Chapter 1:

Introduction to Information Theory

Book: **"Information Theory, Inference, and Learning Algorithms"**

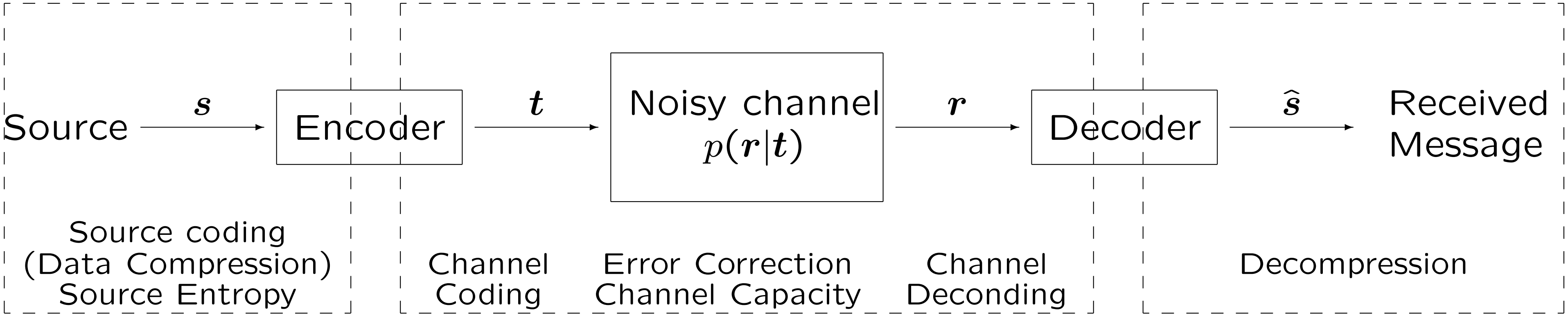
from David MacKay

- Modem $\longrightarrow$ phone line $\longrightarrow$ modem
- ESA $\longrightarrow$ radio waves in space $\longrightarrow$ Beagle 2
- Parental cell $\longrightarrow$ DNA $\longrightarrow$ daughter cells
- RAM $\longrightarrow$ hdd $\longrightarrow$ RAM

How to reduce the probability of error?

Noisy Channels - Error correcting Codes

Examples - System solution - Channel models - Binary symmetric Channel



- Data Compression
  (Removing Redundancy)
  $\rightarrow$ *Source Coding Theorem:* What compression rates are achievable.

- Error Correction (Adding redundancy)
  $\rightarrow$ *Channel Coding Theorem:* What transmission rates are acievable with infinitely small error.

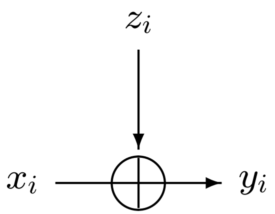
- Encryption: Between Source- and Channel Coding

- Decoding & Encoding should be fast.

Noisy Channels - Error correcting Codes

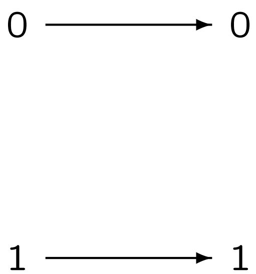
Examples - System solution - Channel models - Binary symmetric Channel

- Gaussian channel:



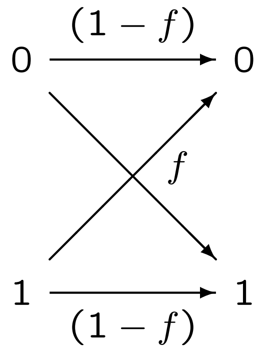
$P(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2}\right\}$

- Noiseless binary channel:



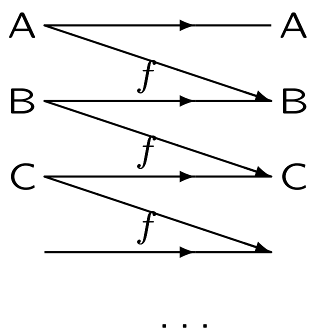
$P(y=0|x=0) = P(y=1|x=1) = 1$

- Binary symmetric channel:



$P(y=0|x=0) = 1-f$ $\quad P(y=0|x=1) = f$
$P(y=1|x=0) = f$ $\quad P(y=0|x=0) = 1-f$

- Noisy typewriter channel



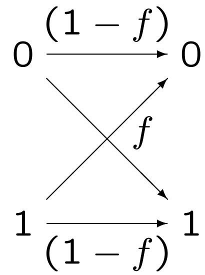
$P(y=A|x=A) = 1-f$ $\quad P(y=B|x=A) = f$
$P(y=B|x=B) = 1-f$ $\quad P(y=C|x=B) = f$
$\dots$ $\quad \dots$

Coding Theory

- The object of coding is to introduce redundancy so that if some of the information is lost or corrupted, it will still be possible to recover the message at the receiver.

Repetition Codes (e.g. $\mathcal{R}_3$: $0 \rightarrow 000$ and $1 \rightarrow 111$)

$\boldsymbol{s}$	0	0	1	0	1	1	0
$\boldsymbol{t}$	000	000	111	000	111	111	000
$\boldsymbol{n}$	000	001	000	000	101	000	000
$\boldsymbol{r}$	000	001	111	000	010	111	000



- Optimal decoding?

  Most probable $p(\boldsymbol{s}|\boldsymbol{r})$

- For a single bit $P(s|r_1r_2r_3) = \frac{P(r_1r_2r_3|s)P(s)}{P(r_1r_2r_3)}$

- if $P(s=1|\boldsymbol{r}) > P(s=0|\boldsymbol{r})$ decode $\hat{s}=1$ else $\hat{s}=0$
  BSC: $P(\boldsymbol{r}|s) = P(\boldsymbol{r}|\boldsymbol{t}(s)) = \prod_{n=1}^{3} P(r_n|t_n(s))$

- Odds ratio:
  $$\frac{P(s=1|\boldsymbol{r})}{P(s=0|\boldsymbol{r})} = \frac{P(\boldsymbol{r}|s=1)}{P(\boldsymbol{r}|s=0)} = \prod_{n=1}^{3} \frac{P(r_n|t_n(1))}{P(r_n|t_n(0))}$$

- assume: $p(0) = p(1) = \frac{1}{2}$

- bin. sym. channel

- $$\frac{P(r_n|t_n(1))}{P(r_n|t_n(0))} = \begin{cases} \frac{(1-f)}{f} & : r_n = 1 \\ \left(\frac{(1-f)}{f}\right)^{-1} & : r_n = 0 \end{cases}$$

Noisy Channels - Error correcting Codes

Repetition Codes - Block Codes - Channel capacity

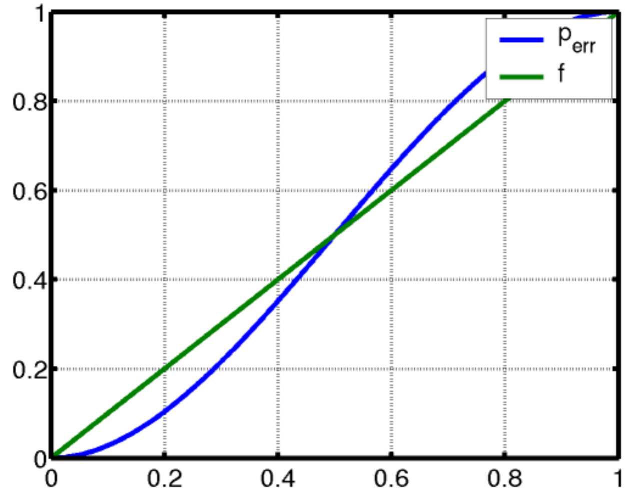
Received sequence $\boldsymbol{r}$	Likelihood ratio $\frac{P(\boldsymbol{r}\,\vert\, s=1)}{P(\boldsymbol{r}\,\vert\, s=0)}$ $\left(\gamma = \frac{1-f}{f} \gg 1\right)$	Decoded sequence $\hat{\boldsymbol{s}}$
000	γ^{-3}	0
001	γ^{-1}	0
010	γ^{-1}	0
100	γ^{-1}	0
101	γ^{1}	1
110	γ^{1}	1
011	γ^{1}	1
111	γ^{3}	1

