united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR.1347 - 4

WORKSHOP ON STATISTICAL PHYSICS AND CAPACITY-APPROACHING CODES

GALLAGER'S CODES AND THEIR DESCENDENTS

D. MacKay
Department of Physics
University of Cambridge
Cavendish Laboratory
Madingley Road
Cambridge CB3 0HE, U.K.

Please note: These are preliminary notes intended for internal distribution only.

Gallager's Codes and their descendants

David MacKay

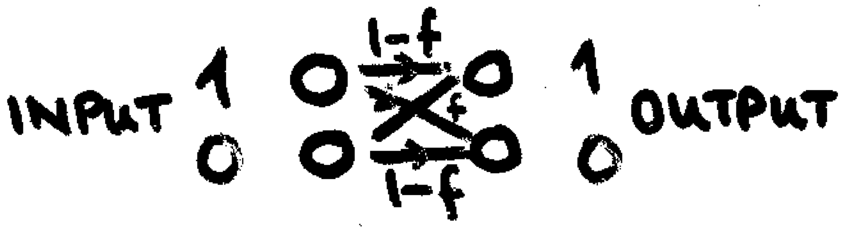
Dept of Physics
University of Cambridge

- LOW-DENSITY PARITY-CHECK CODES
- PRACTICAL DECODING
- THEORY
  - optimal decoder
  - message-passing decoder
- OTHER CODES ON GRAPHS

INFORMATION THEORY + CODING THEORY

Aim: virtually error-free communication over noisy channels

eg: binary symmetric channel



eg $f = 0.1$

Method:

Adds redundancy

Adds noise $\underline{n}$

Good news: Can reduce error probability

Bad news: Redundancy means rate R is reduced

Repetition Code "R3"

s ENCODER t CHANNEL r DECODER $\hat{\mathbf{s}}$

$f = 7.5\%$

The decoder takes the majority vote of the three signals.

Good news: only 1.6% of decoded bits are in error

Bad news: rate of communication reduced to 1/3

Example Encoder:

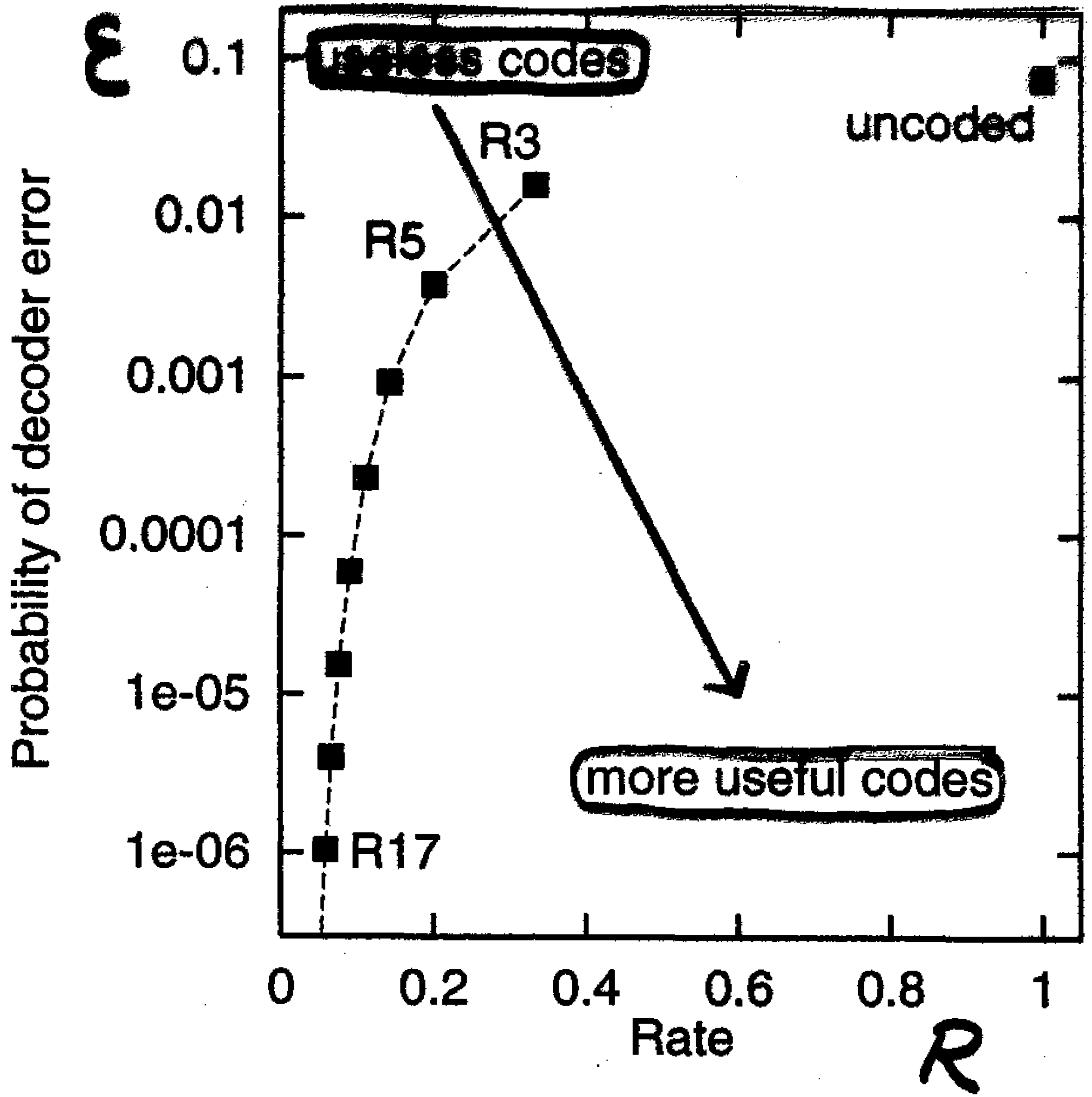
Repetition code R_3

$\underline{s}$	$\underline{t}$
0	000
1	111

Decoder : Majority vote

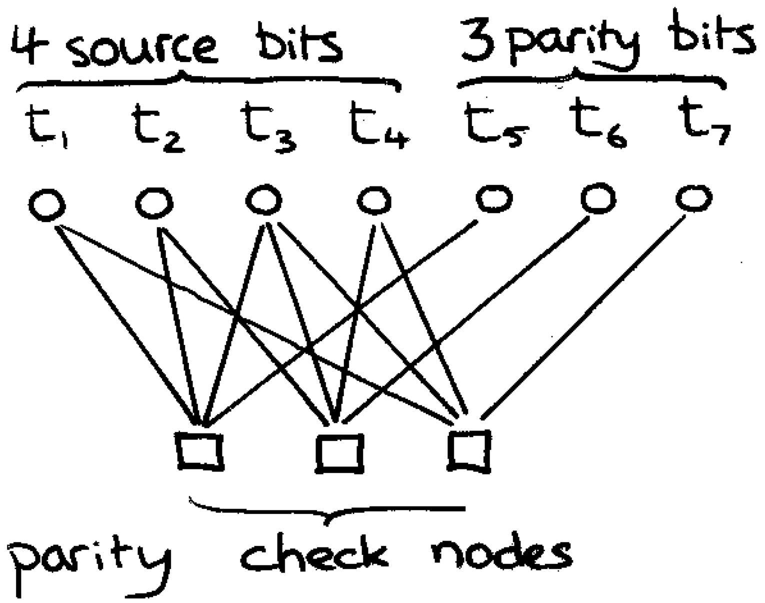
$\underline{r}$	$\hat{\underline{s}}$
000	0
001	0
010	0
100	0
110	1
101	1
011	1
111	1

$$(N, K) = (3, 1)$$


ε
Probability of decoder error
0.1
0.01
0.001
0.0001
1e-05
1e-06
useless codes
R3
R5
R17
uncoded
more useful codes
0
0.2
0.4
0.6
0.8
1
Rate
R

THE (7,4) HAMMING CODE

Graph:



Equations:

$\square \qquad t_5 = t_1 + t_2 + t_3 \mod 2$

$\square \qquad t_6 = t_2 + t_3 + t_4 \mod 2$

$\square \qquad t_7 = t_1 + t_3 + t_4 \mod 2$

Parity check matrix: $N = 7$

$M = 3$

$$\begin{array}{c} \\ \underline{\underline{H}} = \end{array} \begin{array}{c} \circ\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ \square \\ \square \\ \square \end{array}$$

$$\underline{\underline{H}}\,\underline{t} = 0 \mod 2$$

PARITY CHECK MATRIX

$$\underline{\underline{H}} = \begin{array}{c} \left[\begin{array}{cccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \\ \underbrace{\hphantom{1\;1\;1\;0}}_{\substack{K=4 \\ \text{source bits}}} \quad \underbrace{\hphantom{1\;0\;0}}_{\substack{M=3 \\ \text{parity bits}}} \end{array} \begin{array}{l} \uparrow \\ M=3 \text{ constraints} \\ \downarrow \end{array}$$

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

GENERATOR MATRIX

$$\underline{\underline{G}}^T = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\\1\end{array}}\right\} K=4 \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} M=3 \end{array}$$

$$\underline{t} = \underline{\underline{G}}^T \underline{s}$$

Hamming (7,4) code

Encodes $K=4$ source bits into $N=7$ transmitted bits.

PARITY CHECK MATRIX

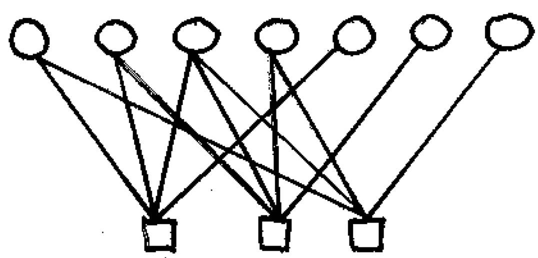
$$\underline{\underline{H}} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \updownarrow M=3 \text{ constraints}$$

$\leftarrow K=4 \rightarrow$ source bits

$\underbrace{\qquad\qquad}_{M=3}$ parity bits

GRAPH

7 transmitted bits



3 constraints

Decoding the Hamming code

received = transmitted + noise

$$\underline{r} = \underline{t} + \underline{n} \mod 2$$

$$\underline{\underline{H}}\,\underline{r} = \underbrace{\underline{\underline{H}}\,\underline{t}} + \underline{\underline{H}}\,\underline{n} \mod 2$$

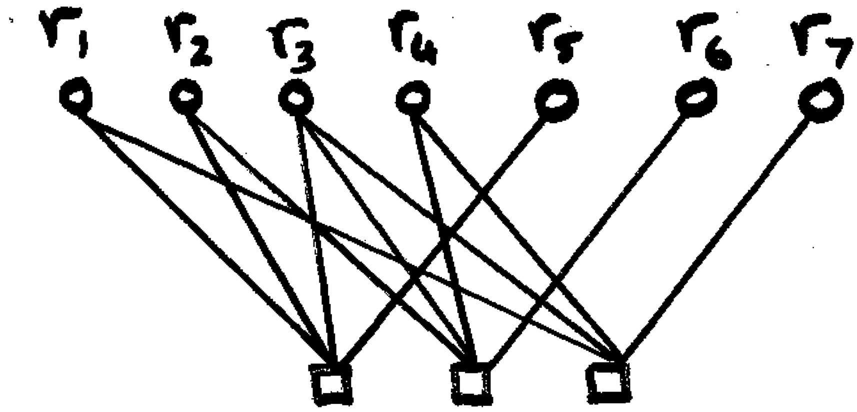
$\underline{\underline{H}}\,\underline{t} = 0$ for all valid transmissions

The "syndrome" of the received signal,

$$\underline{z} \equiv \underline{\underline{H}}\,\underline{r} \mod 2$$

$\Rightarrow$ want to solve for the sparsest $\underline{n}$ such that $\underline{\underline{H}}\,\underline{n} = \underline{z} \mod 2$

Decoder for Hamming Code



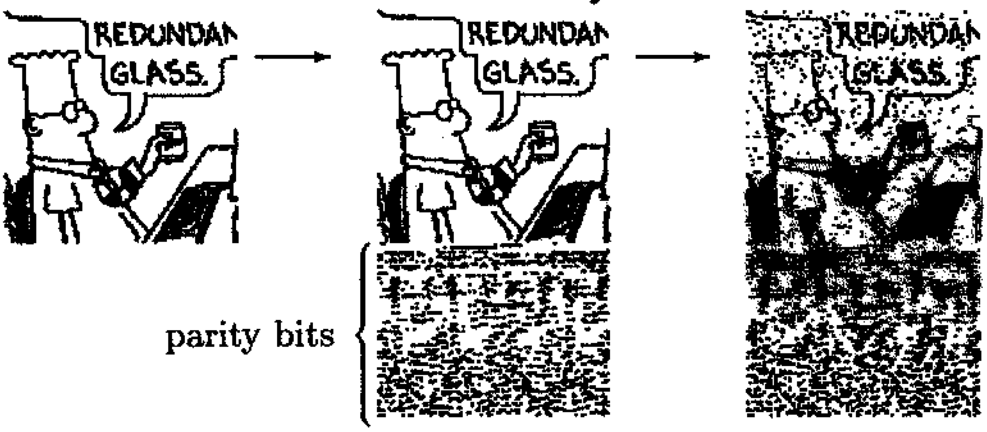
Syndrome: z_1	z_2	z_3	Best guess:
0	0	0	noise = 0
0	0	1	flip r_7
0	1	0	flip r_6
0	1	1	flip r_4
1	0	0	flip r_5
1	0	1	flip r_1
1	1	0	flip r_2
1	1	1	flip r_3

THE Hamming (7,4) code in action

s	t	s	t	s	t	s	t
0000	0000 000	0100	0100 110	1000	1000 101	1100	1100 011
0001	0001 011	0101	0101 101	1001	1001 110	1101	1101 000
0010	0010 111	0110	0110 001	1010	1010 010	1110	1110 100
0011	0011 100	0111	0111 010	1011	1011 001	1111	1111 111

$\mathbf{s}$ ENCODER $\mathbf{t}$ CHANNEL $\mathbf{r}$ DECODER $\hat{\mathbf{s}}$

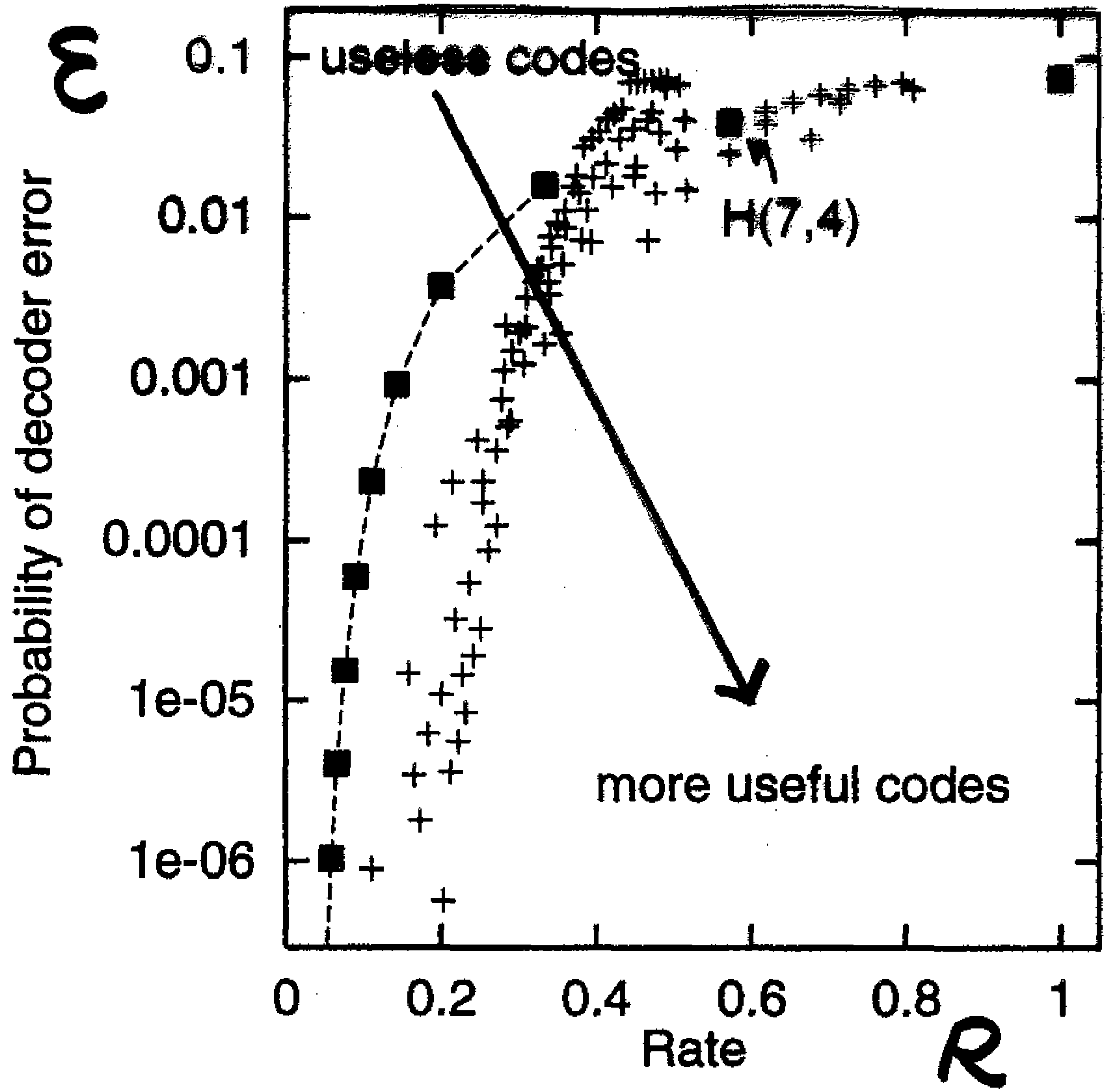
$f = 7.5\%$



The decoder picks the $\hat{\mathbf{s}}$ with maximum likelihood.

