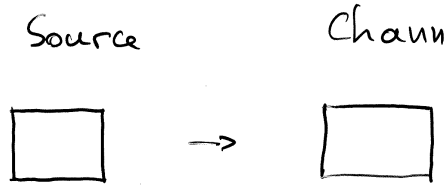
Statistical Mechanics and Error-Correction Codes.



$a_i = 0$ or 1

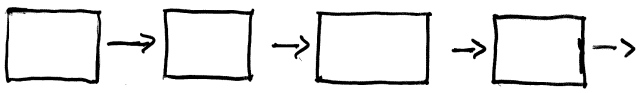
$i = 1, \ldots k$

$$\delta_i = (-1)^{a_i}$$

$$a_k + a_e \rightarrow \delta_k \delta_e$$

Noise during the transmission
=> loss of information.
Shannon: to recover the information
introduce (deterministic) redundancy ("channel encoding")
use this redundancy to infer the message sent ("decoding")
The algorithms which transform the source outputs to redundant messages are called error-correcting codes. Instead of K bits, send N > K bits.
$R = K/N =$ the rate of the code

Source Channel



$a_i = 0$ or 1

$i = 1, \ldots, k$

$\sigma_i = (-1)^{a_i}$

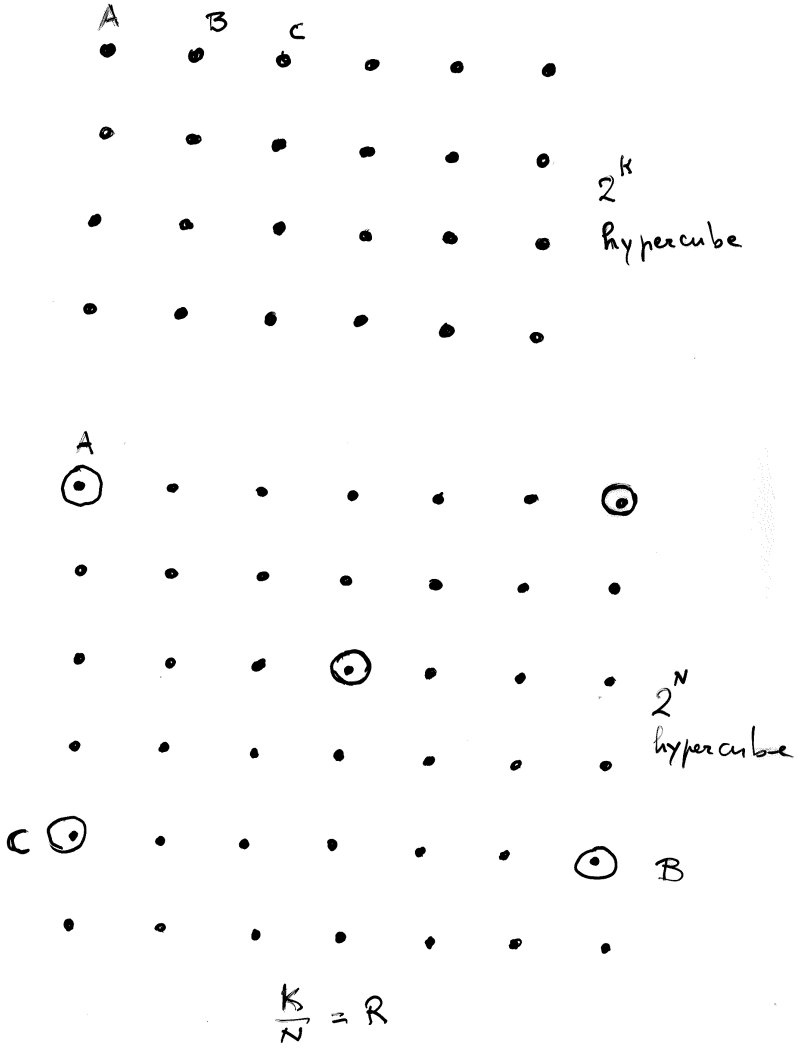
$a_k + a_e \to \sigma_k \sigma_e$

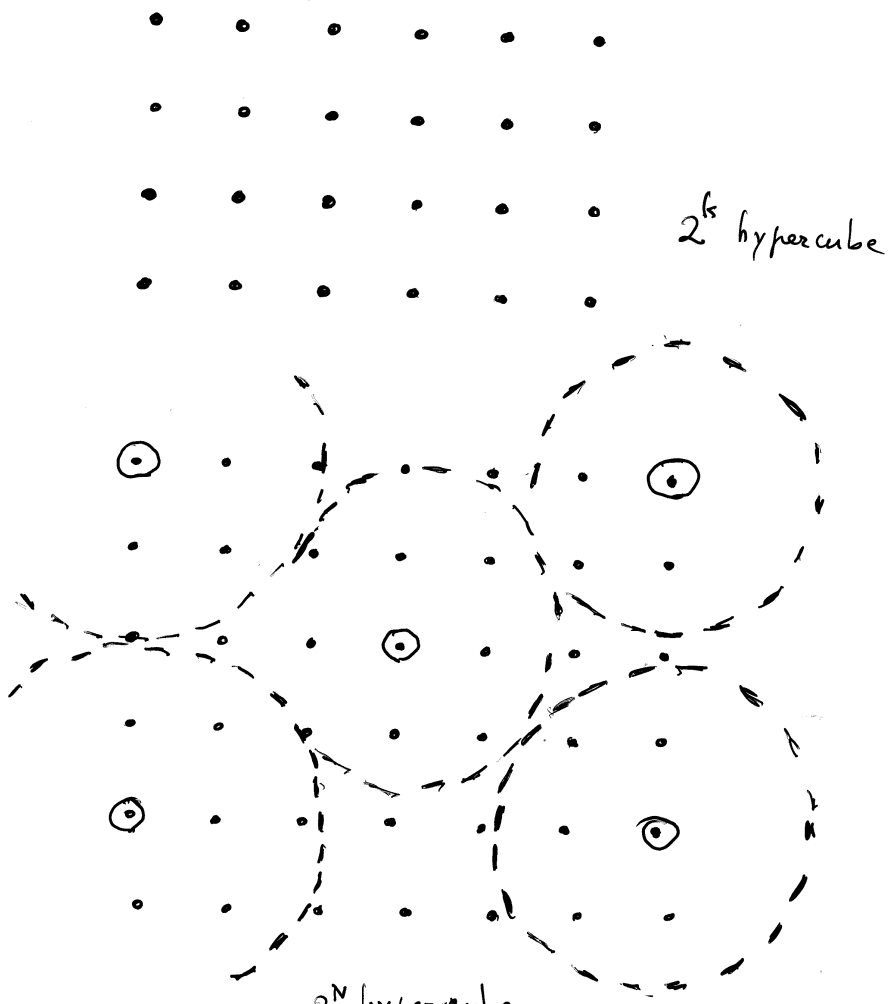
$J_1^{in} \ldots J_N^{in}$ $\qquad$ $J_1^{out} \ldots J_N^{out}$

$N > k$

$\frac{k}{N} = R$ $\qquad$ Rate

Code : A map of 2^k to 2^N


A
B
C
2^K
hypercube
A
2^N
hypercube
C
B
$\frac{K}{N} = R$



$$\frac{K}{N} = R$$

The mathematical theory of communication is probabilistic in nature. Both the production of information and its transmission are considered as probabilistic events.

Message $->$ sequence of K bits

$\vec{\sigma} = \{\sigma_1, \cdots, \sigma_K\},\ \sigma_i = \pm 1$

Assumption: All source words are equally probable

If $\sigma = \pm 1$ is the input because of the noise, the output will be a real number J, in general different from σ.

The statistical properties of the transmission channel are supposed to be known.

Noise during the transmission

$=>$ loss of information.

Channel capacity $\mathcal{C}$ = the maximum information/unit time which can be transmitted through the channel.

The maximum is taken over all possible sources.

Shannon's channel coding theorem 1948.

Consider a random code of rate R and length N

Decoding as above with an appropriate choice of the radius r

P_e the decoding error

If R $< \mathcal{C}$, for any ϵ there exists a $N_0(\epsilon)$ such that for any N $> N_0(\epsilon)$

$$\overline{P_e} < \epsilon$$

error free communication if R $< \mathcal{C}$ and N very large

Very antiintuitive result

A random code, i.e. random map allows error free communication with probability one

What is then the problem ?

DECODING Computational cost of decoding of random codes explodes with N

Choose a decodable code

All codes are good except the ones we know how to decode

This has radically changed with

1993-1994 discovery of turbo codes by Berrou and Glavieux

1997 rediscovery of Low Density Parity Check Codes (LPDC)

Based on random constructions.