$$
\begin{array}{rccccccc}
\boldsymbol{s} & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
\boldsymbol{t} & \overbrace{0\,0\,0} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} \\
\boldsymbol{n} & 0\,0\,0 & 0\,0\,1 & 0\,0\,0 & 0\,0\,0 & 1\,0\,1 & 0\,0\,0 & 0\,0\,0 \\
\hline
\boldsymbol{r} & \underbrace{0\,0\,0} & \underbrace{0\,0\,1} & \underbrace{1\,1\,1} & \underbrace{0\,0\,0} & \underbrace{0\,1\,0} & \underbrace{1\,1\,1} & \underbrace{0\,0\,0} \\
\hat{\boldsymbol{s}} & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
\text{corrected errors} & & \star & & & & & \\
\text{undetected errors} & & & & & \star & &
\end{array}
$$

What do we gain by using $\mathcal{R}_3$?

Two possibilities for errors, which follow the binomial distribution:
$p(e|f,N) = \binom{N}{r} f^e (1-f)^{N-e}$

- All three bits flipped $p_{\#3} = f^3$
- Just two bits flipped
  $p_{\#2} = 3f^2(1-f)$



Probability of error in $\mathcal{R}_3$ is $p_{\mathrm{B}} = p_{\mathrm{b}} = f^3 + 3f^2(1-f) = 3f^2 - 2f^3$

Error rate of $\mathcal{R}_N$ Codes

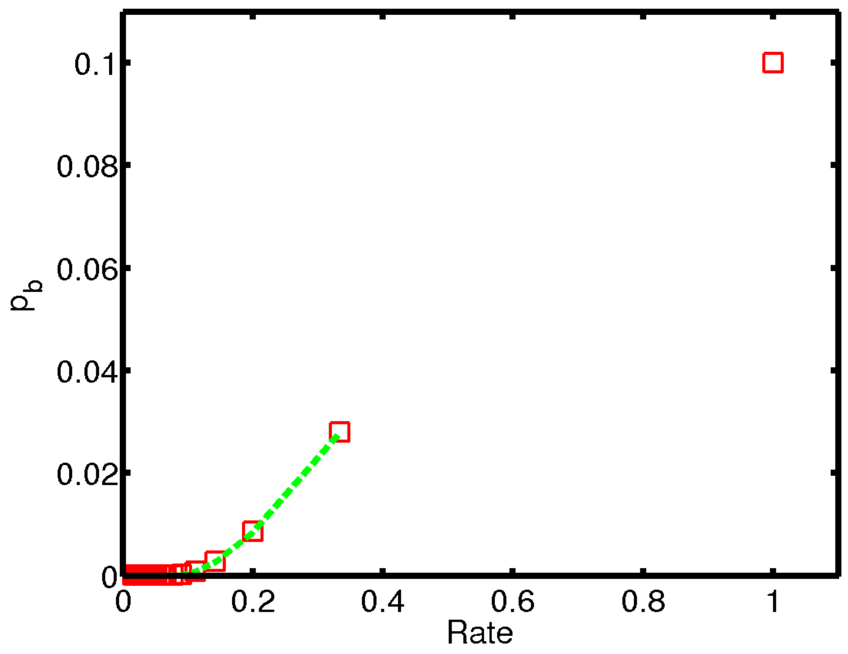
- Error when at least $\lceil N/2 \rceil$ bits in one block are flipped.

$$p_B = \sum_{n=(N+1)/2}^{N} \binom{N}{n} f^n (1-f)^{N-n}$$

- For small f this term is dominated by $n = \frac{(N+1)}{2}$.

- *Def.:* The *(transmission) rate* $R = \frac{\log(\mathcal{M})}{N}$ bits per transmission.

- The rate of $\mathcal{R}_3$ is $R = \frac{1}{3}$.