4% of decoded bits are in error
rate of communication is 4/7

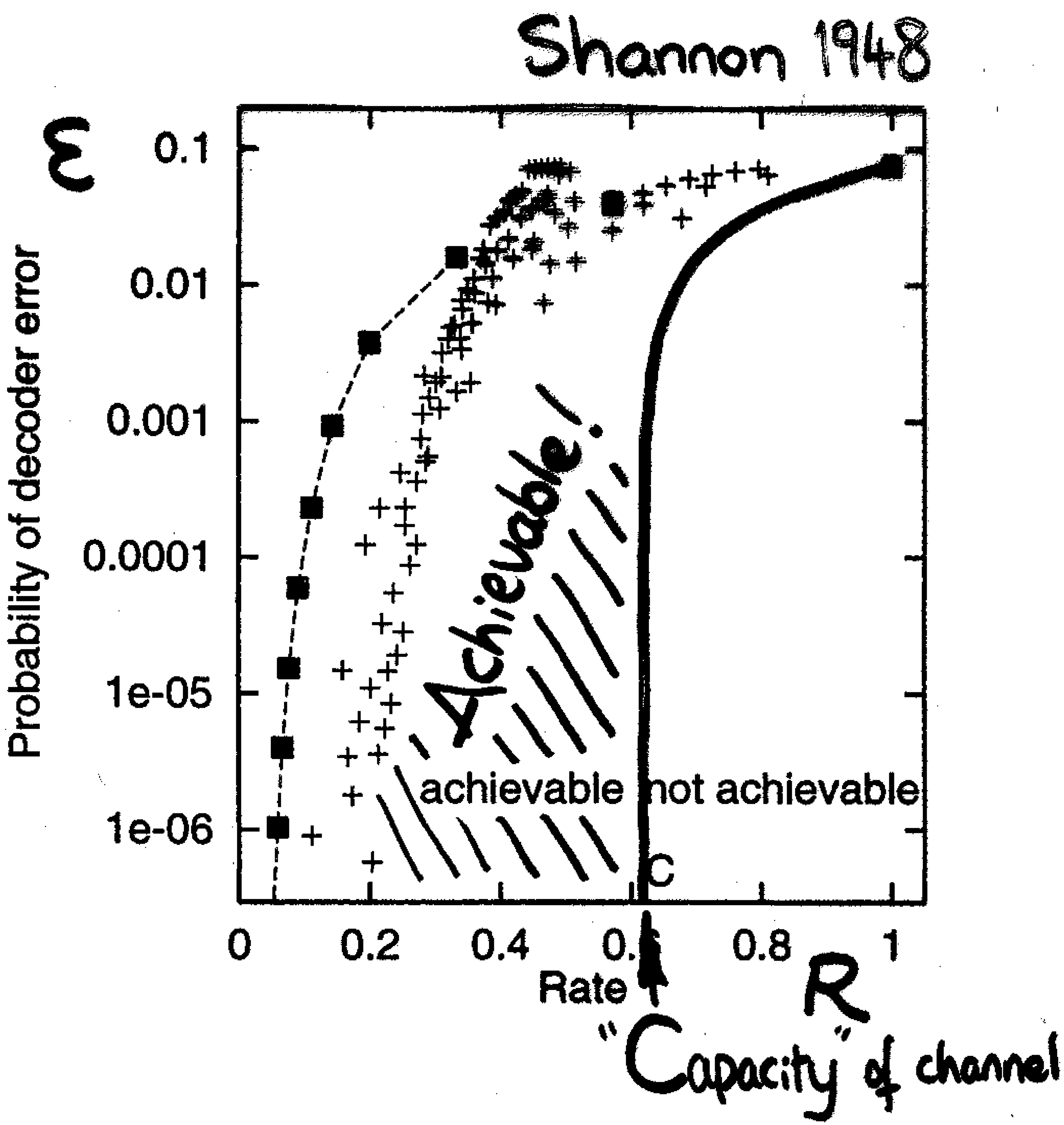
What is achievable?


ε
Probability of decoder error
0.1
0.01
0.001
0.0001
1e-05
1e-06
useless codes
H(7,4)
more useful codes
0
0.2
0.4
0.6
0.8
1
Rate
R

Shannon's proof
was non-constructive

- did not give
  practical encoder
  or decoder.

How to do it?

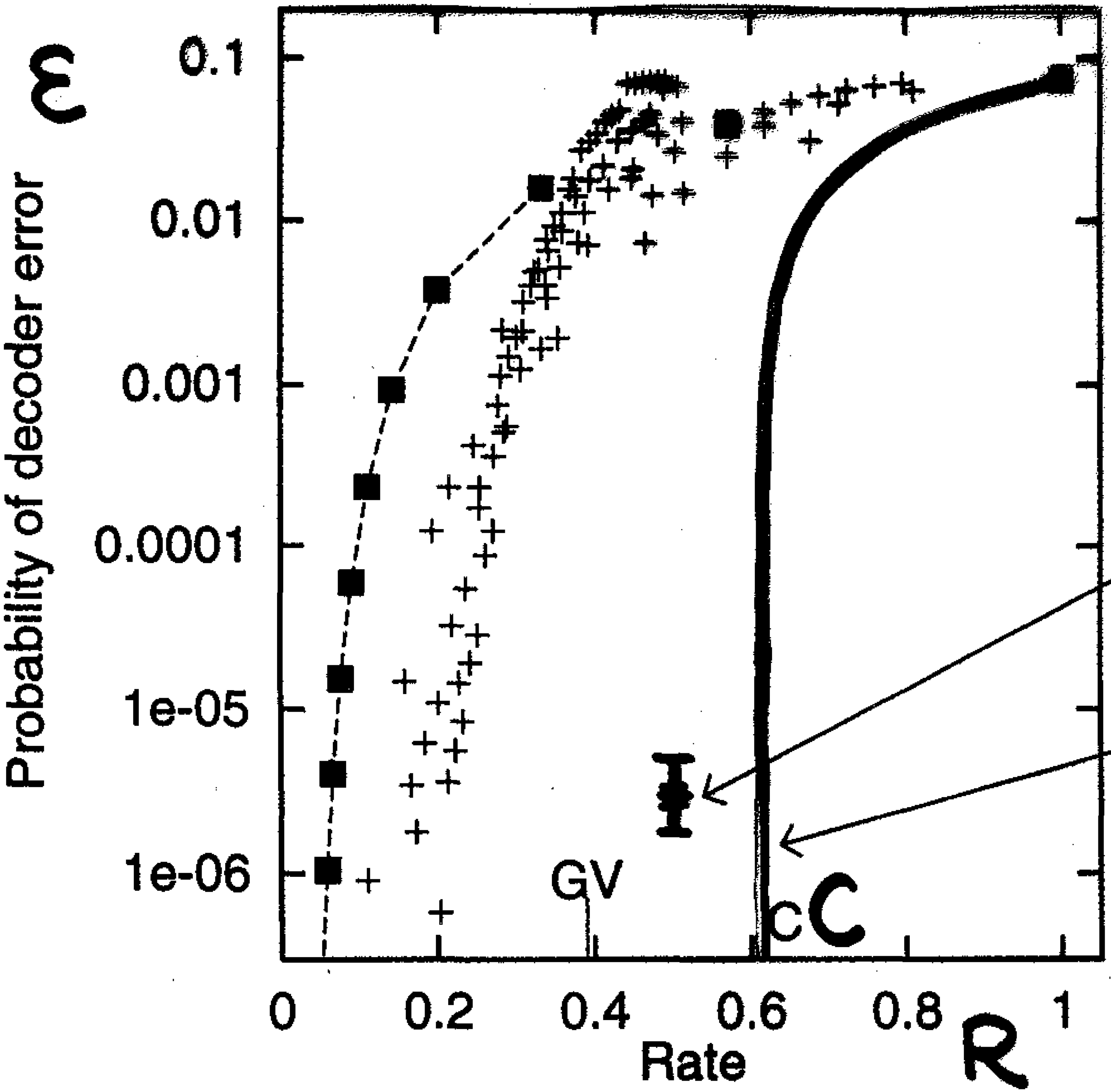

Shannon 1948
ε
Probability of decoder error
0.1
0.01
0.001
0.0001
1e-05
1e-06
Achievable!
achievable
not achievable
C
0
0.2
0.4
0.8
1
Rate
R
"Capacity" of channel

Binary symmetric channel

Capacity $\quad C = 1 - H_2(f)$

$$\left[H_2(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x} \right]$$

f = 7.5%



Our low density parity check code

Shannon limit

17

From the Hamming Code to the Low-Density Parity-check Code (Gallager '62)

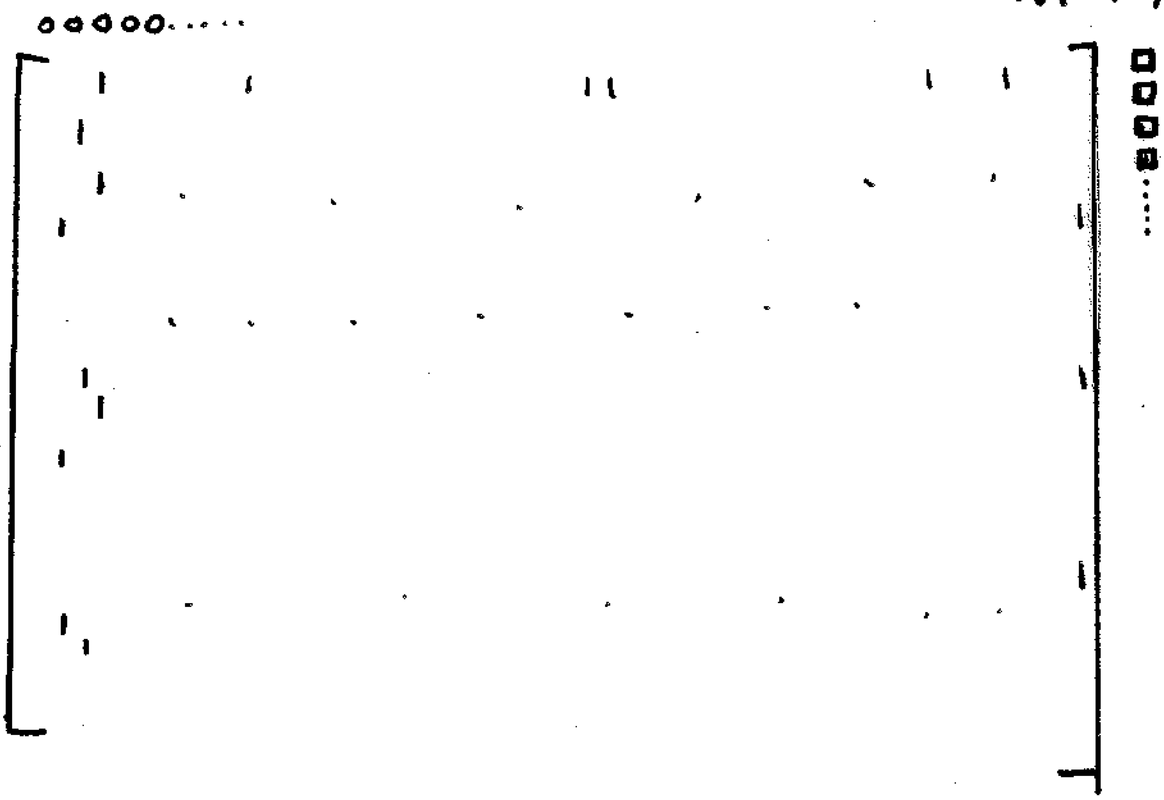
$N=7$

$M=3$

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

eg $N=20{,}000$

$M=10{,}000$

$H=$ 

All columns have weight ~~t~~ j, eg $j=3$

eg $t=3$

LOW DENSITY PARITY CHECK CODES

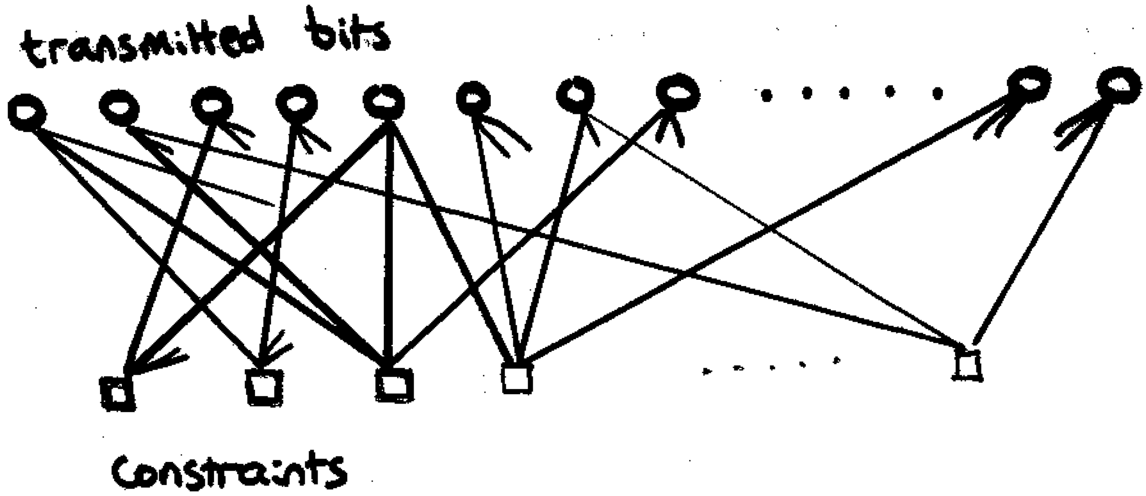
GALLAGER 1962

- Parity check matrix $\underline{\underline{H}}$ is very sparse

$$
\mathbf{H} = \begin{bmatrix}
 & & 1 & & & & & 1 & & 1 & & & 1 & & & \\
 & & & 1 & & & 1 & & 1 & & & & 1 & & & \\
 & 1 & & & 1 & & 1 & & & 1 & & & & & & \\
 & & & 1 & & 1 & & & & & 1 & & & 1 & & \\
 & & 1 & & & & & & 1 & & & & & & 1 & 1 \\
1 & & & & & 1 & & & 1 & 1 & & & & & & \\
 & & & 1 & & & & 1 & & & & 1 & & & 1 & \\
 & 1 & & & & 1 & & & & & & 1 & & & & 1 \\
1 & & & & & & 1 & & & & & & & 1 & & 1 \\
 & & 1 & & 1 & & & & & & & 1 & & 1 & & \\
 & 1 & & & & & & & & & 1 & & 1 & & 1 & \\
1 & & & & 1 & & & 1 & & & 1 & & & & &
\end{bmatrix}
\quad M
$$

eg, every column has weight 3

- The graph is very sparse



TWO EQUIVALENT VIEWPOINTS FOR DECODING

Codeword decoding

$$P(\underline{t}|\underline{r}) \propto P(\underline{r}|\underline{t})\,P(\underline{t})$$