Heuristic decoding. Statistical Mechanics very successful dealing with new codes

A detour via Statistical Mechanics

Ising model

$H = -\sum_{i,j} J_{i,j}\sigma_i\sigma_j$

Ferromagnetic couplings $J_{i,j} = J > 0$

Ground state $\sigma_i = 1$

Gauge transformation

$\sigma_i \to \epsilon_i\sigma_i, \quad J_{i,j} \to J_{i,j}\epsilon_i\epsilon_j = J'_{i,j}$

Ground state of new Hamiltonian

$H' = -\sum_{i,j} J'_{i,j}\sigma_i\sigma_j$

$\sigma_i^0 = \epsilon_i$

Ising type model as Error-Correcting Codes

Let ϵ_i, $i = 1, \cdots, K$ the bits to transmit

sourceword

Let $C^l_{i,j}$, $l = 1, \cdots, N$ a connectivity matrix

Ex of $L \times L$ square lattice lattice: $K = L^2$, $N = 2L^2$

We send over the channel the N bits

$J^{l,in}_{i,j} = C^l_{i,j}\epsilon_i\epsilon_j$ codeword

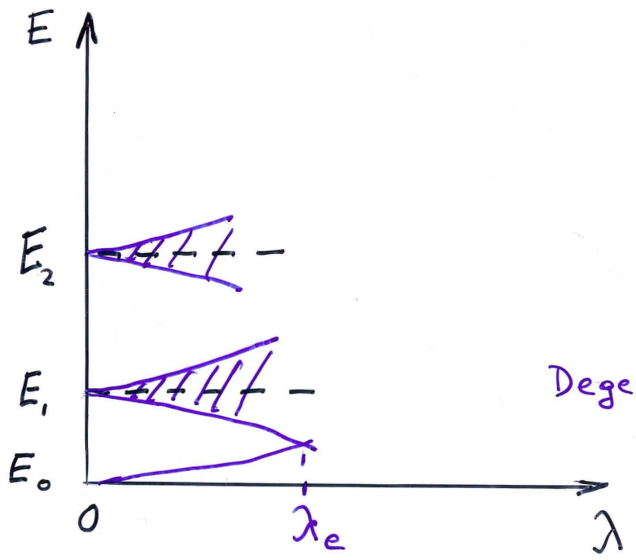
$R = K/N$ rate of the code

Decoding: find the ground state of

$H = -\sum_l C^l_{i,j} J_{i,j}\sigma_i\sigma_j$

The output of the channel is $J_{i,j}^{l,out}$, $l = 1, \cdot, N$

In the absence of noise $J_{i,j}^{l,out} = J_{i,j}^{l,in}$
The ground state $\sigma_i^0 = \epsilon_i$
Noise Suppose $J_{i,j}^{l,out} = J_{i,j}^{l,in} + \lambda \Delta_{i,j}$
$\Delta_{i,j}$ zero mean independant Gaussian random variables and variance one.



$$J_{ij} = J + \Delta_{ij}\,\lambda$$

Error probability

If the noise is symmetric, because of gauge invariance we can assume $\epsilon_i = 1$

Error probability per bit

$P_e = \frac{1-m^0}{2}$ where $m^0 = \frac{1}{K}\sum_i \sigma_i^0$

If we know how to compute the ground state magnetisation we know the error probability of the code

Consider a more general coding scheme
$J^{l,in} = C^{l}_{i_{k_1},\cdots,i_{k_l}} \epsilon_{i_{k_1}}, \cdots, \epsilon_{i_{k_l}}$
Decoding $=>$ find the ground state of
$H^{'} = \sum_l = C^{l}_{i_{k_1},\cdots,i_{k_l}} J^{l,out} \sigma_{i_{k_1}}, \cdots, \sigma_{i_{k_l}}$
Ising model with multispin interactions.
Again $P_e = \frac{1-m^0}{2}$

Like Monsieur Jourdain in Molière's Bourgeois Gentillhome
Monsieur Jourdain discovered he was using prose without knowing it

Linear Codes

Coding

$$\vec{u} = \ G \ \vec{a}$$

$\vec{u}$ length N codeword

$\vec{a}$ length K sourceword

G $N \times K$ matrix with elements zero or one

G is the generator matrix

$u_i = \sum_j g_{ij} a_j$ modulo 2 addition

$J^k = C^k_{i^k_{l_1},\cdots,i^k_{l_k}} \sigma_{i^k_{l_1}}, \cdots, \sigma_{i^k_{l_k}}$

u's are not independent but obey constrains

$\sum_j H_{ij} u_j = 0$ modulo 2, H_{ij} are zero or one

H parity check matrix

$M^k_{i^k_{l_1},\cdots,i^k_{l_k}} J_{i^k_{l_1}}, \cdots, J_{i^k_{l_k}} = 1$

A linear code is given either in terms of the G or H

Relation with Statistical Mechanics

Decoding: a Statistical inference problem

We know the statistical properties of the channel noise.

