Statistical physics
and
statistical inference

Marc Mézard

Ecole normale supérieure
PSL University

Turing lecture 2
ICTS-Bangalore
January 7, 2020

What is inference?

Statistics

Infer a hidden rule, or hidden variables, from data.

Restricted sense : find parameters of a probability distribution

Urn with 10.000 balls. Draw 100, find 70 white balls and 30 black

Best guess for the composition of the urn? How reliable? Probability that it has 6000 white- 4000 black?

If only black and white balls , with fraction x of white,

probability to pick-up 70 white balls is $\binom{100}{70} x^{70}(1-x)^{30}$

Log likelihood of x : $L(x) = 70 \log x + 30 \log(1-x)$

Maximum at $x^* = .7$ Probability of .6 : $e^{L(.6)-L(.7)}$

Bayesian inference

Unknown parameters	x	Prior	$P(x)$
Measurements	y	Likelihood	$P(y\|x)$

Posterior $$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

Bayesian inference

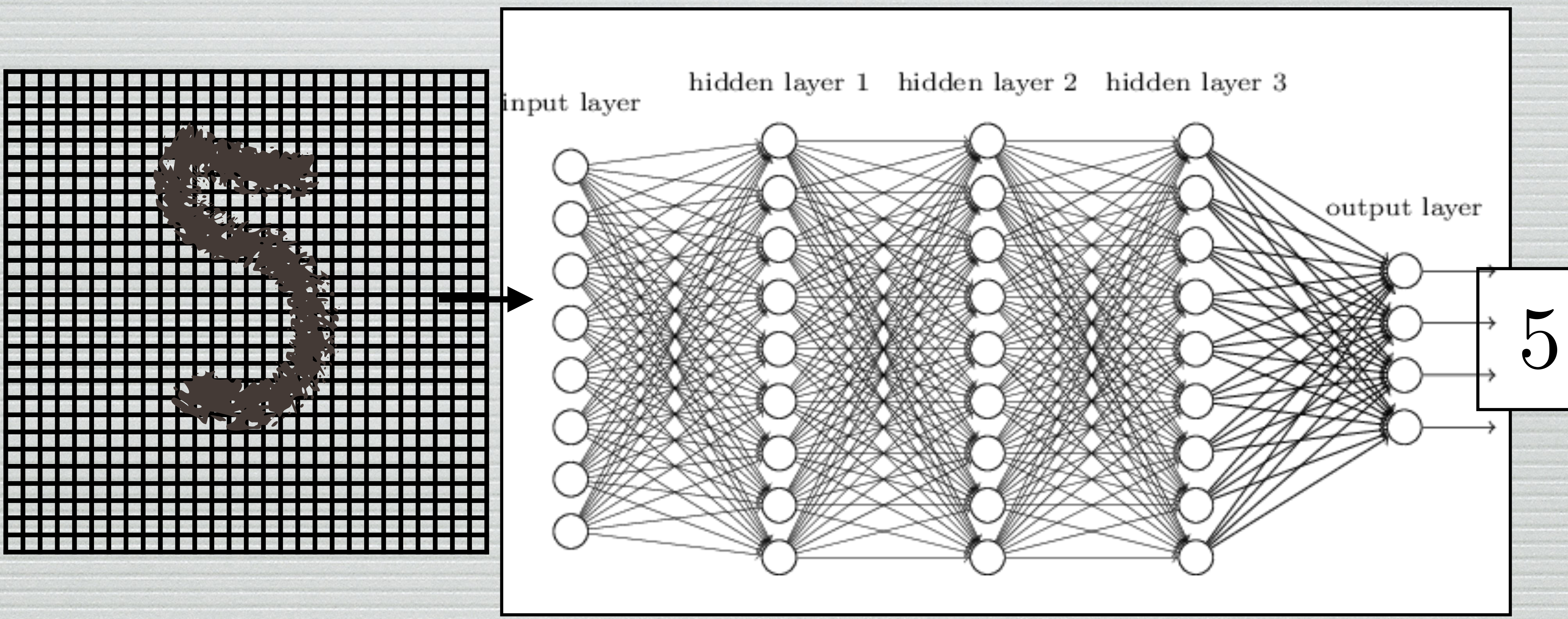
Unknown parameters	x	Prior	$P(x)$
Measurements	y	Likelihood	$P(y\|x)$

Posterior

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

What is inference?