- Concatenated codes: $\mathcal{R}_3^2 = \mathcal{R}_3 \circ \mathcal{R}_3$

$p_b(\mathcal{R}_3^2) \approx 3\left(3f^2\right)^2 = 27f^4$
$p_b(\mathcal{R}_9) \approx \binom{9}{5} f^5 (1-f)^4 \approx 126 f^5$

but $\mathcal{R}_3^2$ requires less computation

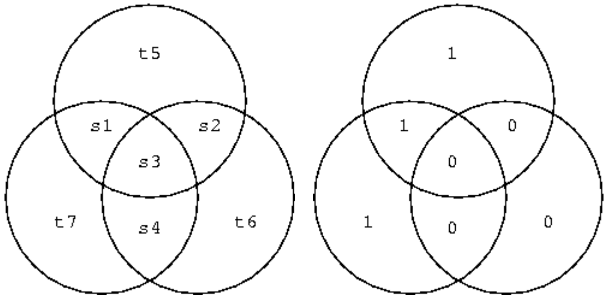
- *Parity check code:*

$$(7,4) - Hamming\,Code:$$

$$\underbrace{}_{}\quad\underbrace{}_{}$$

information bits parity bits

s_1	s_2	s_3	s_4	t_5	t_6	t_7



- linear code: $\boldsymbol{t} = \boldsymbol{G}\boldsymbol{s}$

- $$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{I}_4 \\ \boldsymbol{P} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

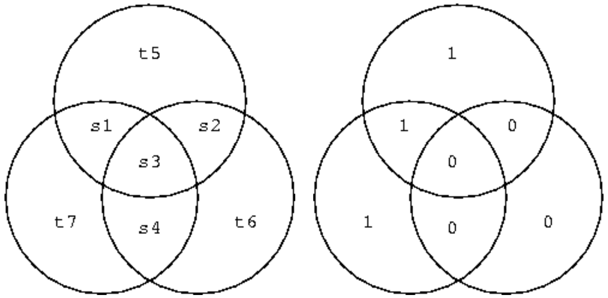
Noisy Channels - Error correcting Codes

Repetition Codes - Block Codes - Channel capacity

- *Parity check code:*

$$(7,4) - Hamming\, Code:$$

$$\underbrace{s_1 \quad s_2 \quad s_3 \quad s_4}_{\text{information bits}} \quad \underbrace{t_5 \quad t_6 \quad t_7}_{\text{parity bits}}$$



- linear code: $\boldsymbol{t} = \boldsymbol{G}\boldsymbol{s}$

- $\boldsymbol{G} = \begin{bmatrix} \boldsymbol{I}_4 \\ \boldsymbol{P} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$

$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

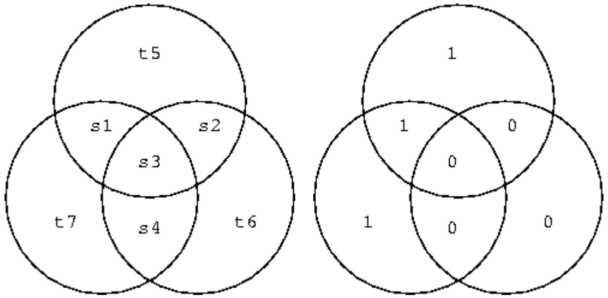
Noisy Channels - Error correcting Codes

Repetition Codes - Block Codes - Channel capacity

- *Parity check code:*

$$(7,4) - Hamming\,Code:$$

$$\overbrace{s_1 \quad s_2 \quad s_3 \quad s_4}^{\text{information bits}} \quad \overbrace{t_5 \quad t_6 \quad t_7}^{\text{parity bits}}$$



- linear code: $\boldsymbol{t} = \boldsymbol{G}\boldsymbol{s}$

- $$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{I}_4 \\ \boldsymbol{P} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

Noisy Channels - Error correcting Codes

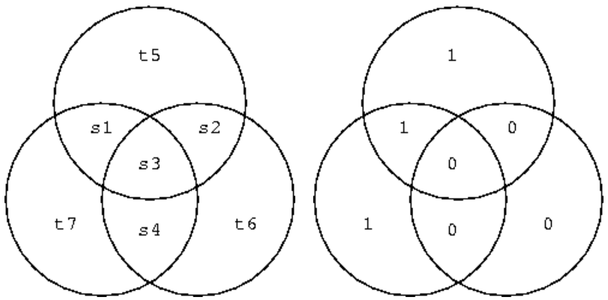
Repetition Codes - Block Codes - Channel capacity