Transition probability $Q(\vec{J^{out}}|\vec{J^{in}})$

We know the code and the channel output $\vec{J^{out}}$

We can assign a probability $P^{code}(\vec{J}|\vec{J^{out}})$ to any codeword $\vec{J}$ that it was sent

Similarly $P^{source}(\vec{\sigma}|\vec{J^{out}})$ to any sourceword $\vec{\sigma}$

Suppose memoryless channel

$Q(\vec{J^{out}}|\vec{J^{in}}) = \prod_i q(J_i^{out}|J_i^{in})$

Bayes theorem. Case of a single bit What is the probability of J when the channel output is J^{out} ?

$p(J|J^{out}) = q(J^{out}|J)/(q(J^{out}|1) + q(J^{out}| - 1))$

As $J = \pm 1$, a simple calculation gives

$\log p(J|J^{out}) = const. + hJ$, where

$$h = \frac{1}{2} \log(\frac{q(J^{out}|1)}{q(J^{out}| - 1)})$$

Hamiltonian of spin in an external magnetic field h

$$P^{code}(\vec{J}|\vec{J^{out}}) =$$

$$\exp(\sum_i h_i J_i) \prod_k \delta(M^k_{i^k_{l_1},\cdots,i^k_{l_k}} J_{i^k_{l_1}}, \cdots, J_{i^k_{l_k}}, 1)$$

Replacing the delta function by a soft constraint

$$\log P^{code}(\vec{J}|J^{\vec{out}}) =$$

$$\sum_i h_i J_i \; + \; U \sum_k M^k_{i^k_{l_1},\cdots,i^k_{l_k}} J_{i^k_{l_1}}, \cdots, J_{i^k_{l_k}}$$

Hamiltonian with ferromagnetic multispin couplings in a random external field

Constrains provide the ferromagnetic couplings, the channel output the external field.

Writing the code letters J in terms of the source letters σ

$$\log P^{source}(\vec{\sigma}|J^{\vec{out}}) = \sum_k h_k C^k_{i^k_{l_1},\cdots,i^k_{l_k}} \sigma_{i^k_{l_1}}, \cdots, \sigma_{i^k_{l_k}}$$

Spin glass Hamiltonian with multispin couplings

Exactly what I proposed earlier

Nature of the couplings determined by the code.

Coupling strength by the channel output

Most probable sequence $\vec{J}$ or $\vec{\sigma}$

=> ground state of the corresponding Hamiltonian

Reminder $P_e = \frac{1-m^0}{2}$

Magnetisation self-averaging

We can compute the error probability of a code as a function of signal to noise by computing the magnetisation of the corresponding spin system.

Methods of Statistical Mechanics of disordered systems.

Example: R= 1/2 convolutional codes

$$J_i^{(1)} = \sigma_i \sigma_{i-1} \sigma_{i-2}, \;\; J_i^{(2)} = \sigma_i \sigma_{i-2}$$

Constraints

$$J_k^{(1)} J_{k+1}^{(1)} J_k^{(2)} J_{k+1}^{(2)} J_{k+2}^{(2)} = 1$$

$$-H = \frac{1}{w^2} \sum_k J_k^{1,out} \tau_k \tau_{k-1} \tau_{k-2} + J_k^{2,out} \tau_k \tau_{k-2}$$

I assumed a Gaussian noise. In that case $h_k = J_k^{out}/w^2$, w^2 is the variance of the noise.
one dimensional spin glass Hamiltonian.
ground state $=>$ $T = 0$ transfer matrix algorithm.
This is the Viterbi algorithm in coding theory (1967)

Convolutional codes One dimensional spin systems with finite range interactions

Complexity of decodindg exponential in the range of the interactions

Zero error probabilty above a certain signal threshold as in Shannon theorem requires a phase transition.

No phase transitions for one dimensional spin systems with short range interactions.

Impossible to reach zero error probability with convolutional codes

Statistical Mechanics of Capacity Approaching Codes

LPDC : Kabashima Kanter and Saad, Montanari

Turbo Codes: Montanari and N.S.