Artificial intelligence,
machine learning

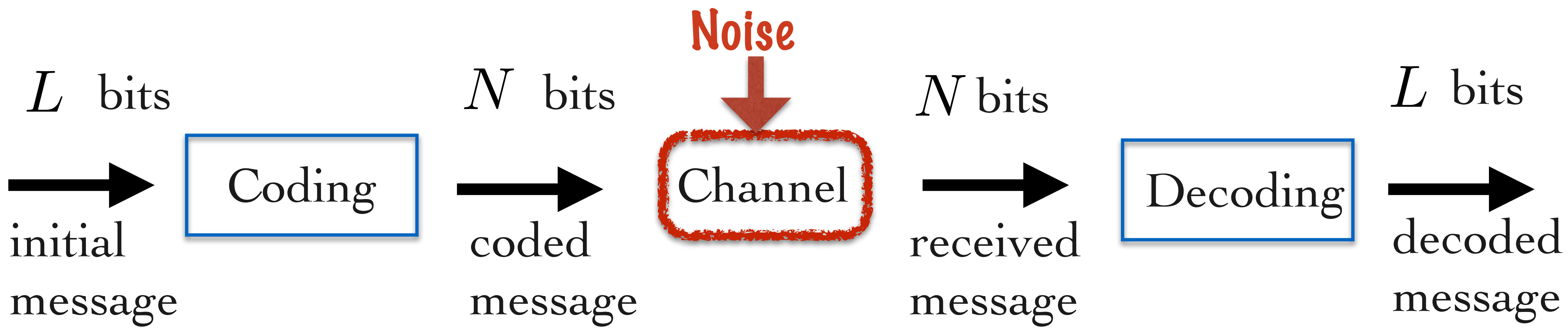


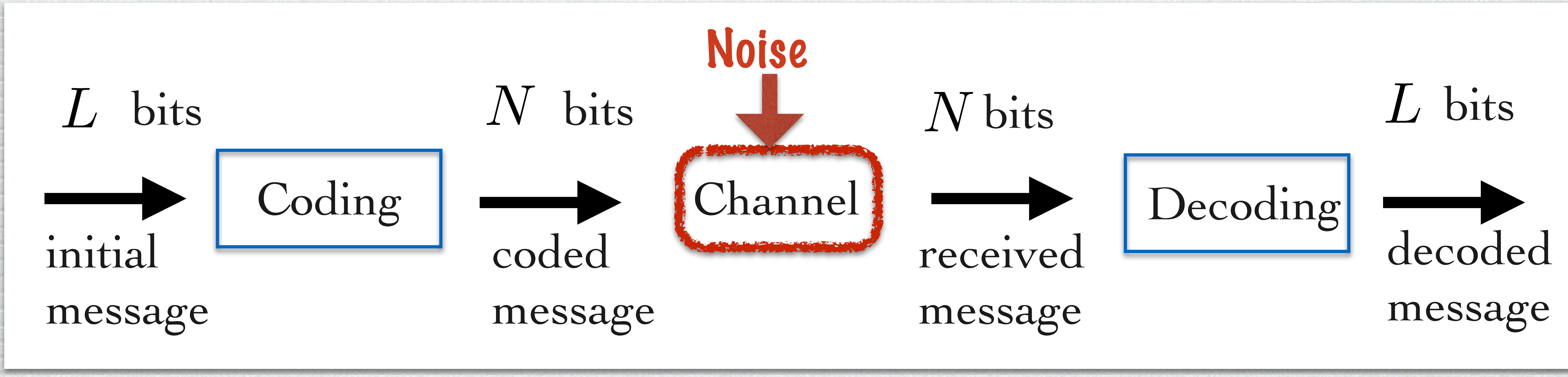
Machine with hundreds of thousands of parameters, trained on very large data base: infer the parameters from data (supervised learning)

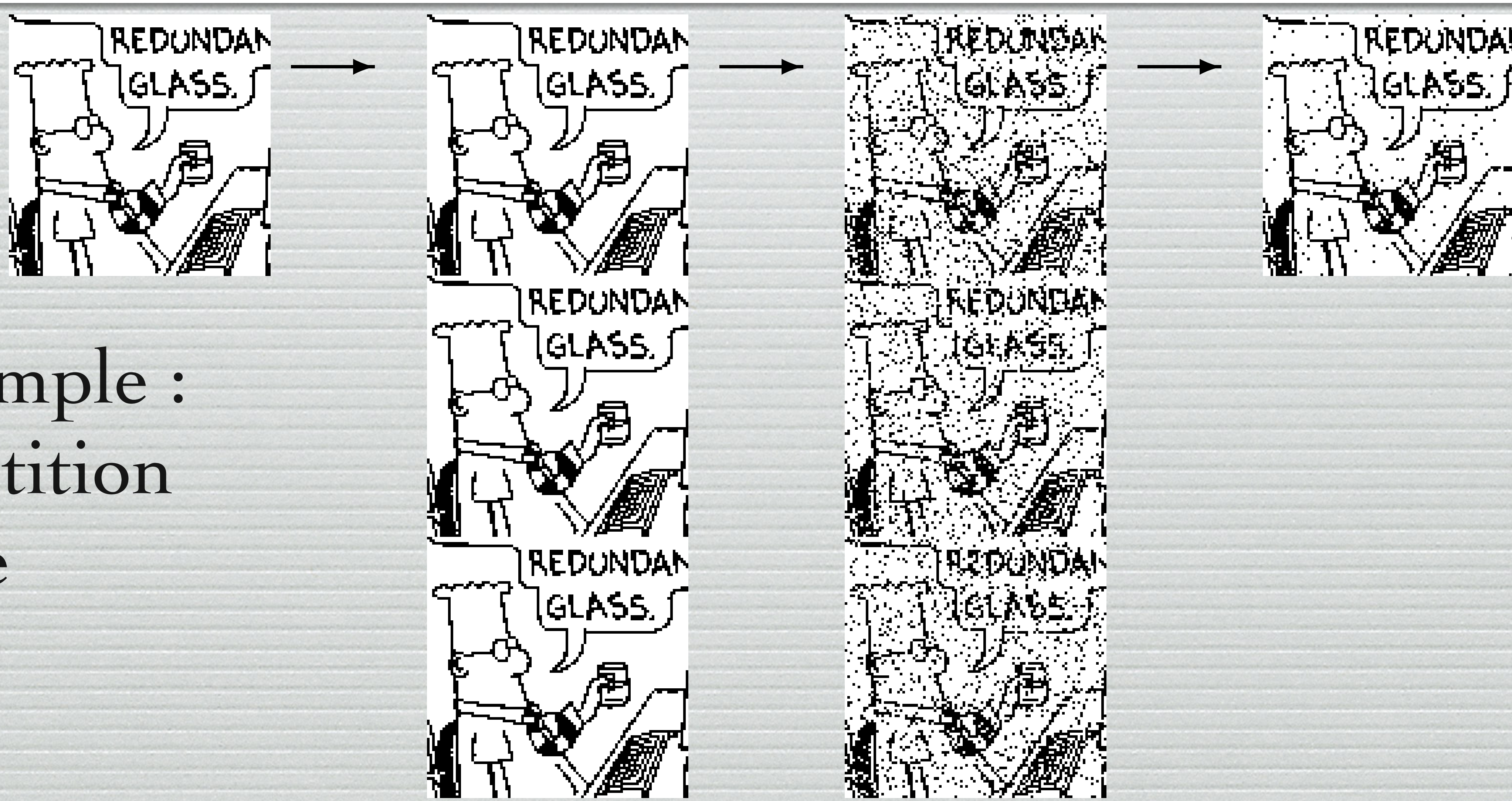
What is inference?