$\prod$ channel likelihoods

uniform over 2^K codewords

$$\propto \prod_{n} \psi_n(t_n) \prod_{m} \psi_m(\{t\}_m)$$

bits

checks

Syndrome decoding

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

$$P(\underline{n}|\underline{z}) \propto P(\underline{z}|\underline{n})\,P(\underline{n})$$

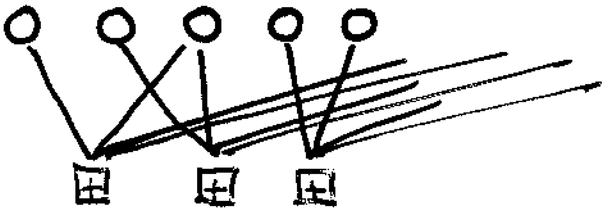
$\delta[\,\underline{z} = \underline{\underline{H}}\underline{n}\,]$

$\prod P(n_n)$

$$\propto \prod_{m} \psi_m(\{n\}_m) \prod_{n} \psi_n(\ldots)$$

$$P(\underline{t}|\underline{r}) \propto \prod_n \psi_n(t_n) \prod_m \psi_m(\{t\}_m)$$

$$\begin{cases} 1 & \Sigma t = \text{even} \\ 0 & \text{otherwise} \end{cases}$$



Replace $t_n = \begin{smallmatrix}1\\0\end{smallmatrix}$ by $x_n = \pm 1$

$$P(\underline{x}) \propto e^{\sum_n h_n x_n + \beta_1 \sum_m J_{ijkl\ldots} x_i x_j x_k x_l}$$

log likelihood ratio

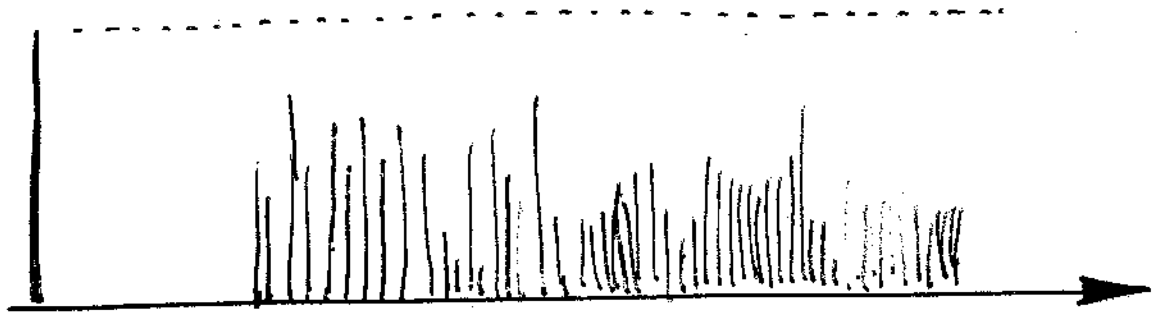
in limit $\beta_1 \to \infty$

Mezard: $$E_J = -\sum_{i_1<\cdots<i_p} J_{i_1 \cdots i_p} S_{i_1} \cdots S_{i_p}$$

Energy Landscape for decoding of good codes

$\log P(\underline{t}|\underline{r})$

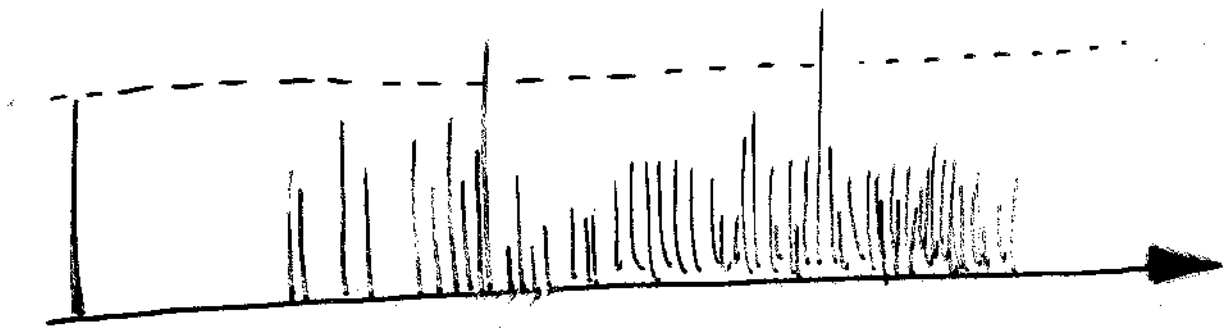
Below threshold



$\underline{t}_{TRUE}$

codewords $\underline{t}$ listed by distance from $\underline{t}_{TRUE}$.

Above threshold



$\underline{t}_{TRUE}$

PRACTICAL DECODING METHODS

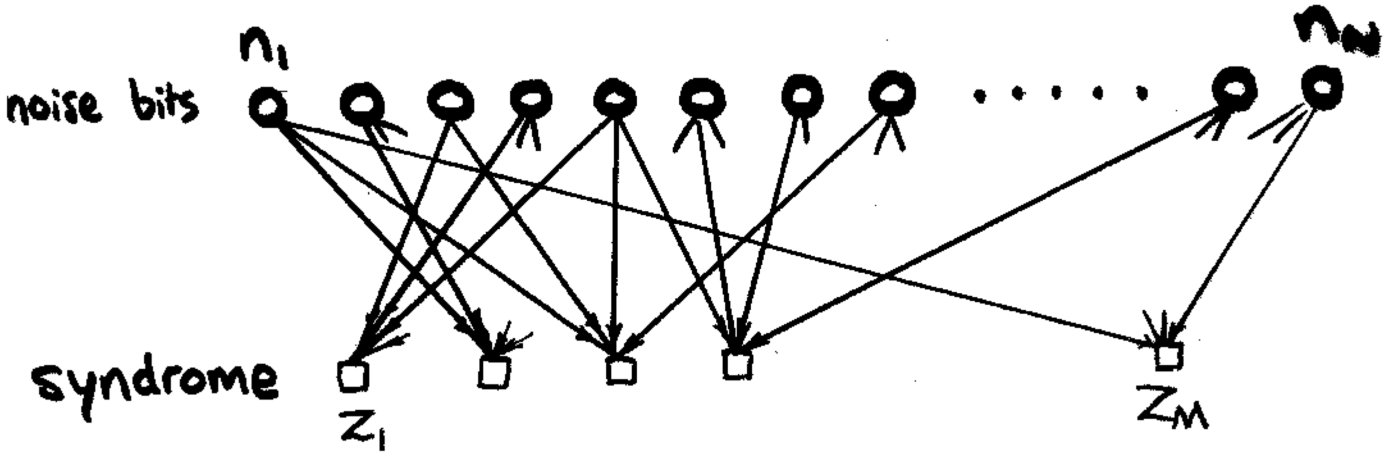
- Sum-product algorithm (aka belief propagation)

- Variational free energy minimization

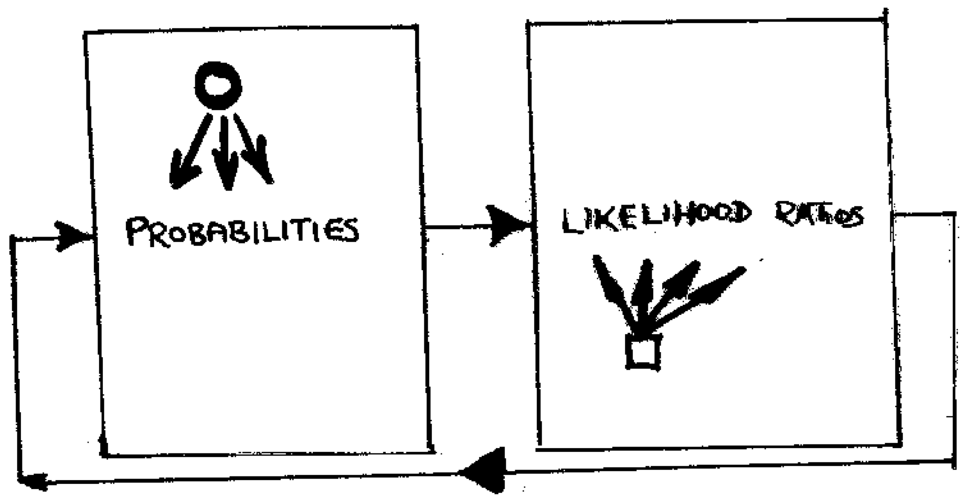
- Subcode sampling (a Monte Carlo method, Neal 2001)

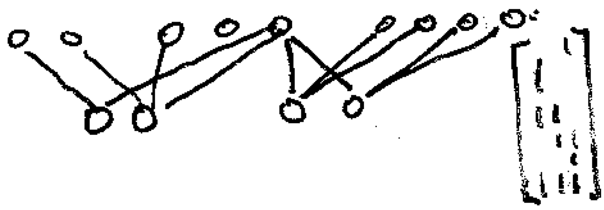
How to Solve the Decoding Problem

(Difficult)



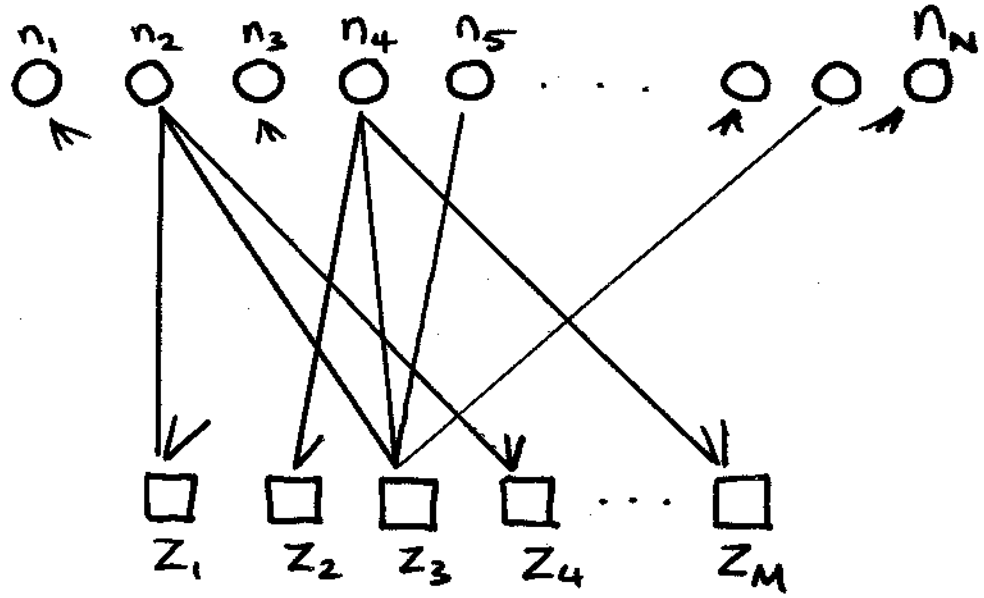
- sum-product algorithm

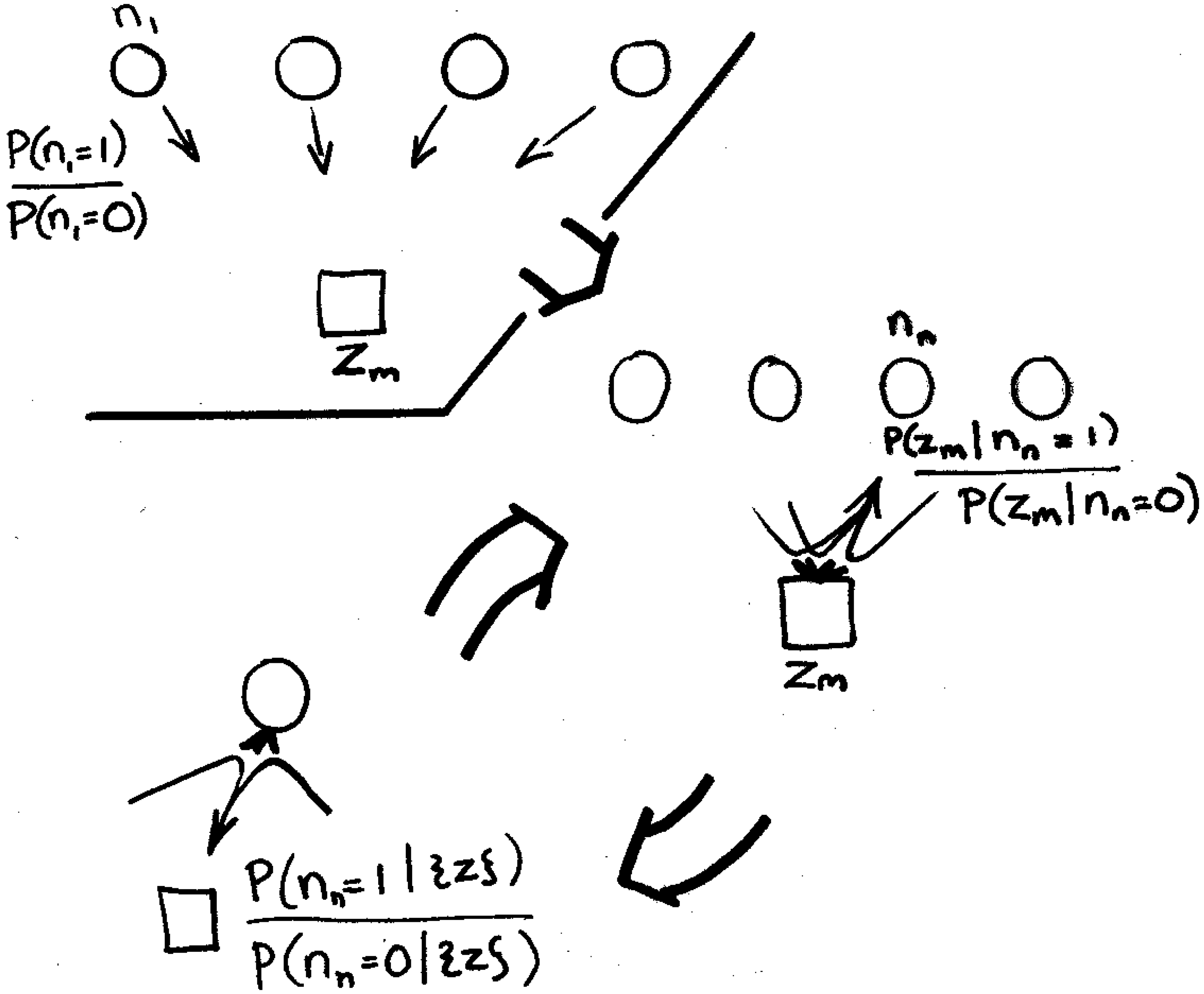




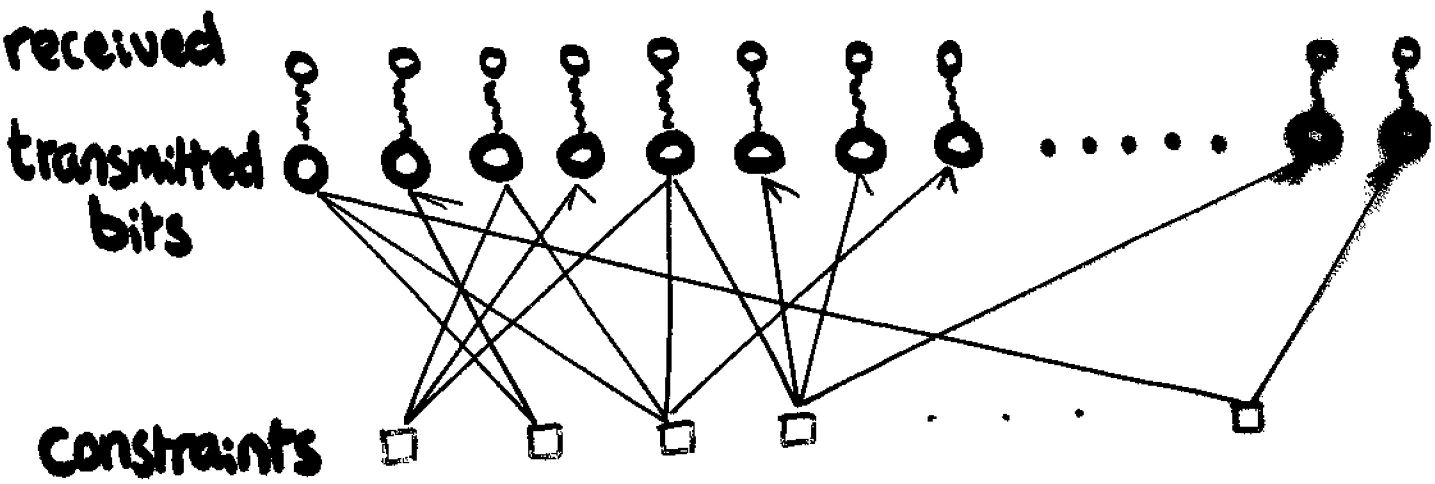
cycle:

DECODING USING THE SUM-PRODUCT ALGM

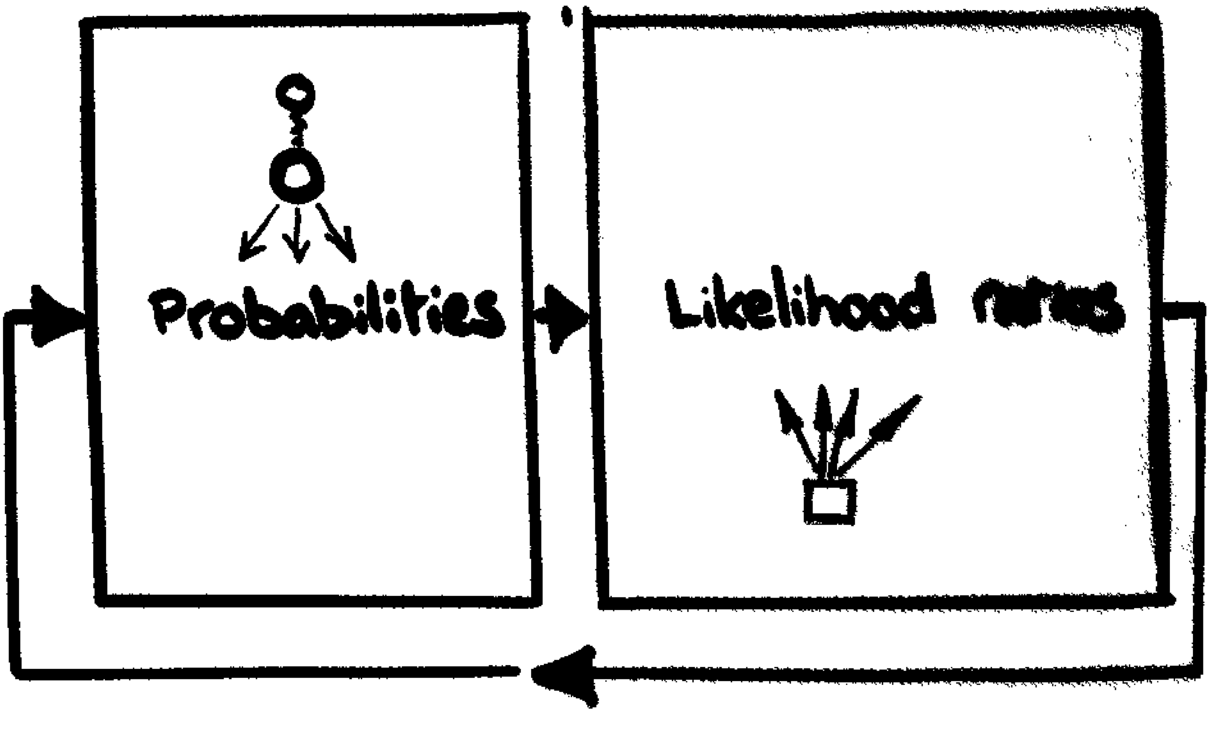




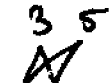
How to Solve the Decoding Problem (DIFFICULT)



- sum-product algorithm



cycle

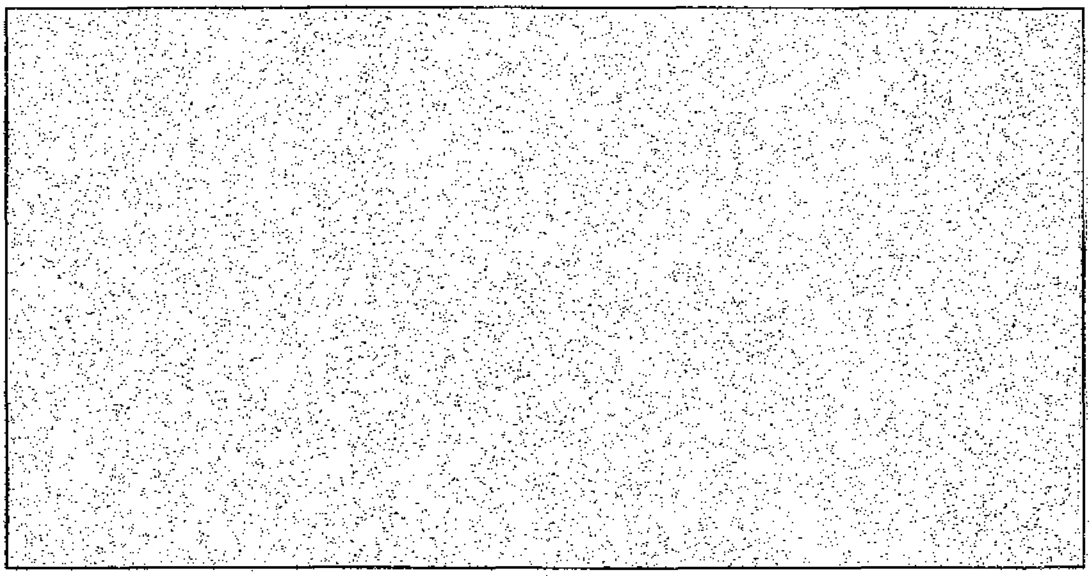


We demonstrate a large code that encodes $K = 10000$ source bits into $N = 20000$ transmitted bits.
Each parity bit depends on about 5000 source bits.
The encoder is derived from a very sparse 10000×20000 matrix $\mathbf{H}$ with three 1s per column.

TRANSMITTED:

parity bits

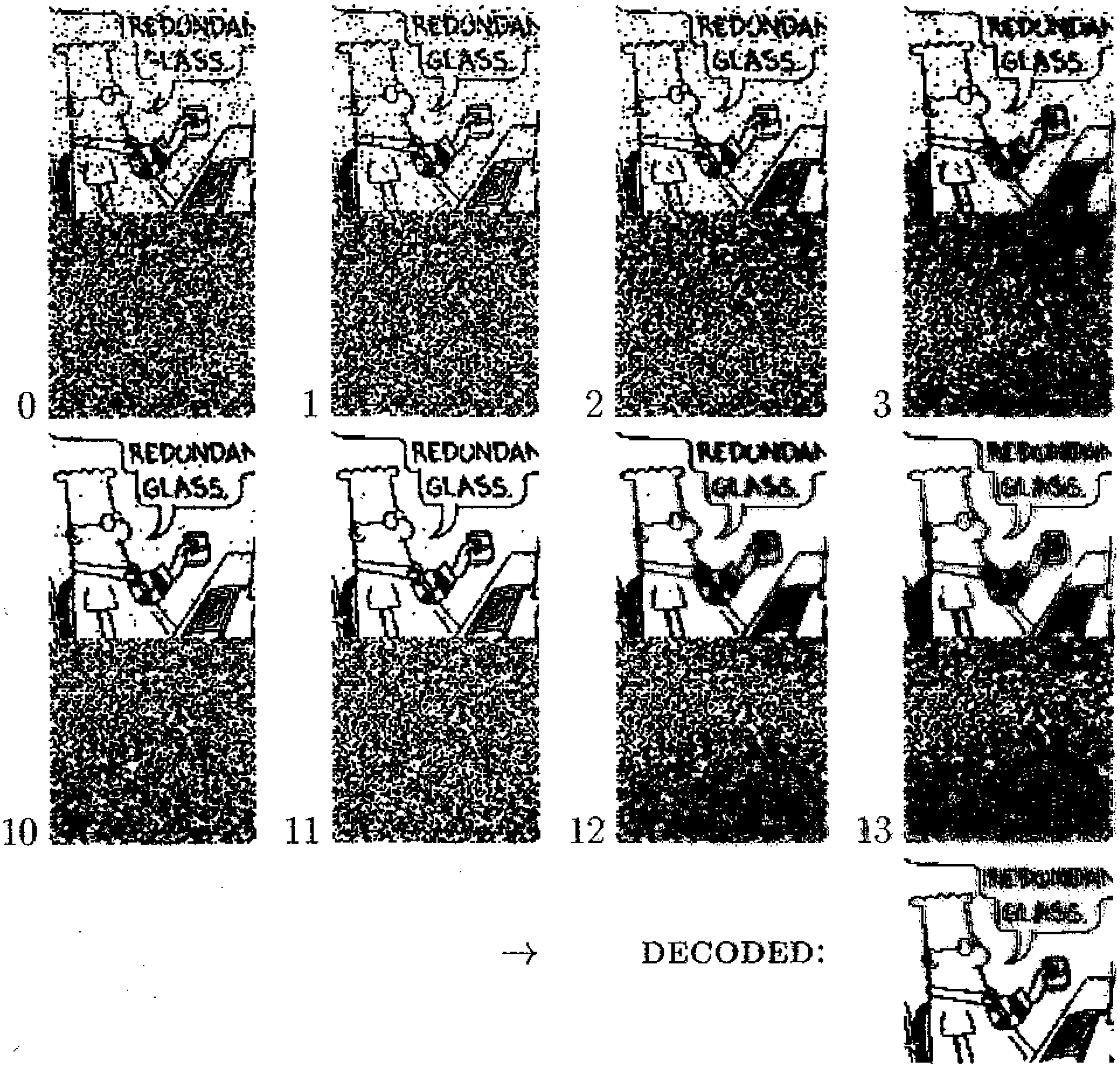
$\mathbf{H} =$



Iterative decoding

After the transmission is sent over a channel with noise level $f = 7.5\%$:

RECEIVED:



This final decoding is error free.

In the case of an unusually noisy transmission, the decoding algorithm fails to find a valid decoding. For this code and a channel with $f = 7.5\%$, such failures happen about once in every 100,000 transmissions.

ERRORS MADE BY Gallager CODES

Each iteration the best guess is checked — if $H\hat{\underline{n}} = \underline{z}$, halt.

Two potential failure modes:

1) Undetected errors
   — when decoder halts in a nearby codeword.

2) Detected errors
   — decoder reaches some max. number of iterations.

ALL* ERRORS ARE DETECTED ERRORS.

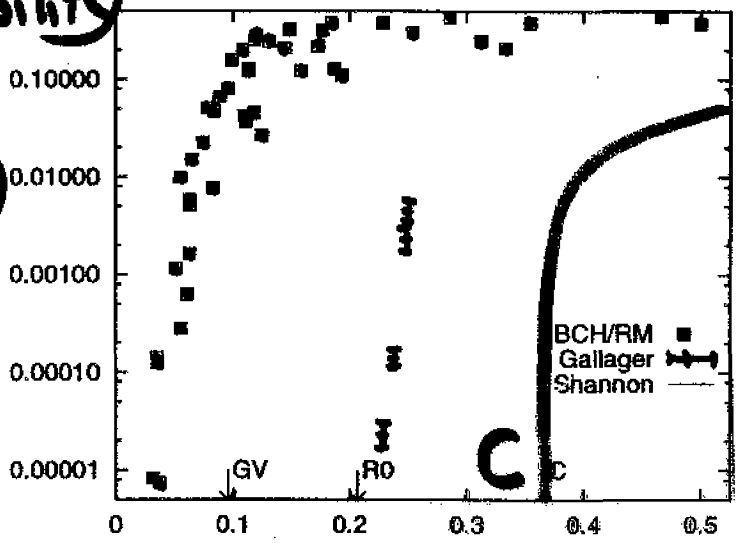
Gallager codes do not have low-weight codewords

* in 10^8 blocks of experiments, for all codes with $N > 400$, $R \in (\frac{1}{4}, \frac{2}{3})$

Block lengths N ≃ 13,000

B.S.C. $f_n = 0.16$

Decoder error probability (log scale)



C

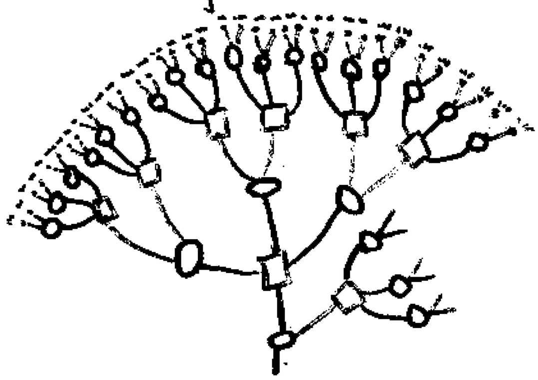
R

THEORETICAL QUESTIONS

- How well would the optimal decoder work?

  & What are the code's distance properties?

- What happens when we run sum-product on an infinite tree?



