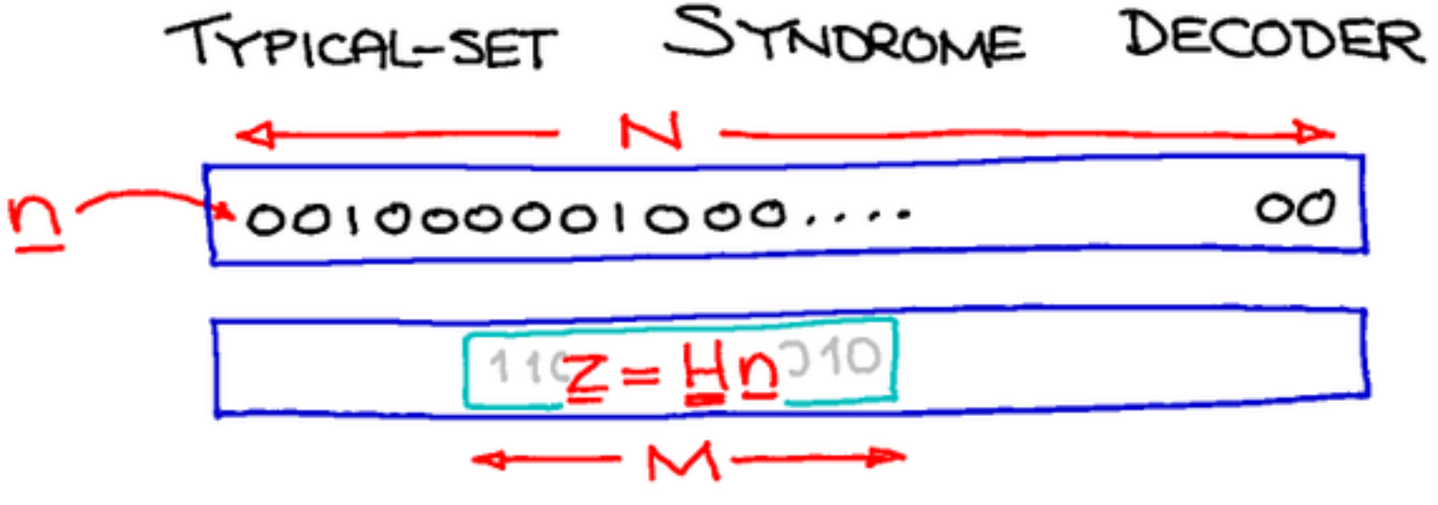
Probability of '1' = f

To have a 99.99% chance of winning, we bought all the typical tickets



Number of tickets in 'typical set'

$|T| \simeq 2^{NH_2(f)}$



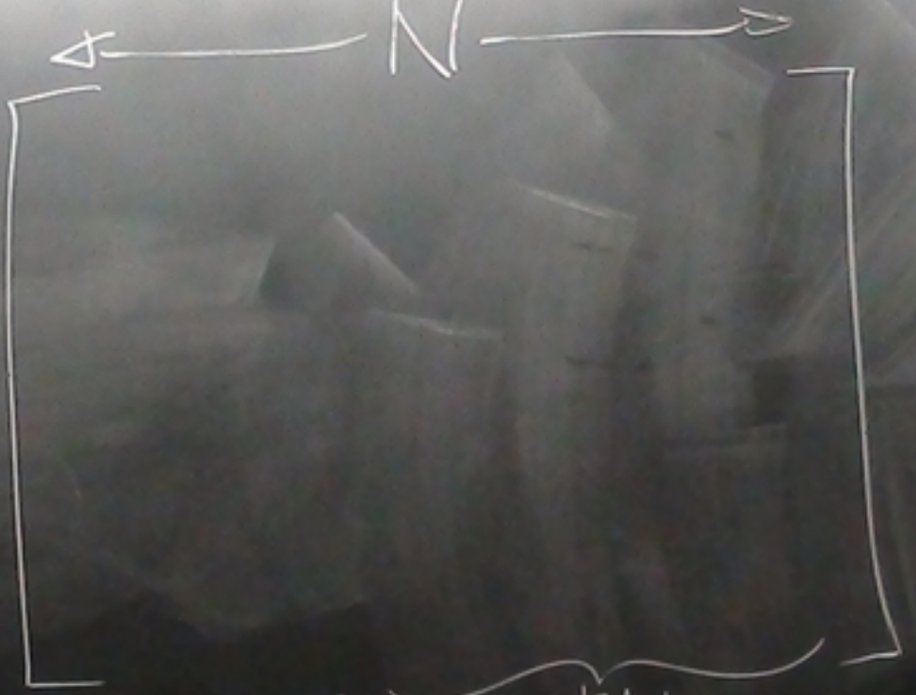
$r = t + n$

$z = Hr$

$z = Hn$?