Information theory, communication, signal processing

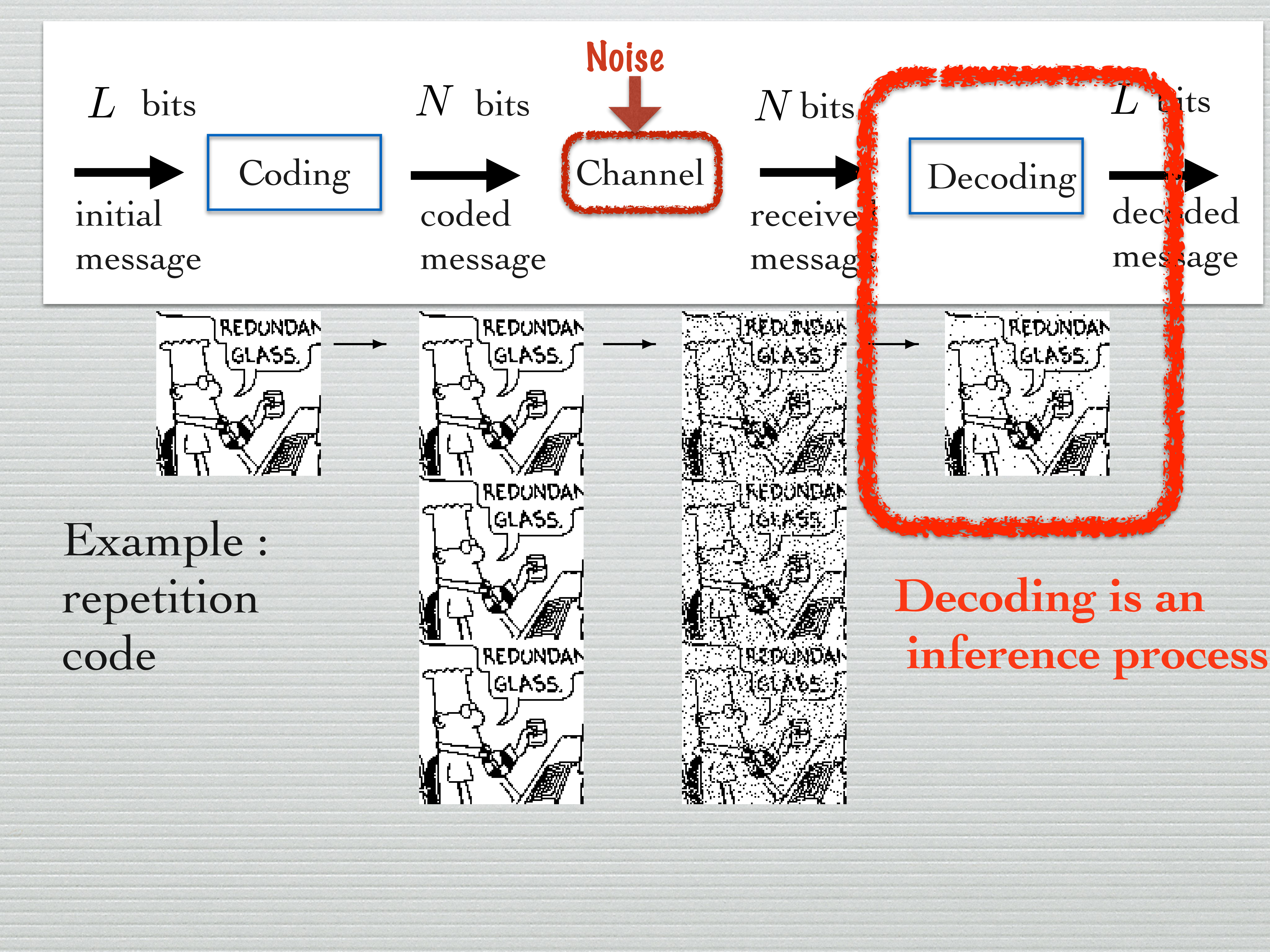
Information transfer : error correction by the use of redundancy

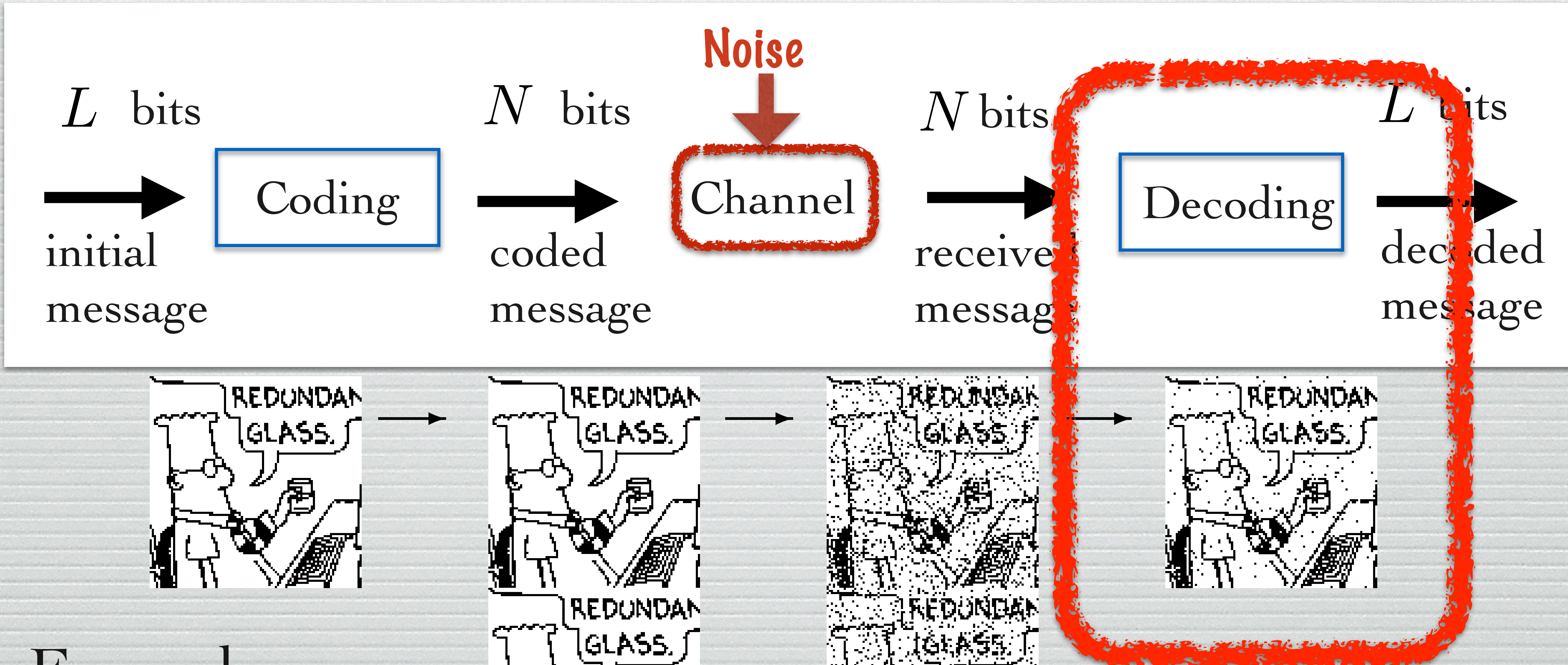

Noise
L bits
initial message
Coding
N bits
coded message
Channel
N bits
received message
Decoding
L bits
decoded message