- *Parity check code:*

$$(7,4) - Hamming\,Code:$$

information bits $\quad$ parity bits

$$\underbrace{s_1 \quad s_2 \quad s_3 \quad s_4}_{} \quad \underbrace{t_5 \quad t_6 \quad t_7}_{}$$



- linear code: $\boldsymbol{t} = \boldsymbol{G}\boldsymbol{s}$

- $\boldsymbol{G} = \begin{bmatrix} \boldsymbol{I}_4 \\ \boldsymbol{P} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$

$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$	$\boldsymbol{s}$	$\boldsymbol{t}$
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

Decoding scheme

- Minimal *distance* between code words is 3
- For the binary symmetric channel and equiprobable source vectors $\boldsymbol{s}$ One deconding scheme is to take the "closest" vector

$$\min_{\boldsymbol{s}} d(\boldsymbol{r}, \boldsymbol{t}(\boldsymbol{s}))$$

$\Longrightarrow$ Search all possible sources.

- *parity-check matrix:*

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{I}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

- $\forall$ code words:

$$\boldsymbol{H}\boldsymbol{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- syndrome vector: $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{r}$

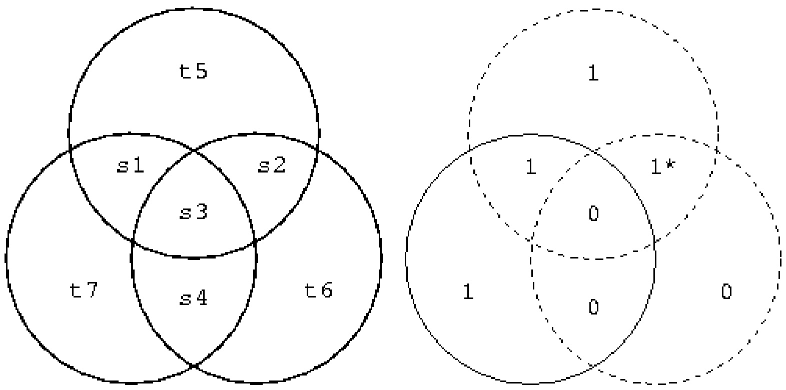
Correcting Errors

- Example:

$$\begin{array}{rl} \text{Transmit } \boldsymbol{s} = & 1000 \\ \text{Encoded } \boldsymbol{t} = & 1000101 \\ \text{Noise } \boldsymbol{n} = & 0100000 \\ \hline \text{Received } \boldsymbol{r} = & 1100101 \end{array}$$

- $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{r} = \begin{bmatrix} 1\,1\,0 \end{bmatrix}^{\mathsf{T}}$

- Pictoral solution:



Syndrome $\boldsymbol{z}$	000	001	010	011	100	101	110	111
Unflip this bit	*none*	r_7	r_6	r_4	r_5	r_1	r_2	r_3

Properties of the (7,4)-Hamming codes

- 8 syndromes (7 errors, 1 for the zero noise) are most probably caused by one error.
- What if $\boldsymbol{n}$ has wight 2?
- Example:

$$\begin{array}{rl} \text{Transmit } \boldsymbol{s} = & 1000 \\ \text{Encoded } \boldsymbol{t} = & 1000101 \\ \text{Noise } \boldsymbol{n} = & 0100010 \\ \hline \text{Received } \boldsymbol{r} = & 1100111 \end{array}$$

- $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{r} = \begin{bmatrix} 1\,0\,0 \end{bmatrix}^{\mathsf{T}} \rightarrow$ flip r_5
  $\hat{\boldsymbol{s}} = 1100011$

- codeword *distance*: 3
  $\rightarrow$ Only when 2 or more bits are flipped we get errors.

  Block error: $p_B = \sum_{r=2}^{7} \binom{7}{r} f^r (1-f)^{7-r}$.