& are finite trees related? with cycles

Finding optimal decoder performance

- Typical set method
- Gallager's method
- Replica method

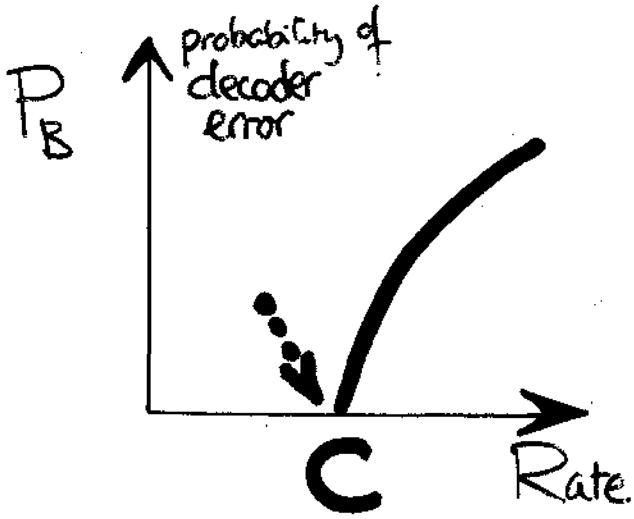
Code families

Very Good codes

can achieve

$P_B \to 0$

as $R \to C$

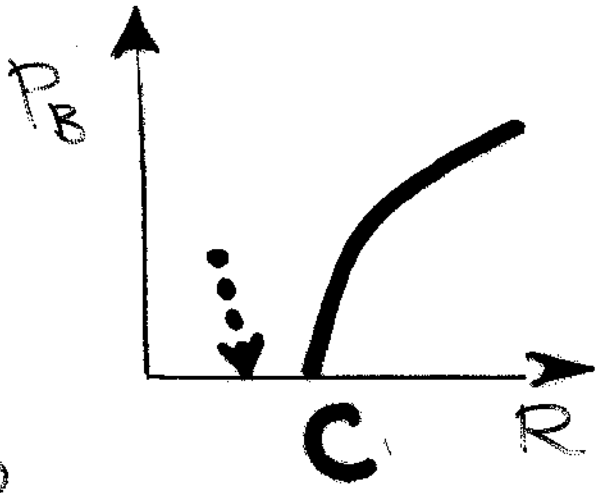


Good codes

can achieve

$P_B \to 0$

at some $R > 0$

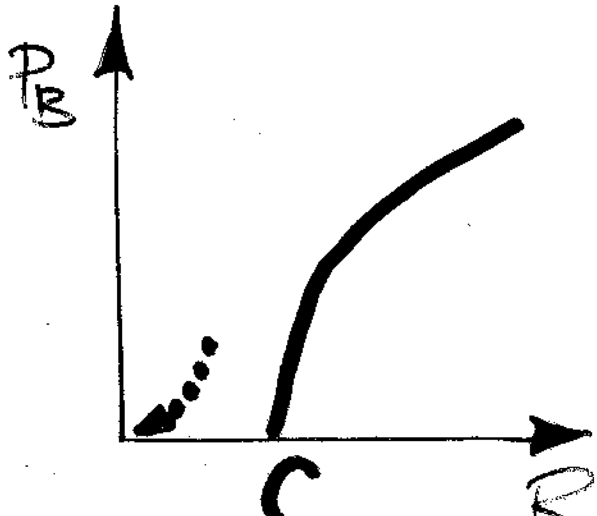


Bad codes

only achieve

$P_B \to 0$

as $R \to 0$



Codes have <u>Good distance</u>

if $\frac{d}{N} \to$ const. as $N \to \infty$

Codes have <u>Very Good distance</u>

if $\frac{d}{N} \to \frac{d_{GV}}{N}$

$$H_2\left(\frac{d_{GV}}{N}\right) = 1 - R$$

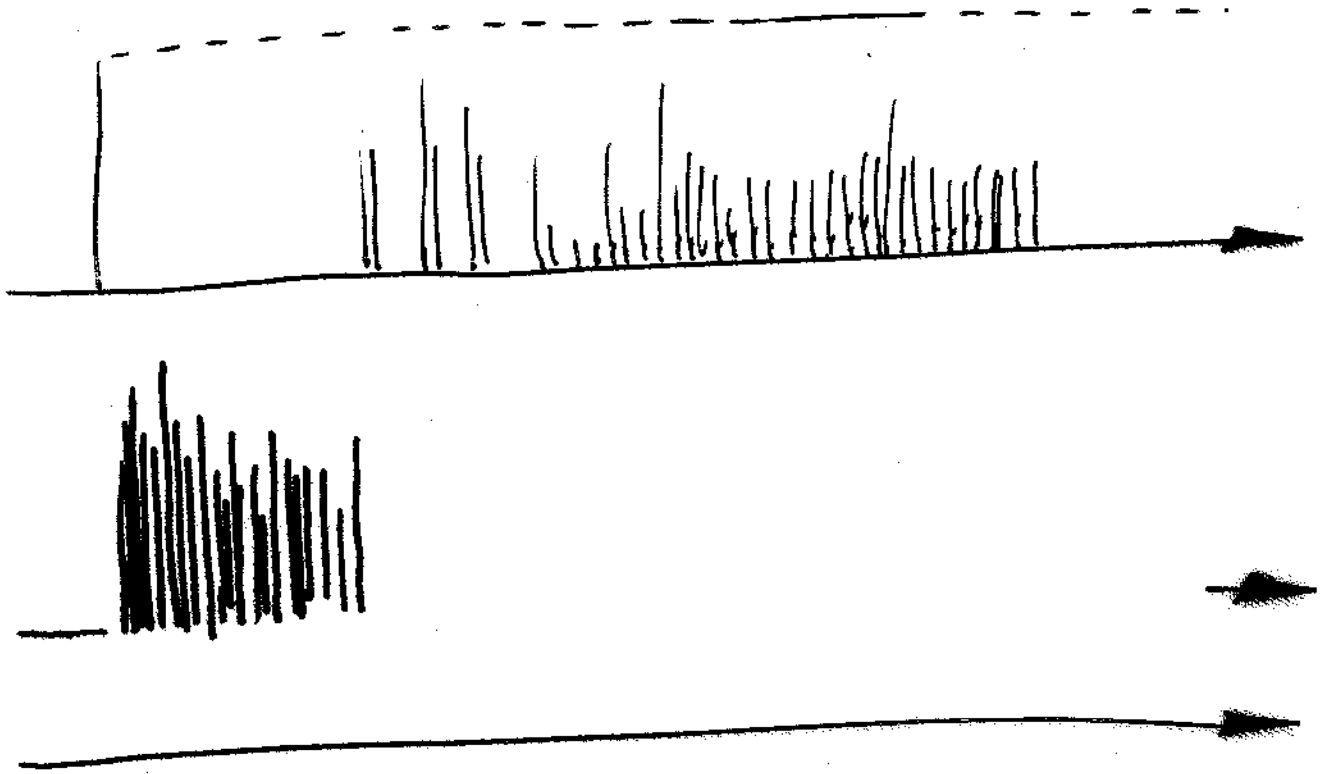
Codes have <u>Bad distance</u>

if $\frac{d}{N} \to 0$

& <u>very bad distance</u>

if $d =$ const.

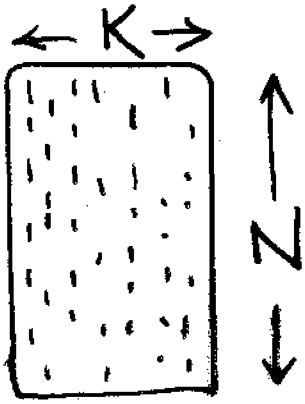
Good codes vs. bad codes



Low-density Generator-Matrix Codes are bad codes

$$\underline{t} = \underline{\underline{G}}^{T} \underline{s}$$

$$\underline{\underline{G}}^{T} =$$



If every column has weight g,
then $\exists$ at least K
codewords with weight g,

and

$$P(\text{block error}) \sim K f^{\lceil g/2 \rceil}$$

because $g/2$ flips can cause confusion about one bit.

THEORY

Gallager codes with fixed $\overset{j}{t} \geq 3$ are GOOD, and have GOOD DISTANCE.

Gallager 1962 15
MacKay 1999 25

As $\overset{j}{t}$ increases*, Gallager codes become VERY GOOD*, and have VERY GOOD DISTANCE.

* $\left(\frac{t}{M} \text{ still vanishing}\right)$

* for a wide range of channels with and without memory.

MacKay 1999

Upper bound on Optimal Decoder performance

f  $\underline{n} \to \underline{z} = \underline{\underline{H}}\,\underline{n}$

$\underline{n}$: N bits $\underline{z}$: M bits

$$R = 1 - \frac{M}{N}$$

Perfect decoding implies no uncertainty:

$$H(\underline{n}|\underline{z}) = 0.$$

$$H(\underline{n}) - H(\underline{n}|\underline{z}) = H(\underline{z}) - H(\underline{z}|\underline{n})$$

Mutual information identity

$\underline{n} \to \underline{z}$ Deterministic $\Rightarrow$ $H(\underline{z}|\underline{n}) = 0$

$\Rightarrow$ Perfect decoding implies

$$\underbrace{H(\underline{n})}_{N H_2(f)} = H(\underline{z})$$

$$H(\underline{z}) = H(z_1) + H(z_2|z_1) + H(z_3|z_1 z_2) \ldots$$

$$\leqslant H(z_1) + H(z_2) + H(z_3) \ldots$$

$$= M H_2(p_z)$$

p_z: probability that a syndrome bit is 1.

Because row-weights are finite, $p_z < 1/2$

so $H(\underline{z}) < M$ and $\underbrace{\left(1 - \frac{M}{N}\right)}_{R} < \underbrace{\left(1 - H_2(f)\right)}_{C}$

Conclusion of Upper bound argument

No codes based on finite-degree graphs can reach the Shannon limit.

The minimum distance 'd' of a code is the smallest Hamming distance between two codewords.

eg (7,4) Hamming Code has $d = 3$.

Why distance?

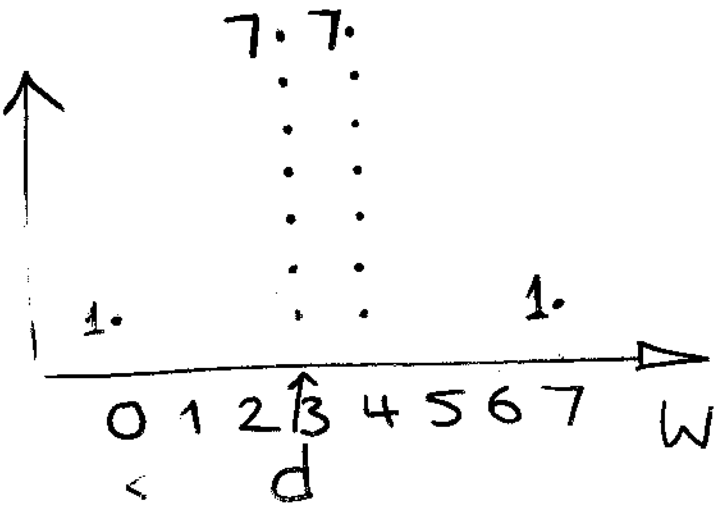
The optimal decoder can guarantee to correct up to $t = \left\lfloor \frac{d-1}{2} \right\rfloor$ errors

Distance isn't everything!

N The optimal decoder of a good code can correct many more than $d/2$ errors, with probability $\rightarrow 1$

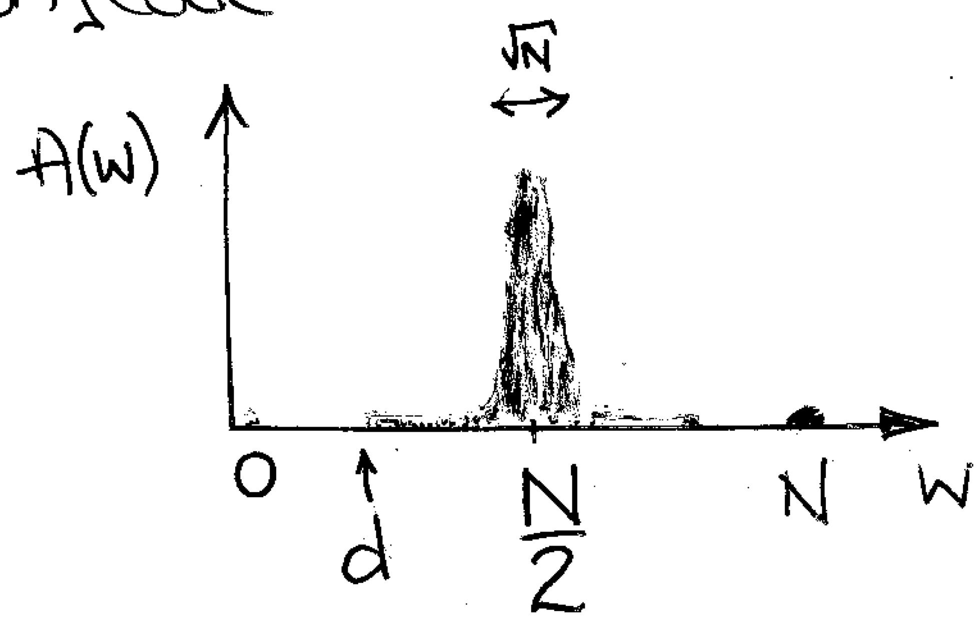
The weight enumerator function of a linear code $A(w)$ counts how many codewords have weight w.

(7,4) H:



A Random linear Code

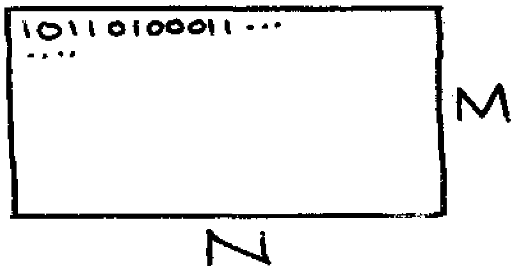
blocklength N



Optimal decoder performance depends only on $A(w)$

example – Random linear code

$\underline{\underline{H}} =$



$R \doteq 1 - \frac{M}{N}$

$$A(w) = \sum_{\underline{t}:\ \text{weight}(\underline{t}) = w} \delta(\underline{\underline{H}}\,\underline{t} = \underline{0})$$

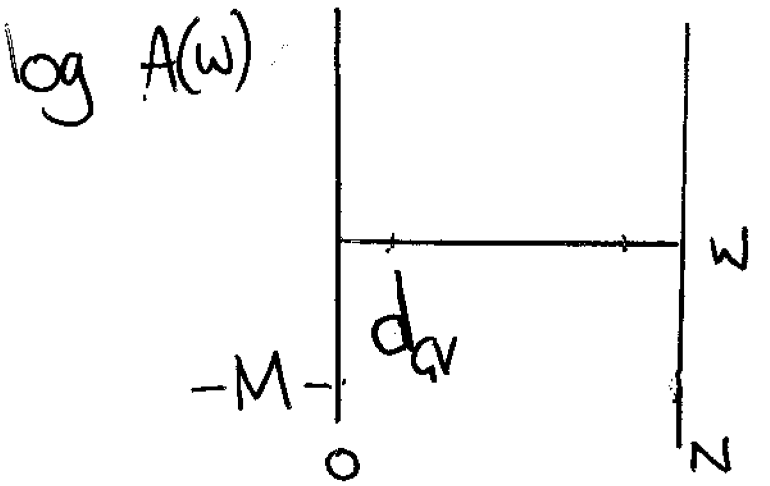
$$\langle A(w) \rangle = \sum_{\underline{t}:\ w} \langle \delta(\underline{\underline{H}}\,\underline{t} = \underline{0}) \rangle$$

$$= \binom{N}{w} \left(\frac{1}{2}\right)^{M}$$

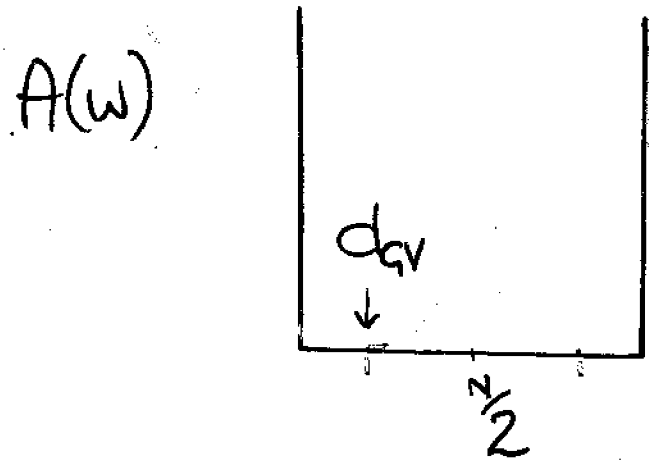
for $w > 1$

$$A(w) = \binom{N}{w} 2^{-M}$$

$$\log A(w) \simeq N H_2(w/N) - M$$



Gilbert-Varshamov distance $d_{GV}(R, N)$



$$H_2\left(\frac{d_{GV}}{N}\right) = 1 - R$$

Typical set decoder

Given a channel model, define 'T' to be those with typical probability.

For BSC, $T \simeq 2^{NH_2(f)}$

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

The typical set decoder guesses $\hat{\underline{n}}$ if $\exists$ unique $\hat{\underline{n}} \in T$ s.t. $\underline{\underline{H}}\,\underline{n} = \underline{\underline{z}}$

otherwise it declares an error.

Probability of error of Typical Set Decoder for Random Linear Codes

$$P_{TS} = P(\underline{n} \text{ is not typical}) \xrightarrow{} 0 \text{ for large } N$$

$$+ \, P(\text{another typical } \hat{\underline{n}} \text{ satisfies } \underline{\underline{H}}(\underline{n}-\hat{\underline{n}}) = \underline{0})$$

$$P_{TS} \leq \sum_{\hat{\underline{n}} \in T} \delta[\, \underline{\underline{H}}(\underline{n}-\hat{\underline{n}}) = \underline{0} \,]$$

$\underline{n}$

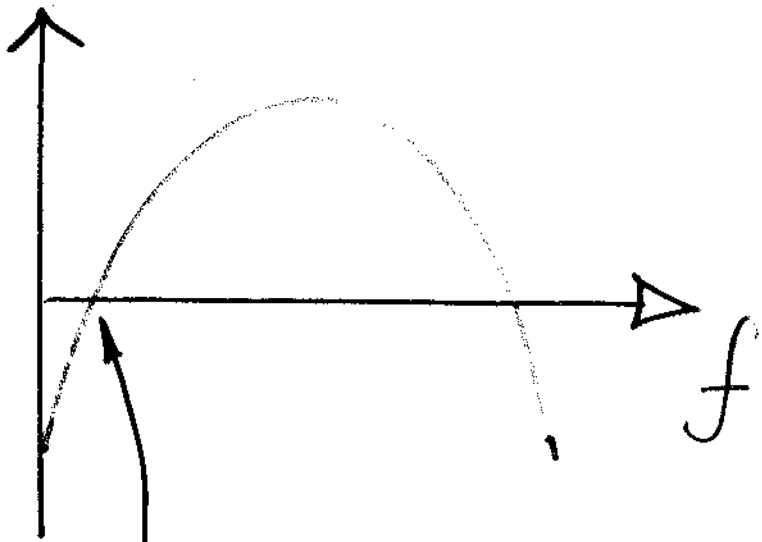
$$\langle P_{TS} \rangle \leq \Big\langle \sum_{\hat{\underline{n}} \in T} \delta[\, \underline{\underline{H}}(\underline{n}-\hat{\underline{n}}) = \underline{0} \,] \Big\rangle_{\underline{n}}$$

average over linear codes

$$= |T| \; P[\, (\underline{n}-\hat{\underline{n}}) \text{ is a codeword} \,]$$

$$= 2^{NH_2(f)} \; 2^{-M}$$

$$\log \langle P_{TS} \rangle \leq N H_2(f) - M$$



$$f_{crit} : \quad H_2(f) = \frac{M}{N}$$

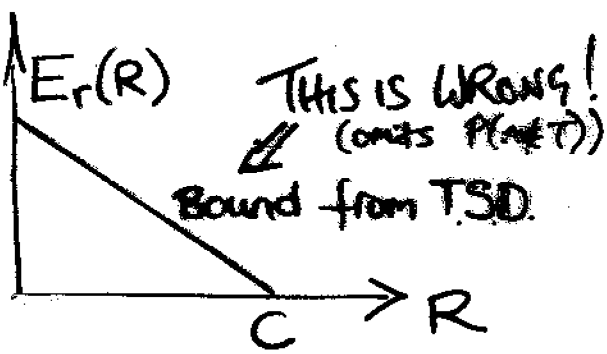
$$\rightarrow R_{crit}(f) = 1 - \frac{M}{N}$$

$$\text{Capacity} = 1 - H_2(f)$$

$$\langle P_{TS} \rangle \lesssim 2^{-N[\,C(f) - R\,]}$$

$$e^{-N\underbrace{E_r(R)}_{\text{Reliability function}}}$$

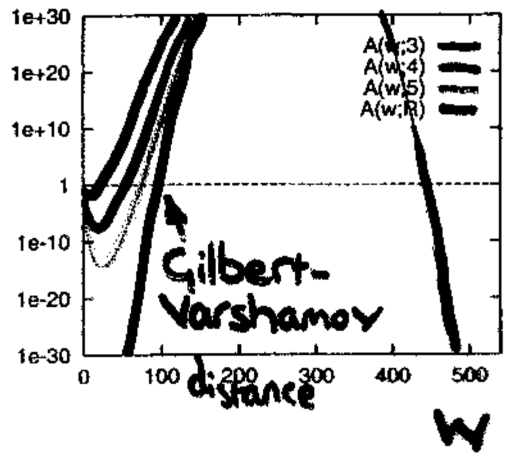
or 'error exponent'