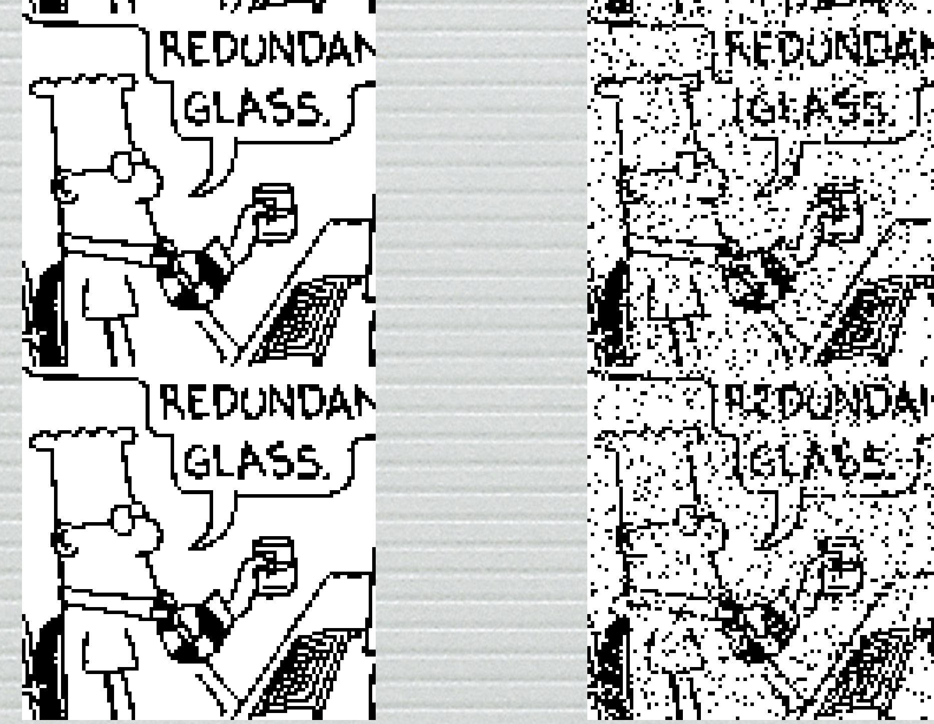


Example : repetition code


L bits
initial message
Coding
N bits
coded message
Noise
Channel
N bits
receive
messag
Decoding
L
its
dec ded
mes age
REDUNDAN
GLASS.
Example : repetition code
Decoding is an inference process



Example : repetition code



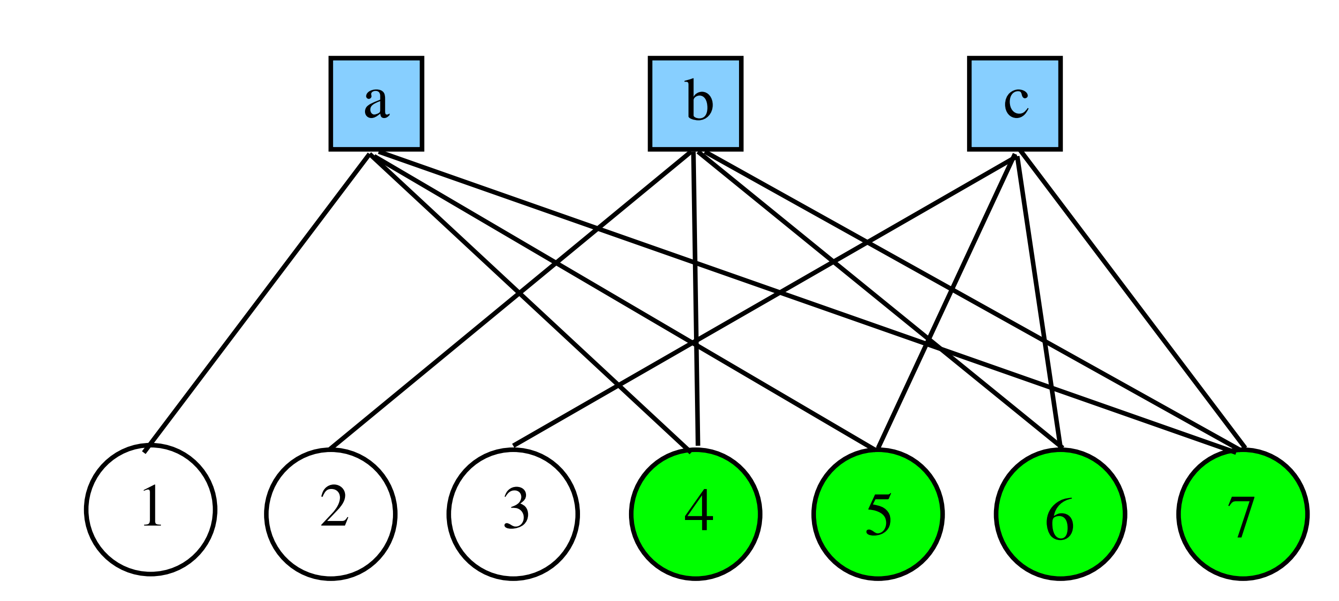
Decoding is an inference process

More powerful codes : multi-bits interactions

Efficient codes : parity checks (LDPC codes)

Add redundancy, with structure allowing to decode

$x_i \in \{0,1\}$

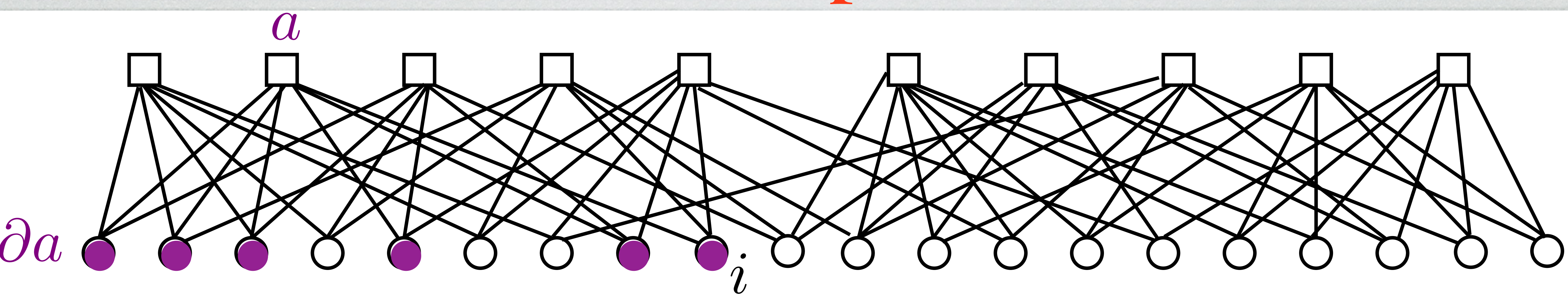


$$a: \quad x_1 + x_4 + x_5 + x_7 = 0 \pmod 2$$
$$b: \quad x_2 + x_4 + x_6 + x_7 = 0 \pmod 2$$
$$c: \quad x_3 + x_5 + x_6 + x_7 = 0 \pmod 2$$