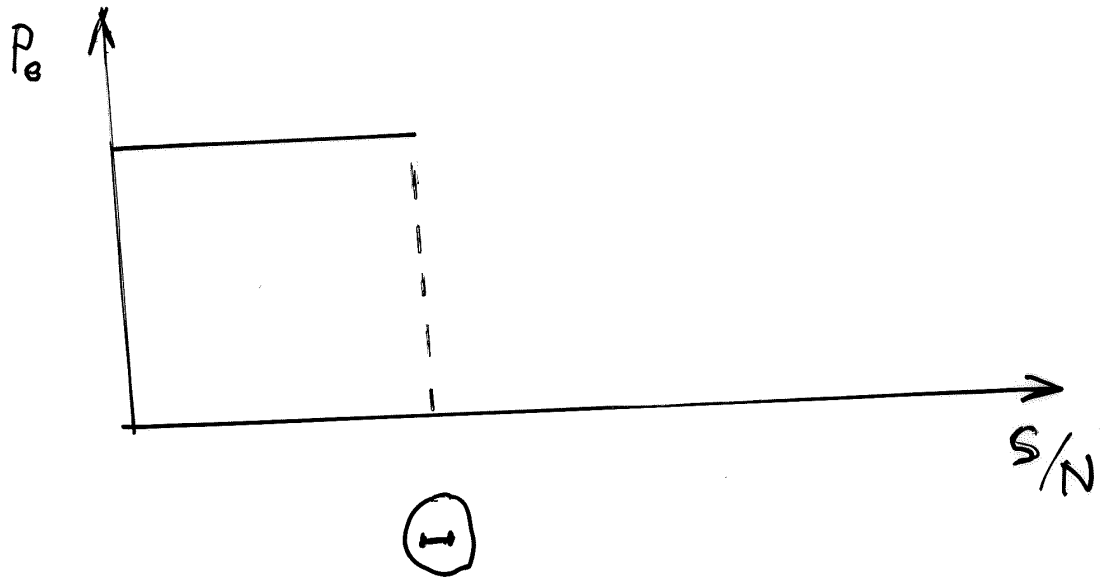
LPDC and Turbo Codes are linear Codes

Phase diagramme studied by the replica method

In both cases there is a Phase Transition

In coding language there is a signal to noise s/n threshold Θ such that for

$$s/n > \Theta \qquad P_e = 0$$



(1)

Calculation of Θ does not
depend on the decoding algorithm

Assumes that thermal equilibrium was reached

Decoding

Low Density Parity Check Codes (LPDC): Gallager 1962

LPDC defined through the Parity Check matrix H $H\vec{u} = 0$ modulo 2

H is a sparse random K $\times$ N matrix

Each column of H has l elements equal to one and all other elements equal to zero. Each row has m non zero elements. The rate of the code is $R = 1 - m/l$

They correspond to ferromagnetic spin models with l-spin infinite strength interactions on a random sparse graph in a random external magnetic field

Gallager proposed an approximate iterative decoding algorithm Equivalent to computing the local magnetisations by iterating the Thouless Anderson Palmer (TAP) or cavity equations of spin-glasses

It is hoped that this procedure will converge to a fixed point after a reasonable number of iterations.

The number of iterations will depend on the amount of noise. If the noise is too strong there will be no convergence.

This algorithm, would be exact in a graph without loops.

It is approximate because of the presence of loops on a random graph.

The same algorithm was rediscovered recently in computer science, where it is called belief propagation