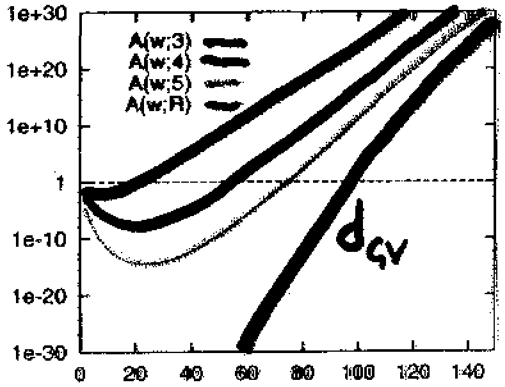


Average weight enumerator functions of random Gallager codes

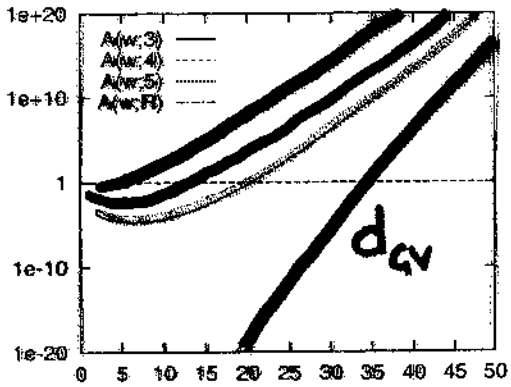
A(w)

R=$\frac{1}{3}$





1/3



R=$\frac{2}{3}$

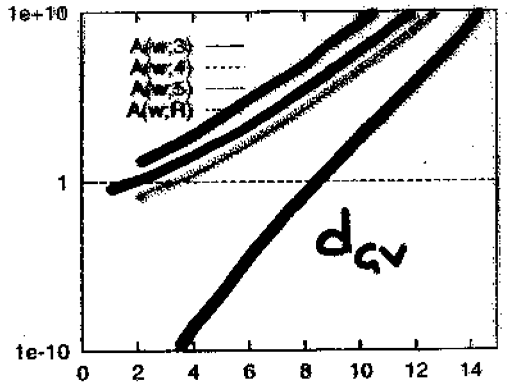
—— Gallager, j=3

—— j=4

—— j=5

—— Random linear code

2/3



R=$\frac{8}{9}$

8/9

N=540 bits

Probability of error of Typical Set Decoder

$$P_{TS} = P(\underline{n} \notin T) + P(\underline{n} \in T,\ \exists\ \underline{\hat{n}} \in T : \underline{\underline{H}}(\underline{n} - \underline{\hat{n}}) = 0)$$

$(\underline{\hat{n}} \neq \underline{n})$

$$P_{TS} \leq P(\underline{n} \notin T) + \sum_{\underline{\hat{n}} \in T} \sum_{\underline{n}} P(\underline{n})\, \delta(\underline{\underline{H}} \underbrace{(\underline{n} - \underline{\hat{n}})} = 0)$$

let this vector have weight w

\# of typical $\underline{\hat{n}}$ s.t. $|\underline{n} - \underline{\hat{n}}| = w$

$$\langle P_{TS} \rangle \simeq \sum_{w} g(w)\ P[\underline{\underline{H}} \underline{\Delta} = \underline{0}]$$

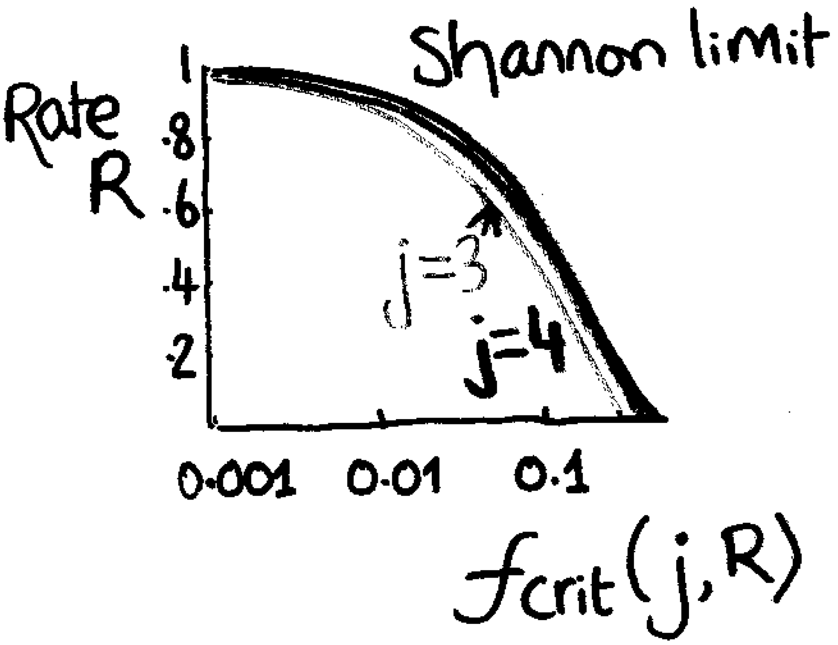
$$\frac{A(w)}{\binom{N}{w}}$$

The Bottom line

From $\langle A(w) \rangle$ we can find lower and upper bounds on the threshold $f_{crit}(j,k)$ below which $P^{error}_{TS} \to 0$

for given (j,k)

Lower bounds for ensemble M



Aside:

Compare $R = 1 - H_2\left(\frac{d_{av}}{N}\right)$

with the Shanon limit

$$C = 1 - H_2(f)$$

Conclusion:

Random linear codes can cope with noise levels f up to $f_{max} = \frac{d_{av}}{N}$

Whereas a bounded-distance decoder can only cope with

$$f_{bdd} = \frac{d/2}{N}$$

Other sparse-graph codes

- Codes over $GF(q)$
- Irregular constructions

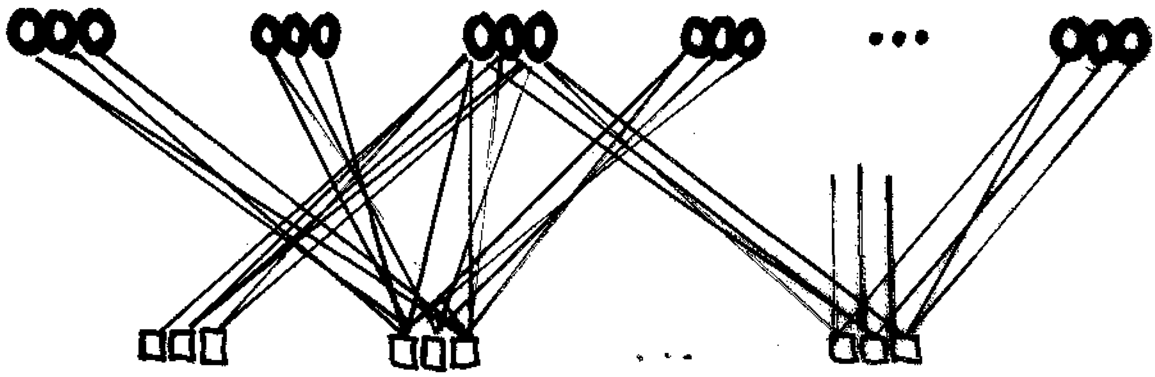
- Graphs with STATE VARIABLES
  - Turbo codes
  - Repeat-accumulate codes

- Graphs with more complex constraints
  - Tanner codes

- MN codes
  - nonlinear codes using source-redundancy

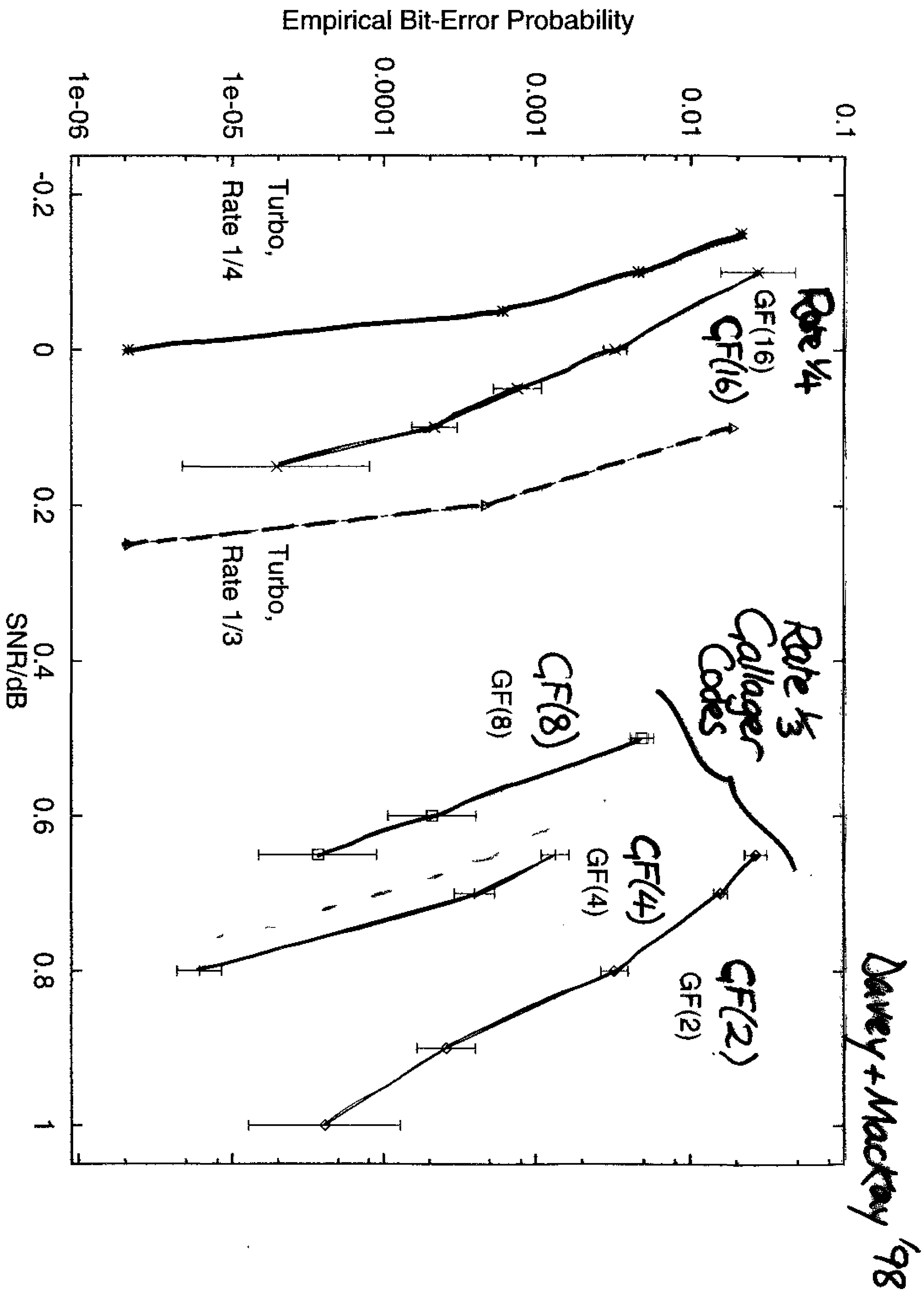
IMPROVING GALLAGER CODES I

- Clump bits together and
  track correlations during decoding



(Davey and MacKay)

Gallager codes over GF(q)

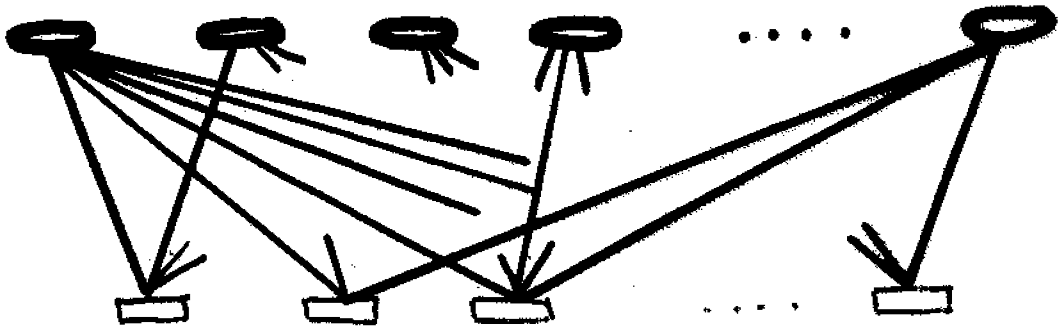


Empirical Bit-Error Probability
0.1
0.01
0.001
0.0001
1e-05
1e-06
-0.2
0
0.2
0.4
0.6
0.8
1
SNR/dB
Turbo, Rate 1/4
Turbo, Rate 1/3
GF(16)
GF(8)
GF(4)
GF(2)
Rate ¼
GF(16)
Rate ⅓
Gallager Codes
GF(8)
GF(4)
GF(2)
Davey + Mackay '98

IMPROVING GALLAGER CODES II

- Make graph <u>irregular</u>

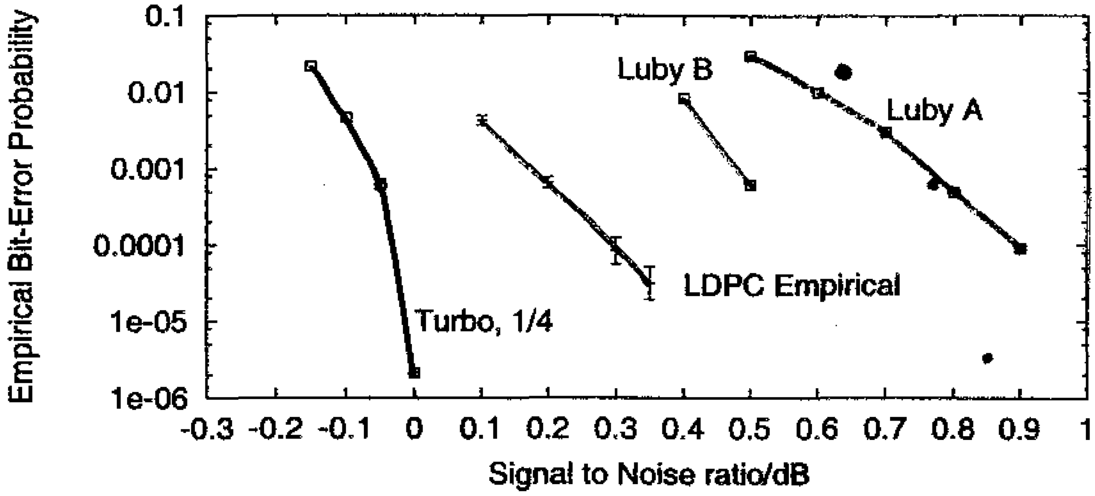
(Luby, Mitzenmacher, Shokrollahi, & Spielman ;
Davey and MacKay)