2^4 codewords

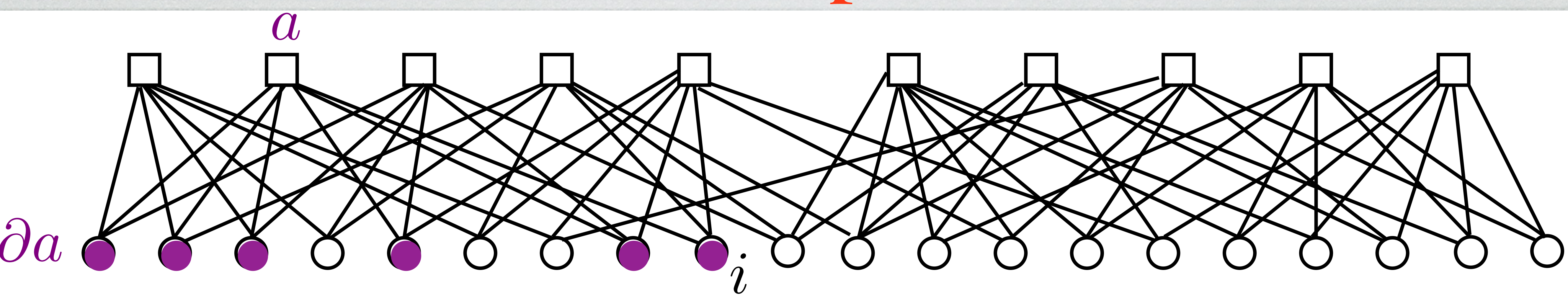
among 2^7 words

Error decoding: « crystal hunting » inference problem



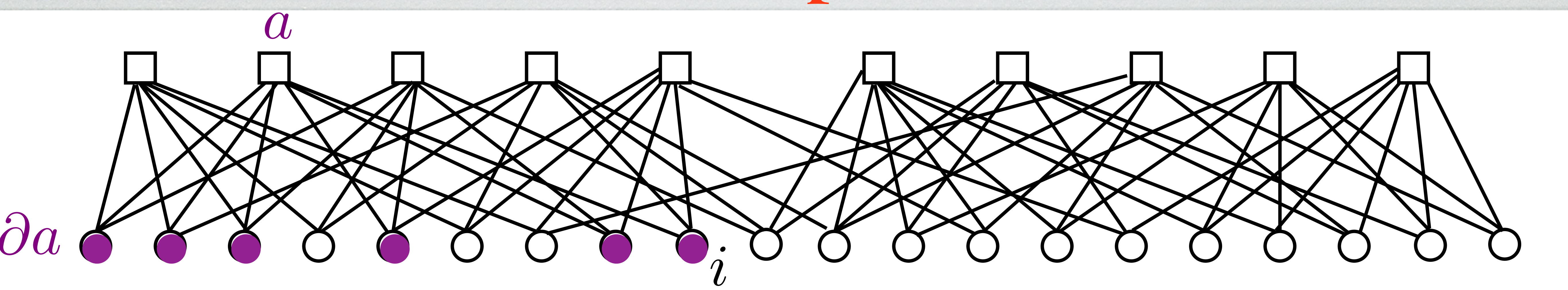
$$P(x_1,\ldots,x_N|y_1,\ldots,y_N) = \frac{1}{Z}\prod_i \psi_i(x_i|y_i)\ \prod_a \mathbb{I}\left(\sum_{i\in\partial a} x_i = 0\ (mod\ 2)\right)$$

Error decoding: « crystal hunting » inference problem



$$P(x_1,\dots,x_N|\underbrace{y_1,\dots,y_N}_{\text{received}}) = \frac{1}{Z}\prod_i \psi_i(x_i|y_i) \prod_a \mathbb{I}\left(\sum_{i\in\partial a} x_i = 0 \ (mod\ 2)\right)$$

Error decoding: « crystal hunting » inference problem

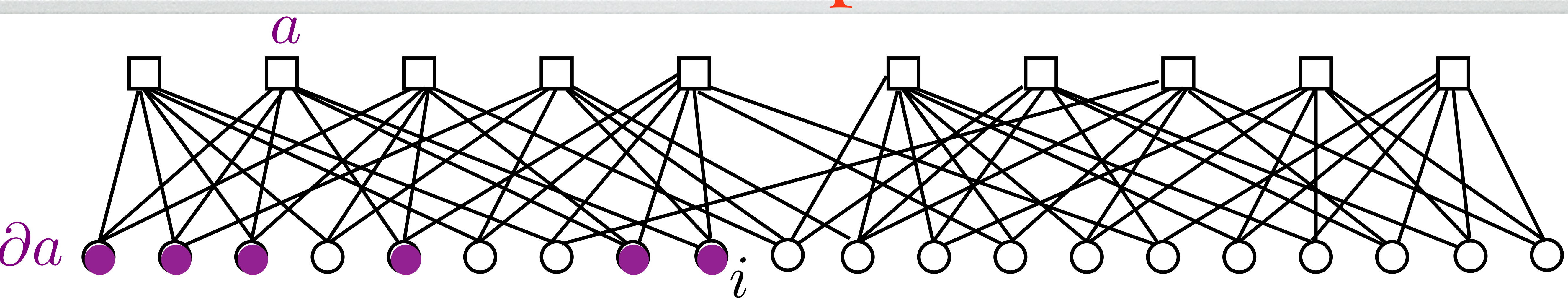


$$P(x_1,\ldots,x_N|y_1,\ldots,y_N) = \frac{1}{Z}\prod_i \psi_i(x_i|y_i) \prod_a \mathbb{I}\left(\sum_{i\in\partial a} x_i = 0 \ (mod\ 2)\right)$$