$\hat{n}$

$H =$



$K = N - M$ set M bits

$t_1\ t_2\ t_3 \ldots\ t_K\ t_{K+1} \ldots\ t_N$

$\underline{\underline{H}}\underline{t} = \underline{0}$

Probability of error of this code w/ $\underline{\underline{H}}$ $= P_I + P_{II}$

P_I: Prob that $\underline{n}$ is not in the bag

↑ indep of $\underline{\underline{H}}$

P_{II}: $\underline{n}$ is in the bag, but other $\underline{\tilde{n}}$ are also in the bag

w/

$$\underline{\underline{H}}\,\underline{\tilde{n}} = \underline{z}$$

$$\underline{\underline{H}}\,(\underline{n} - \underline{\tilde{n}}) = \underline{0}$$

f this
code
w/ $\underline{\underline{H}}$.

Prob that
$\underline{r}$ is not
in the bag

↑

indep of $\underline{\underline{H}}$

$\underline{r}$ is in the
but other
also in the

w/

H

$T = 2^{NH_2(f)^t}$ → 0
by picking N
large

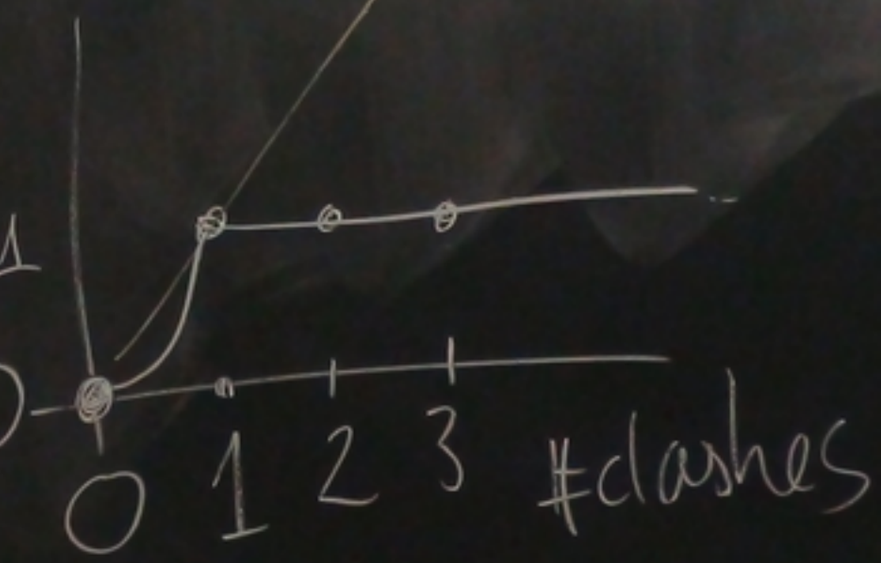
H(

$$P_{\text{II}}(H) = \sum_{\substack{n \\ n \in \text{bag}}} P(n)\ \mathbb{1}\Big(\underbrace{\exists\, \tilde{n}:\ \tilde{n} \neq n,\ \tilde{n} \in \text{bag},\ H(n-\tilde{n})=0}\Big)$$