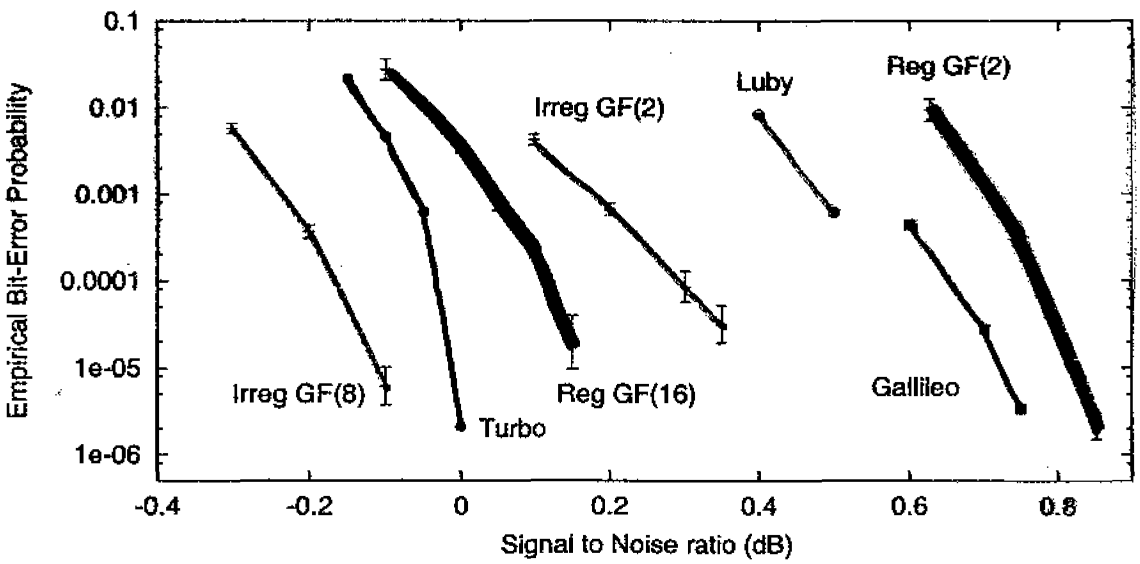
a privileged bit participates in many constraints

1999: Richardson, Shokrollahi & Urbanke

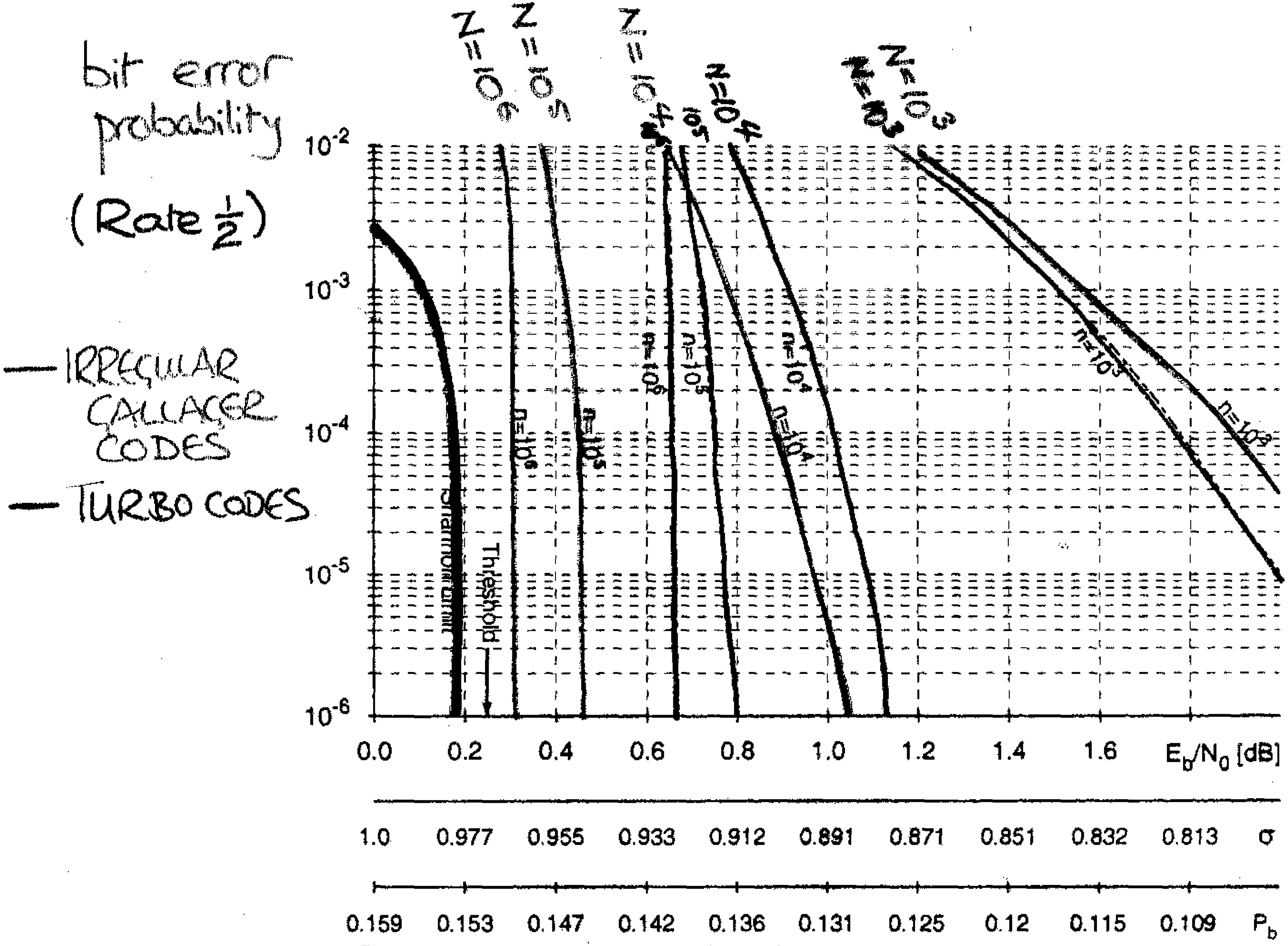
IRREGULAR BINARY GALLAGER CODES



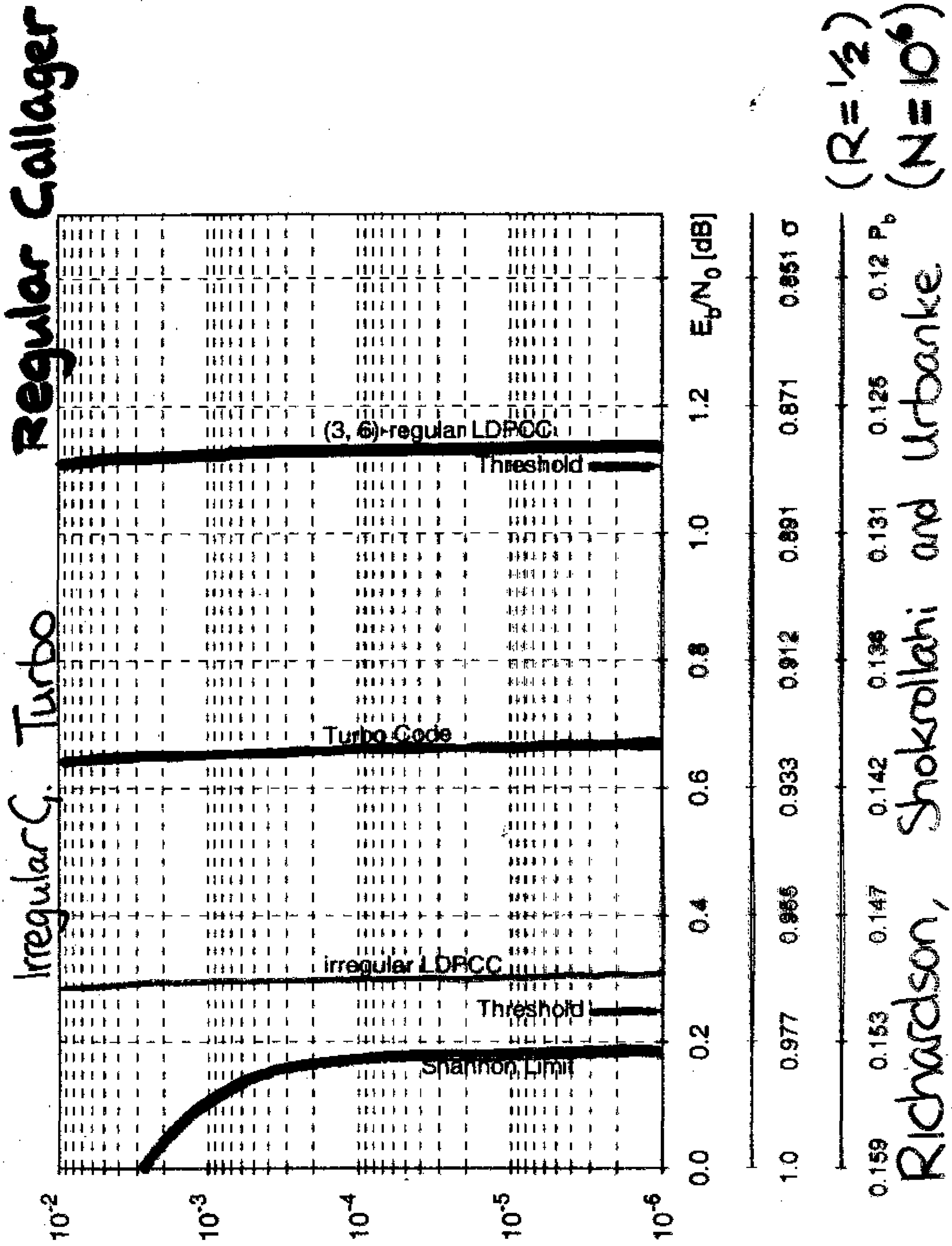
R = 1/4



s of
any
city,
that
se-
the
rror
the
y of
ugh,



Richardson, Shokrollahi and Urbanke

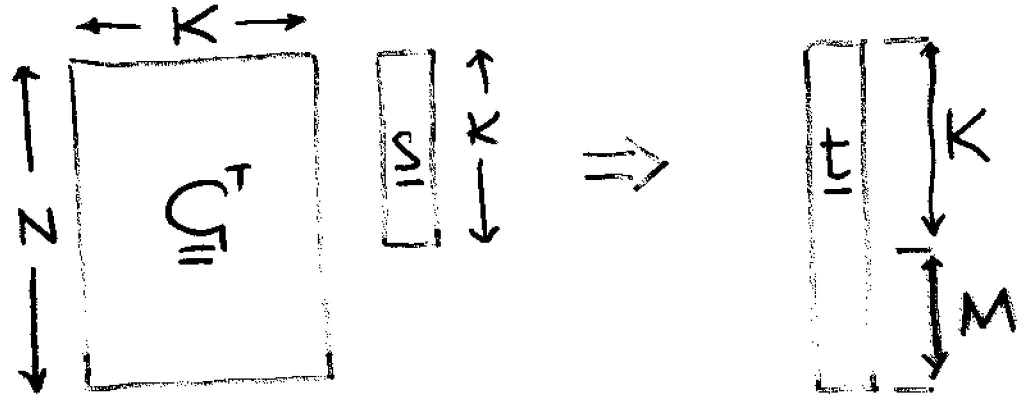


Regular Gallager
Irregular G.
Turbo
(3, 6)-regular LDPCC
Threshold
Turbo Code
irregular LDPCC
Threshold
Shannon Limit
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
0.0 0.2 0.4 0.6 0.8 1.0 1.2 E_b/N_0 [dB]
1.0 0.977 0.955 0.933 0.912 0.891 0.871 0.851 σ
0.159 0.153 0.147 0.142 0.136 0.131 0.125 0.12 P_b
Richardson, Shokrollahi and Urbanke
(R=1/2)
(N=10^6)

MN Codes

MacKay+Neal '95

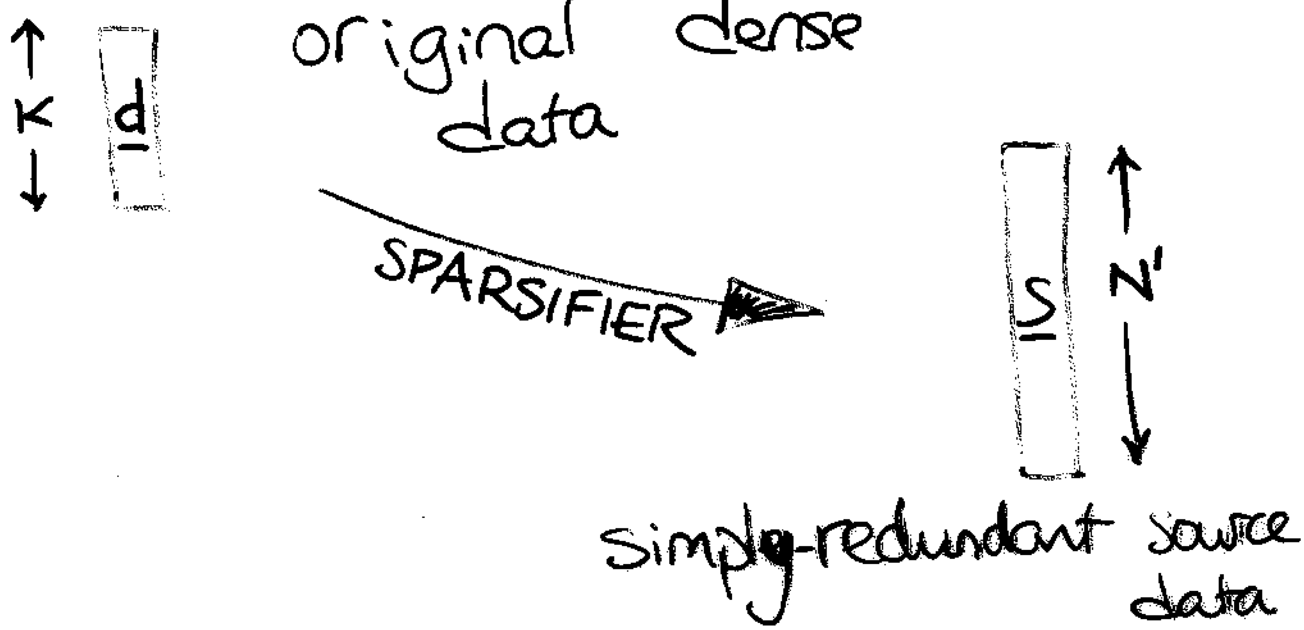
Ordinary linear codes
add redundancy to
arbitrary source data
by appending parity bits.

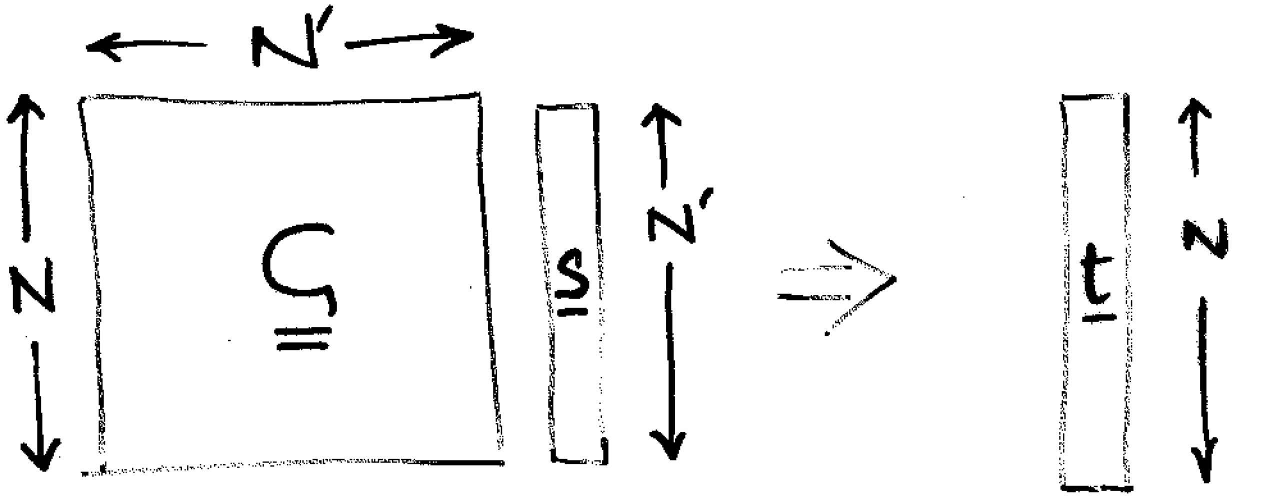


Idea: instead, make
redundant source data,
and encode by multiplying
by a matrix $\underline{\underline{G}}$
which mixes the source bits.