  Bit error: $p_b = \frac{3}{7} p_B$

  The leading term for small f is $21f^2$
  $\Longrightarrow p_B \approx O(f^2)$

- The rate is $R = \frac{4}{7}$

Symmetry of the (7,4)-Hamming code

- Parity check matrix

  $$H = \begin{bmatrix} P & I_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

- $(t_1 t_2 t_3 t_4 t_5 t_6 t_7) \to (t_5 t_2 t_3 t_4 t_1 t_6 t_7)$

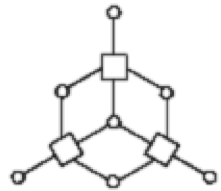
  $$\to H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

- Adding two parity constaints leads to a new one
  $(1) + (2) = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$
  which checks $t_5 + t_1 + t_4 + t_6 =$ even.

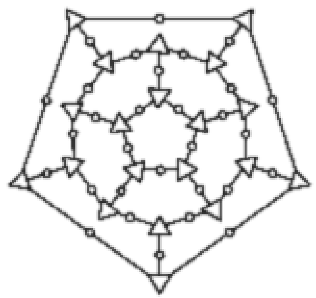
  $$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

  But $\{\boldsymbol{t} : \boldsymbol{H}\boldsymbol{t} = \boldsymbol{0}\}$.

- (7,4)-Hamming Code



- (30,11)-Hamming Code



Howmany bit errors are corrected

- Example: Can (14,8)-Hamming Code two errors?
- Count the error patterns:

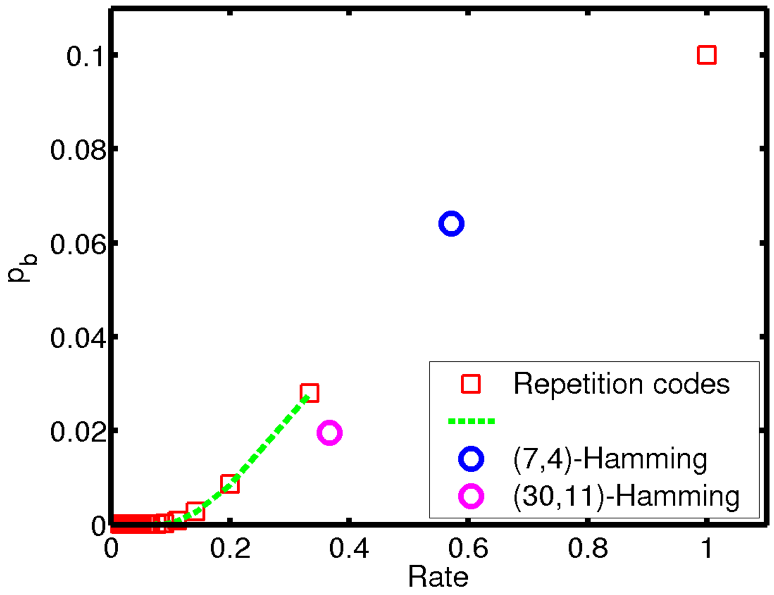
$$\binom{N}{0} + \binom{N}{1} + \binom{N}{2}$$

  for $N = 14$ there are 106 patterns.
- Every error musst give rise to one syndrom.
- For M parity pits, there are 2^M syndroms.

  For $M = 6$ this is 64.
- $\rightarrow$ The (14,8)-Hamming Code does not correct two erros.

  (The (30,11) does)

Performance of codes

- 

- Which points in the plain can be achieved?

- It was thought that to get error $\rightarrow 0$ the rate $\rightarrow 0$.

Noisy Channels - Error correcting Codes

Repetition Codes - Block Codes - Channel capacity

- > *Noisy-Channel Coding Theorem:*
  >
  > $\forall\ \epsilon > 0$ and $R < C$, there exists a code of sufficiently large length N, with rate $\geq R$ and block error $< \epsilon$.

- The *capacity* of channel Q is

$$C(Q) = \max_{p(X)} \{I(X;Y)\}$$

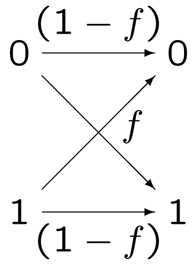
It is maximized by some optimal input distribution $p^*(X)$.