received

A priori knowledge of the channel

Error decoding: « crystal hunting » inference problem



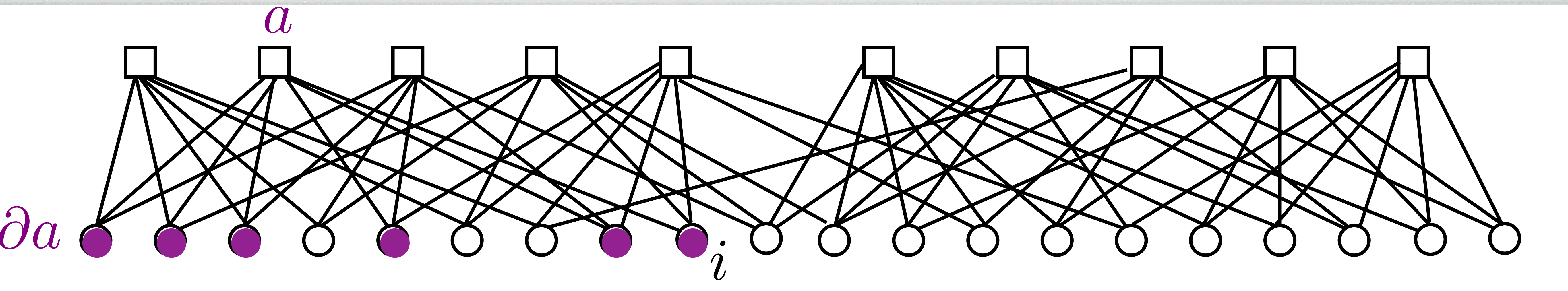
$$P(x_1,\ldots,x_N|y_1,\ldots,y_N) = \frac{1}{Z}\prod_i \psi_i(x_i|y_i) \prod_a \mathbb{I}\left(\sum_{i\in\partial a} x_i = 0 \ (mod\ 2)\right)$$

received

A priori knowledge of the channel

Parity check constraints

Error decoding: « crystal hunting » inference problem



$$P(x_1,\ldots,x_N|\boxed{y_1,\ldots,y_N}) = \frac{1}{Z}\prod_i \psi_i(x_i|y_i)\prod_a \mathbb{I}\left(\sum_{i\in\partial a} x_i = 0\ (mod\ 2)\right)$$

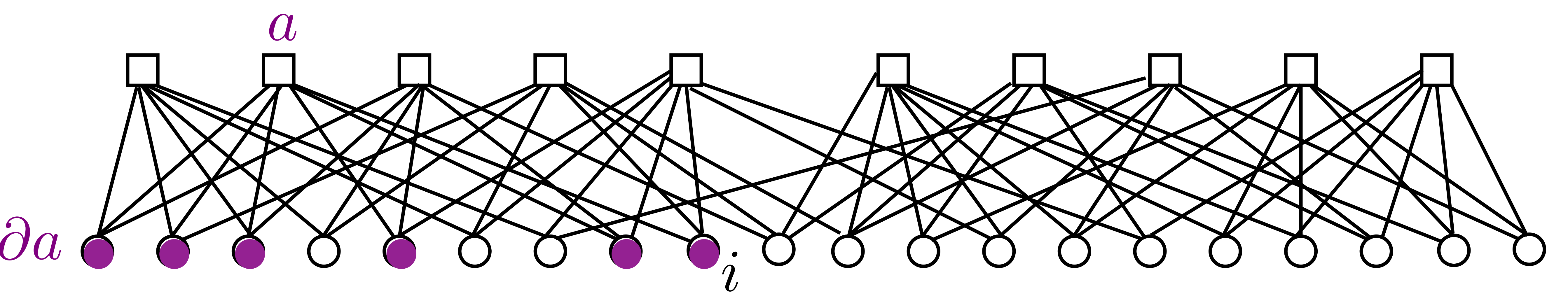
received

A priori knowledge of the channel

Parity check constraints

Spin glass problem with multispin interactions, discontinuous glass transition (1 step RSB)

Error decoding: inference problem



$$P(x_1, \ldots, x_N | y_1, \ldots, y_N) = \frac{1}{Z} \prod_i \psi_i(x_i|y_i) \prod_a \mathbb{I}\left(\sum_{i \in \partial a} x_i = 0 \ (mod\ 2) \right)$$

One possible decoding algorithm: use belief-propagation mean-field equations relating the local fields

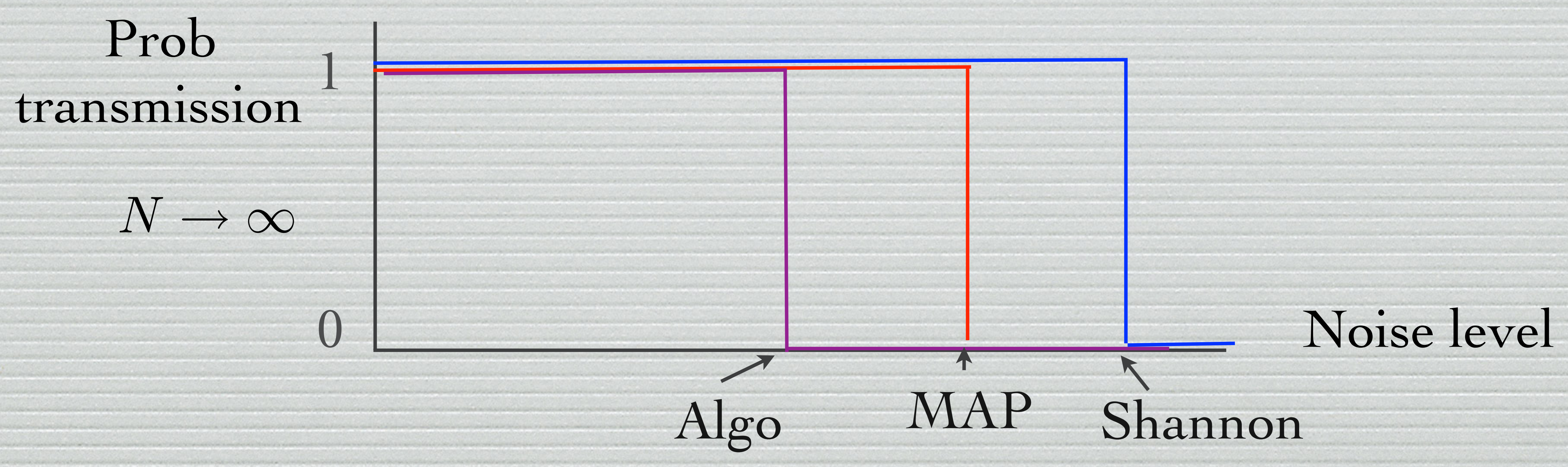
Solve them iteratively (Gallager)

Phase Transitions in Error correcting codes

Shannon 1948 (**random** code **ensemble**)