MN Code

Idea:





These are non-linear codes.

MN CODE DECODING PROBLEM

Instead of

$$\underline{\underline{G}}\;\underbrace{\underline{s}}_{\text{DENSE}} + \underbrace{\underline{n}}_{\text{SPARSE}} = \underline{r}$$

INFER $\underline{s}$ & $\underline{n}$

we have

$$\underline{\underline{G}}\;\underbrace{\underline{s}}_{\text{SPARSE}} + \underbrace{\underline{n}}_{\text{SPARSE}} = \underline{r}$$

INFER $\underline{s}$ & $\underline{n}$

MN CODE DETAILS

Define
$$\underline{\underline{G}} = \underline{\underline{C}}_n^{-1}\,\underline{\underline{C}}_s$$

$\underline{\underline{C}}_n$ & $\underline{\underline{C}}_s$ are VERY SPARSE matrices

Then

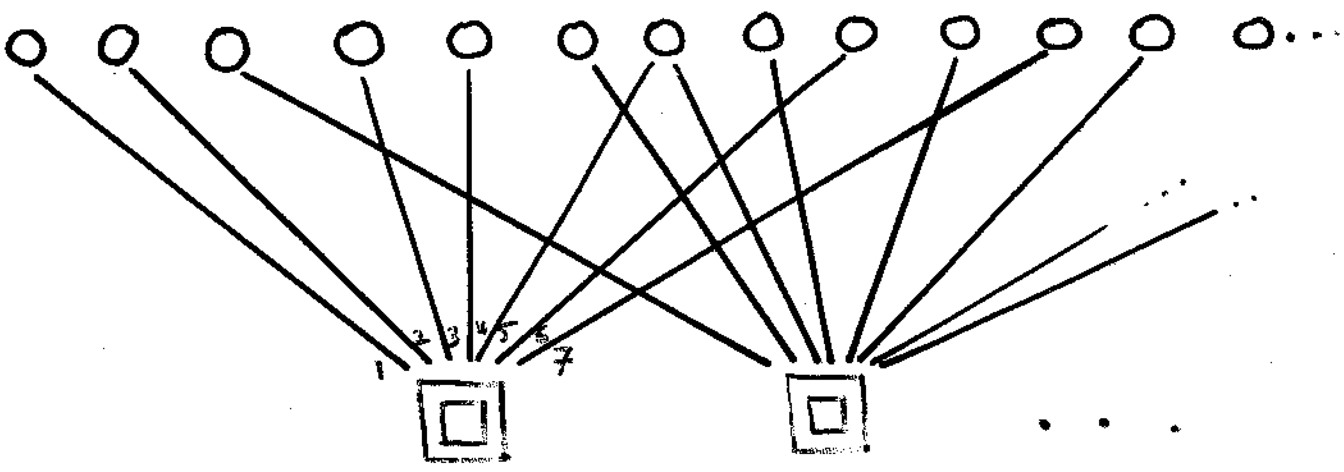
$$\underline{\underline{G}}\,\underline{s} + \underline{n} = \underline{r}$$

$$\Rightarrow \quad \underbrace{\overset{\longleftarrow N' \longrightarrow \quad \longleftarrow N \longrightarrow}{\begin{bmatrix} \underline{\underline{C}}_s & \underline{\underline{C}}_n \end{bmatrix}}}_{\text{VERY SPARSE matrix}}{}_{\,(N \text{ rows})} \begin{bmatrix} \underline{s} \\ \underline{n} \end{bmatrix} \begin{array}{l} \updownarrow N' \\ \updownarrow N \end{array} = \underline{\underline{C}}_n\,\underline{r} \quad \updownarrow N$$

SPARSE vector

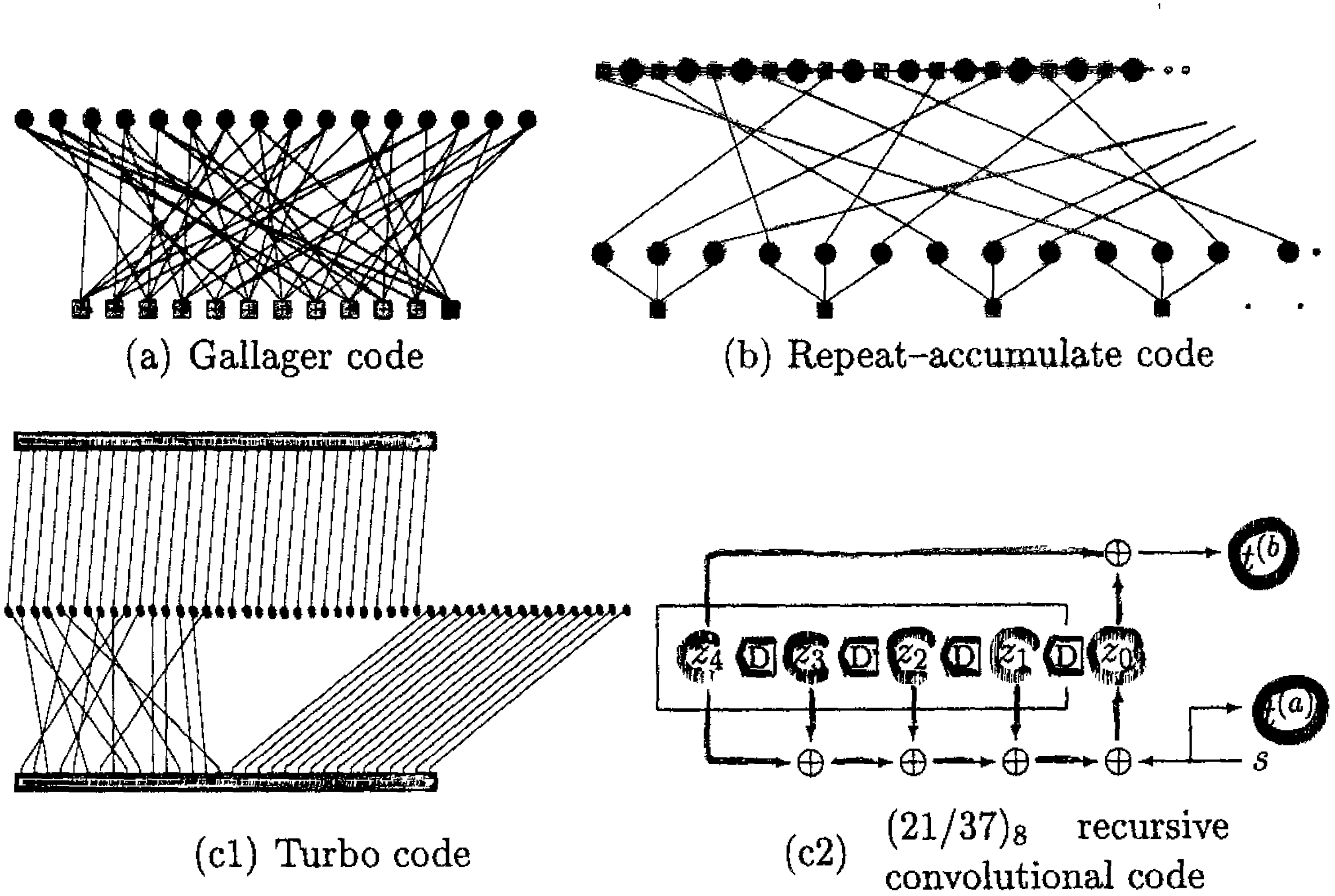
[c.f. Kanter, Saad, Kabashima et al]

TANNER CODES



Complex Constraints:

eg these seven bits are a codeword of (7,4)H



(a) Gallager code

(b) Repeat–accumulate code

(c1) Turbo code

(c2) $(21/37)_8$ recursive convolutional code

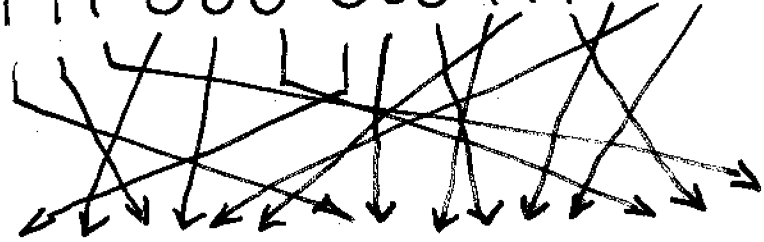
Figure 1. Graphs of three sparse graph codes.

Repeat-Accumulate Codes

$s = 1\,0\,0\,1\,0$

repeat $\quad R_3$

111 000 000 111 000



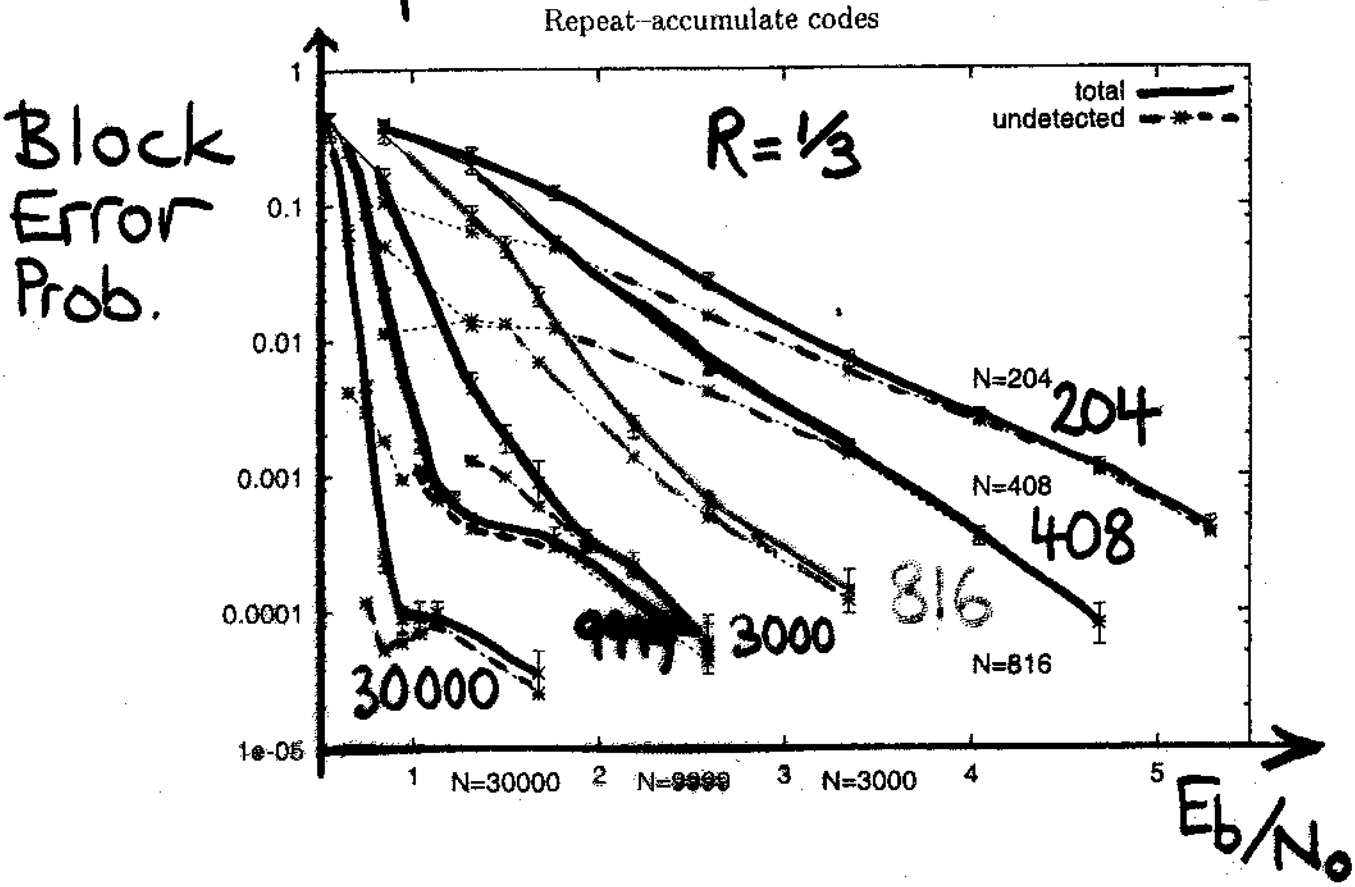
permute

p $\quad$ 001001 101000011

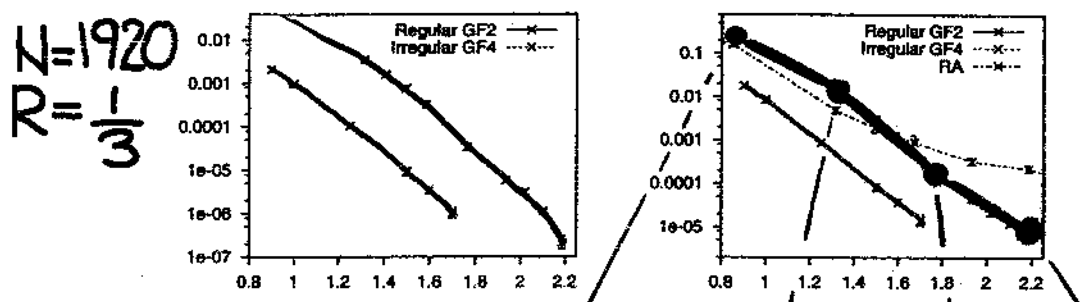
accumulate $\quad t_n = t_{n-1} + p_n$

001110110000010

Repeat-accumulate Codes



DECODING TIMES HAVE A POWER-LAW DISTRIBUTION

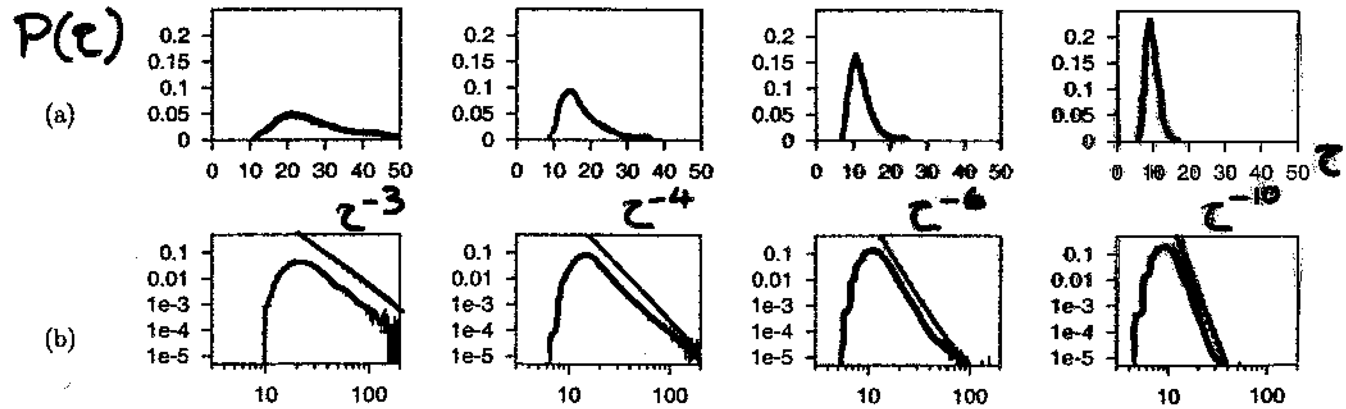


Profile

Columns		Rows	
2	449	5	635
3	389	6	5
4	72		
5	1		
11	2		
13	4		
19	39		

(c)

Figure 1. (a,b) Performance of Gallager codes with $N = 1920$, $R = 1/3$, as a function of E_b/N_0. In (b) we also show the performance of a repeat–accumulate code with $N = 3000$. (c) The profile of the irregular code over GF(4).



x/σ	0.90	0.95	1.00	1.05
(E_b/N_0)/dB	0.846	1.315	1.761	2.185
P_w	0.19	0.014	1.7×10^{-4}	9.0×10^{-6}
Power, p	3	4	6	10

Figure 2. Histograms showing the frequency distribution of decoding times for the binary Gallager code from figure 1: (a) linear plot; (b) log–log plot. The graphs show the number of iterations taken to reach a valid decoding; the value of P_w gives the frequency with which no valid decoding was reached after 1000 iterations. The power p which gives a good fit of the power law distribution $P(\tau) \propto \tau^{-p}$ (for large τ) is also shown.

David MacKay

http://wol.ra.phy.cam.ac.uk/