#clashes $\geqslant 1$

#clashes

$\mathbb{1}(\#\text{clashes} \geqslant 1)$:



$$\leq \sum_{\underline{n}} P(\underline{n}) \sum_{\substack{\tilde{n} \neq n \\ \tilde{n} \in \text{bag}}} \mathbb{1}\left[H(n - \tilde{n}) = 0\right]$$

$$\leq \sum_{\underline{n}} P(\underline{n}) \sum_{\substack{\tilde{n} \neq n \\ \tilde{n} \in \text{bag}}} \mathbb{1}\left[\underline{\underline{H}}(n-\tilde{n}) = 0\right]$$

$$\underline{x} = \underline{n} - \tilde{\underline{n}}$$

$$\left\langle P_{\Pi} \right\rangle_{\underline{\underline{H}}} = \sum_{H} P(\underline{\underline{H}}) \sum_{\underline{n} \in \text{bag}} P(\underline{n}) \sum_{\substack{\tilde{n} \text{ in bag} \\ \tilde{n} \neq n}} \mathbb{1}\left(\underline{\underline{H}}\,\underline{x} = 0\right)$$

$$= \sum_{n \in \text{bag}} P(\underline{n}) \sum_{\substack{\tilde{n} \in \text{bag} \\ \tilde{n} \neq n}} \sum_{H} P(H)\, \mathbb{1}\left(H\,\underline{x} = 0\right) \quad \text{where } \underline{x} = n - \tilde{n}$$

Q

If $\mathbf{x}$ is a fixed non-zero binary vector of length N and $\mathbf{h}$ is a random $\left(\frac{1}{2}, \frac{1}{2}\right)$ binary vector of length N, what is the probability that

$$\mathbf{h}^{\mathsf{T}}\mathbf{x} = 0 \bmod 2?$$

e.g.

$\mathbf{x}$:	0	1	1	1
$\mathbf{h}$:	h_1	h_2	h_3	h_4

What is Prob that $\underline{h}^{(i)} \cdot \underline{x} = 0$?

eg $\underline{h} \cdot \underline{x} = h_2 + h_3 + h_4 \mod 2$

Q

If $\mathbf{x}$ is a fixed non-zero binary vector of length N and $\mathbf{H}$ is a random $\left(\frac{1}{2}, \frac{1}{2}\right)$ binary matrix,

$$\mathbf{H} = \begin{bmatrix} \leftarrow & \mathbf{h}^{(1)} & \rightarrow \\ \leftarrow & \mathbf{h}^{(2)} & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & N & \rightarrow \end{bmatrix} \begin{matrix} \uparrow \\ M \\ \downarrow \end{matrix}$$

what is the probability that

$$\mathbf{H}\mathbf{x} = \mathbf{0} \bmod 2?$$

What is Prob that $\underline{h}^{(1)} \cdot \underline{x} = 0$? $= \frac{1}{2}$

eg $\underline{h} \cdot \underline{x} = h_2 + h_3 + h_4 \mod 2$

What is Prob $\left(\underline{H} \underline{x} = 0 \right)$? $= \left(\frac{1}{2} \right)^M$ ✓

$$\left\langle P_{\underline{H}}(H) \right\rangle_H = \sum_{n \in \text{bag}} P(n) \sum_{\substack{\tilde{n} \in \text{bag} \\ \tilde{n} \neq n}} \left(\frac{1}{2} \right)^M \leq 1 \times 2^{N H_2(f)} \left(\frac{1}{2} \right)^M$$

$\frac{1}{2}$

$$P_{\text{prob of error}} \leq P_I + \frac{1}{2^{M - NH_2(f)}}$$

1

$\left(\frac{1}{2}\right)^M$

vanishes if $M \gg NH_2(f)$

ie $\frac{M}{N} > H_2(f)$

$1 - \frac{M}{N} < 1 - H_2(f)$

$1 - \frac{M}{N}$ $\qquad$ $1 - H_2(f)$

if $R < C$

QED

$\underline{z} = \underline{\underline{H}}(\underline{t} + \underline{n})$

$= \underline{\underline{H}}\,\underline{n}$

$\underline{n} \rightarrow$ 0000010100 000

#1's is $Nf \pm \alpha\sqrt{Nf}$

$\underline{t}$ $\boxed{s_1 s_2 s_3 \ldots\ldots s_K t_{K+1} \ldots\ldots t_N}$

K $\qquad$ M

Homework recommendations

- Noisy channels - Chapters 8, 9, 10 (10.1-10.4 only)
  - Exercises 9.17 (p155); 10.12 (172); 15.12 (235)
  - and (if you want more practice) 15.11, 15.13, 15.15

- Invent a channel to pose to your colleagues:
  - 'what's the capacity of _this_?'