Typical structured code **ensemble** (e.g. LDPC), with optimal decoding

Typical structured code **ensemble** , with fast BP-based decoding algorithm

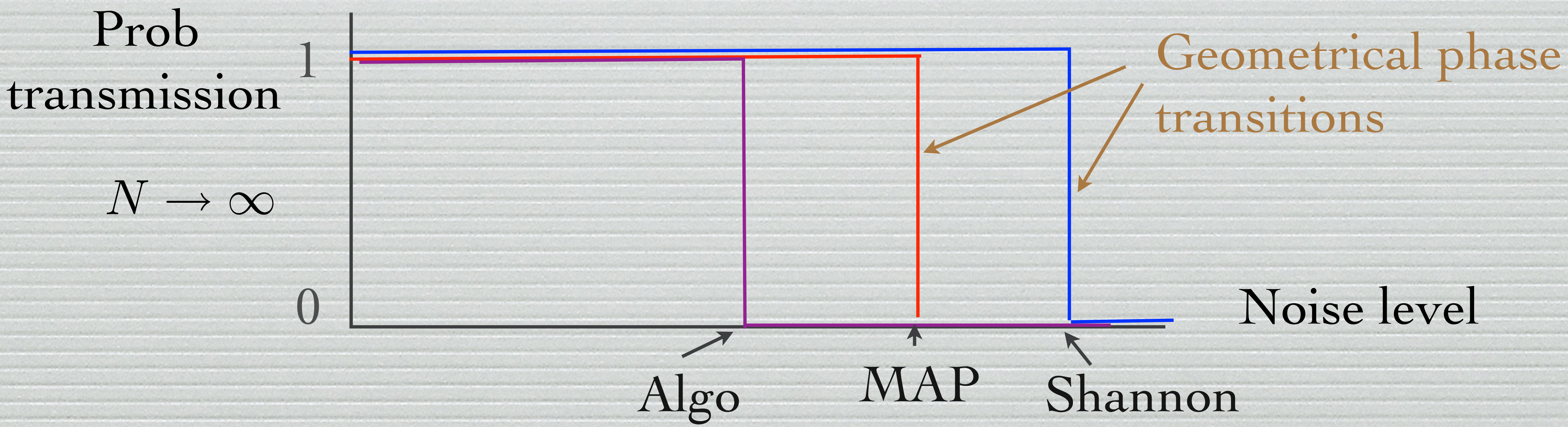


Phase Transitions in Error correcting codes

Shannon 1948 (**random** code **ensemble**)