Big similarity with the physics of glasses

At low temperature thermal equilibrium is not reached Aging phenomena

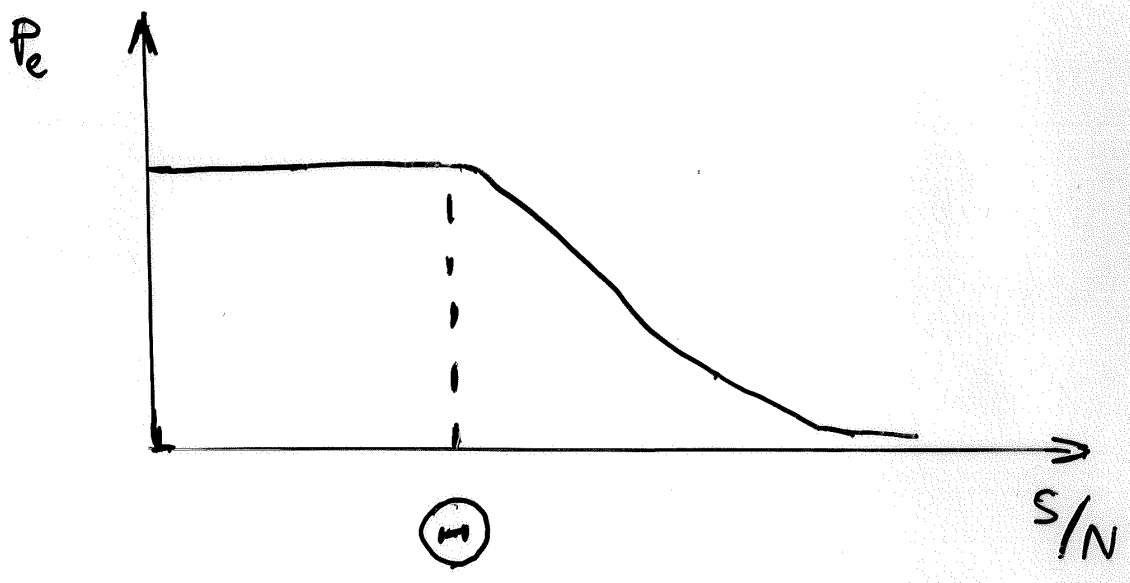
The same behaviour is expected for the decoding algorithm.

In the particular case of the erasure channel it was proven by Montanari Richardson and Urbanke

that there is another threshold $\theta <$ ⊖

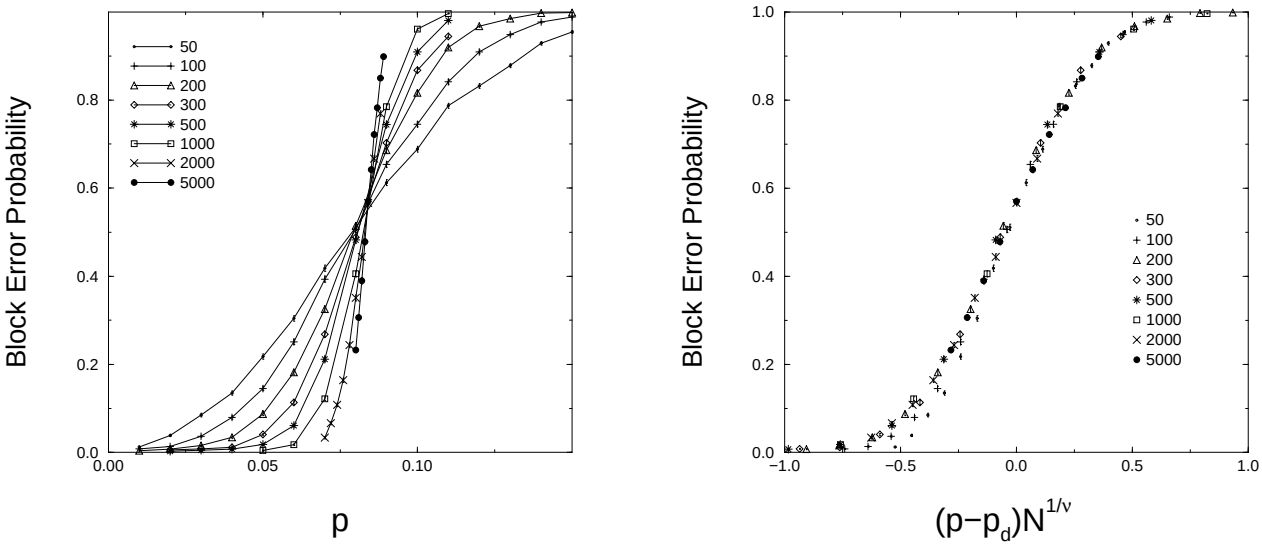
For $s/n < \theta$ the algoritm does not find the optimal solution. Very new phenomenon in communication theory

What about finite size effects ?


P_e
(m)
S/N

There is finite size scaling

$$P(\frac{s}{n}, N) = f(z), \quad z = N^{1/\nu}(\frac{s}{n} - cN^{-2/3})$$



The block error probability for a $(6, 3)$ regular Gallager code for several block lengths (see the legend). On the right the *scaling plot* (see text) for the same quantity.

Observed numerically in the gaussian channel
proved mathematically by Montanari Richardson
and Urbanke for the erasure channel
Surprising result
finite size scaling is known for static phase transitions
This is NOT a static but a dynamic
phase transition
Finite size scaling completely unexpected by information theorists
IMPORTANT PRACTICAL APPLICATIONS