- Proof outline
  - Average block error of *all* random codes.
  - Jointly typical sequences:

$$\left| \frac{1}{N} \log \left\{ \frac{1}{p(\boldsymbol{x}, \boldsymbol{y})} \right\} - H(X,Y) \right| < \beta$$

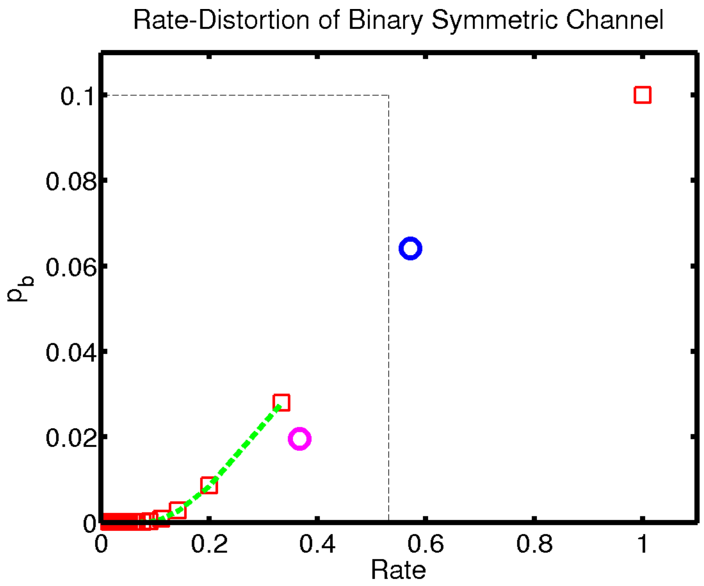
probability of $\boldsymbol{x}, \boldsymbol{y}$ being jointly typical $\to 1$ for $N \to \infty$.

- Example: binary symmetric channel with $f = 1/10$:



$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum_{i \in \{0,1\}} p(x=i) H(Y|x=i) \\
&= H(Y) - \left[f \log(\frac{1}{f}) + (1-f) \log(\frac{1}{1-f}) \right] \\
&\leq 1 - \left[f \log(\frac{1}{f}) + (1-f) \log(\frac{1}{1-f}) \right] \\
&\to C(bsc) = 0.5310
\end{aligned}$$

Rate-distortion Theory

- 

- Communication with error above C.

- Noiseless Channel

$$0 \longrightarrow 0$$

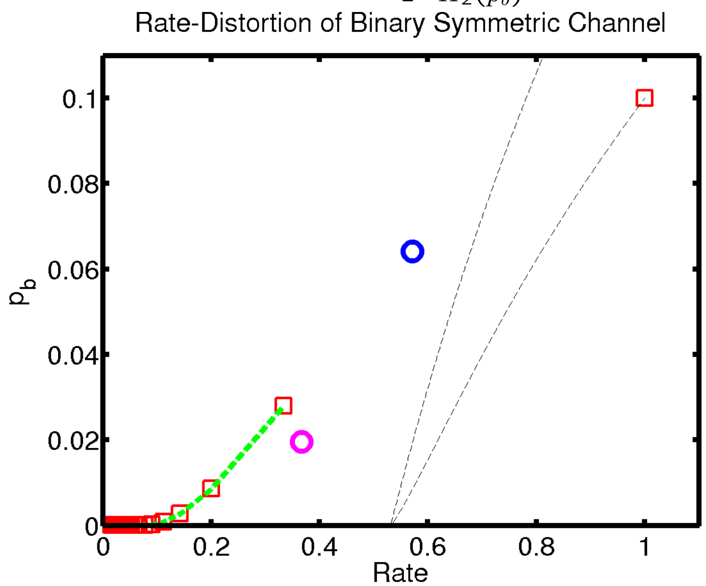
$$1 \longrightarrow 1$$

- $C = 1$ bit per channel use.

- Force communication at $R > C$.

- How to achieve the smallest possible p_b? $\rightarrow$ Communicate only $\frac{1}{R}$ and let receiver guess the missing fraction $(1 - \frac{1}{R})$.
$\rightarrow p_b = \frac{1}{2}(1 - \frac{1}{R})$

- Shannon's limit $R = \frac{C}{1-H_2(p_b)}$.



Conclusion

- Repetition Codes
- Hamming Codes
  Linear, parity checking
- Channel coding Theorem
- Rate-distortion Theory