Typical structured code **ensemble** (e.g. LDPC), with optimal decoding

Typical structured code **ensemble** , with fast BP-based decoding algorithm

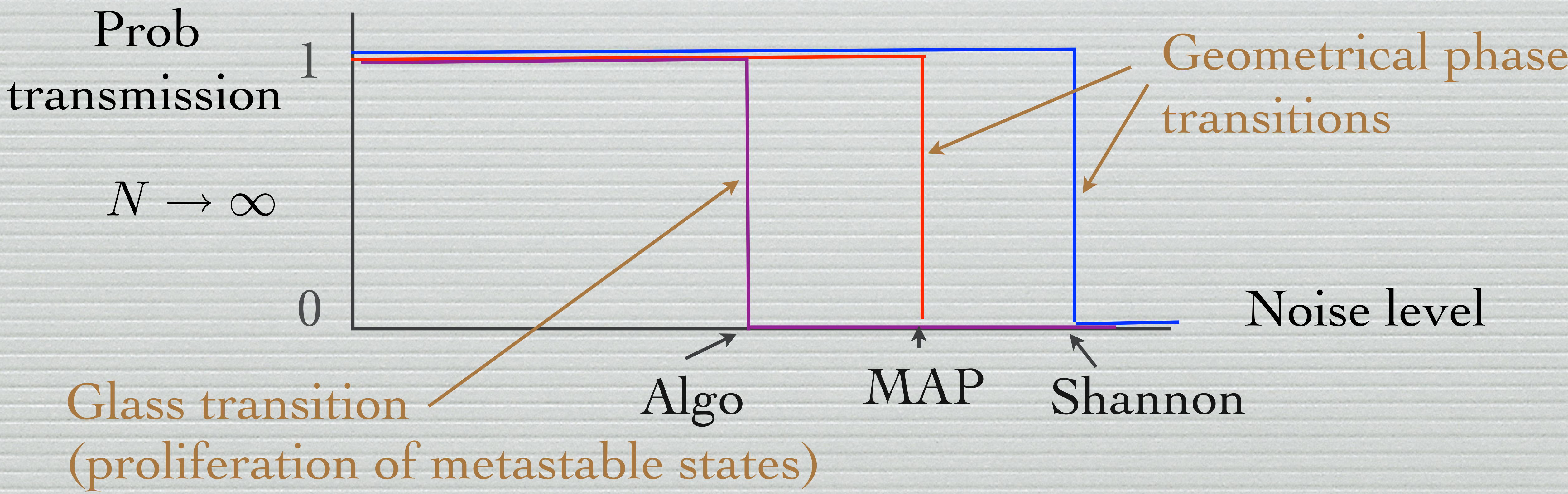


Phase Transitions in Error correcting codes

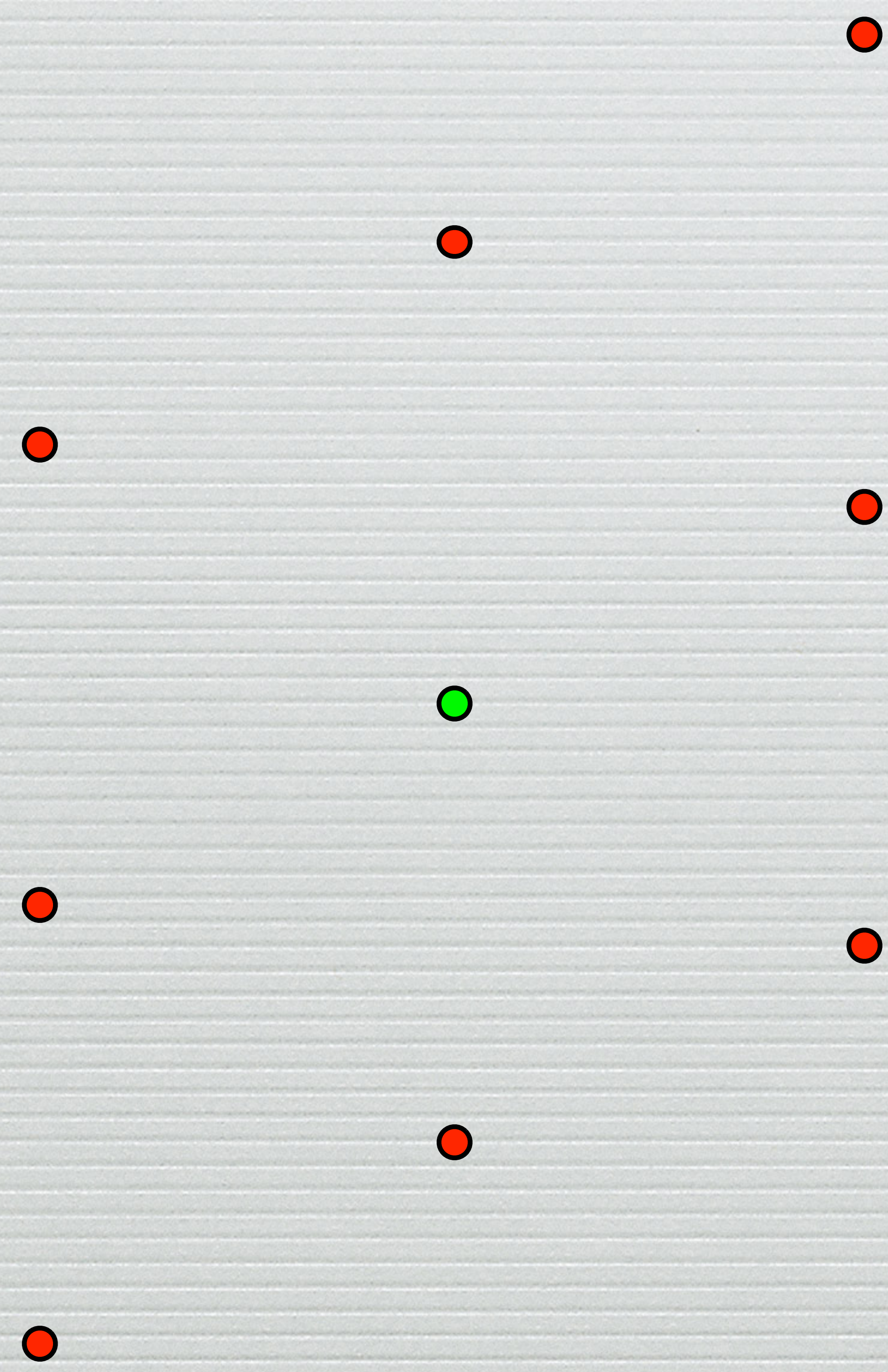
Shannon 1948 (**random** code **ensemble**)

Typical structured code **ensemble** (e.g. LDPC), with optimal decoding

Typical structured code **ensemble** , with fast BP-based decoding algorithm

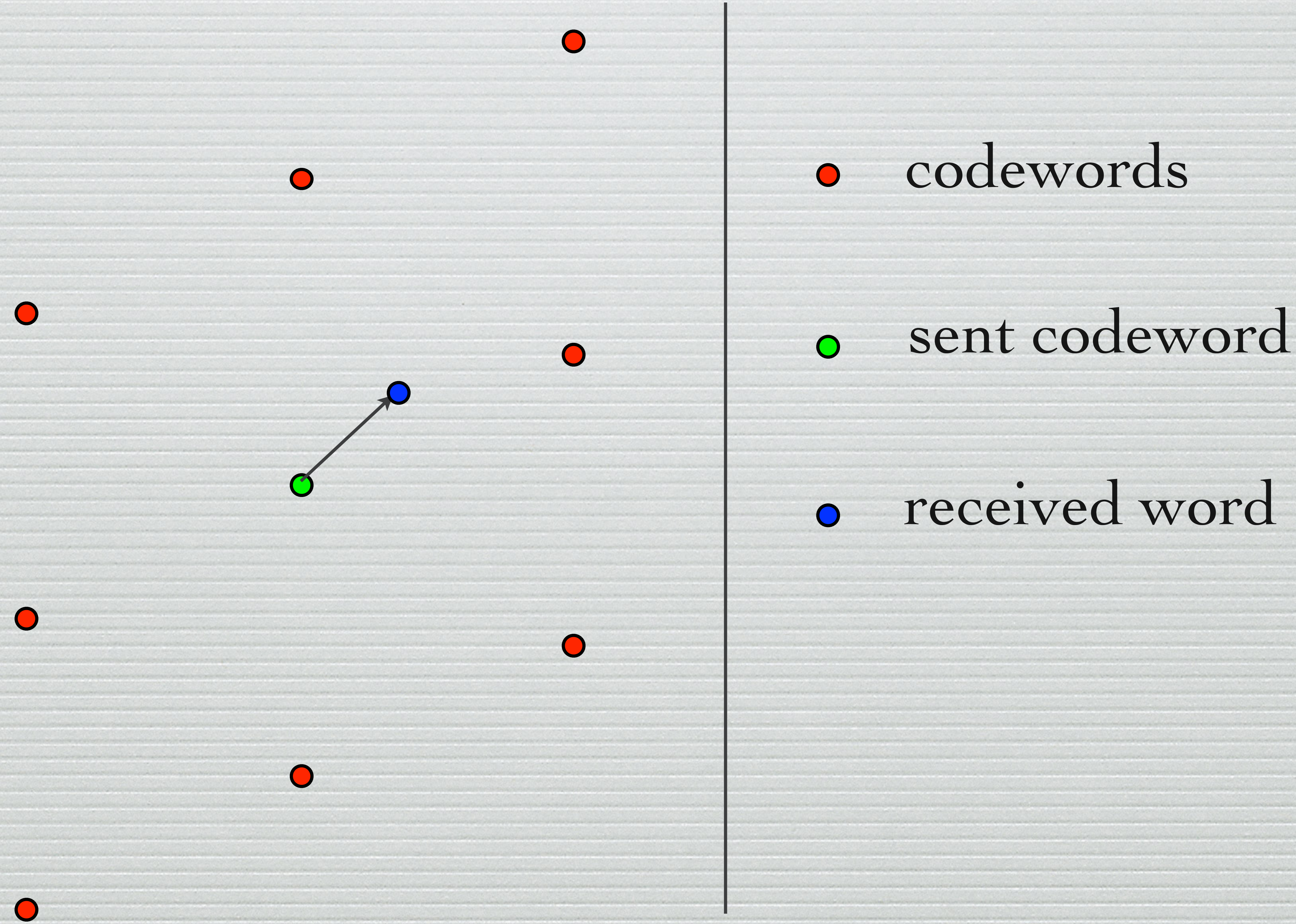


Error correction: decoding

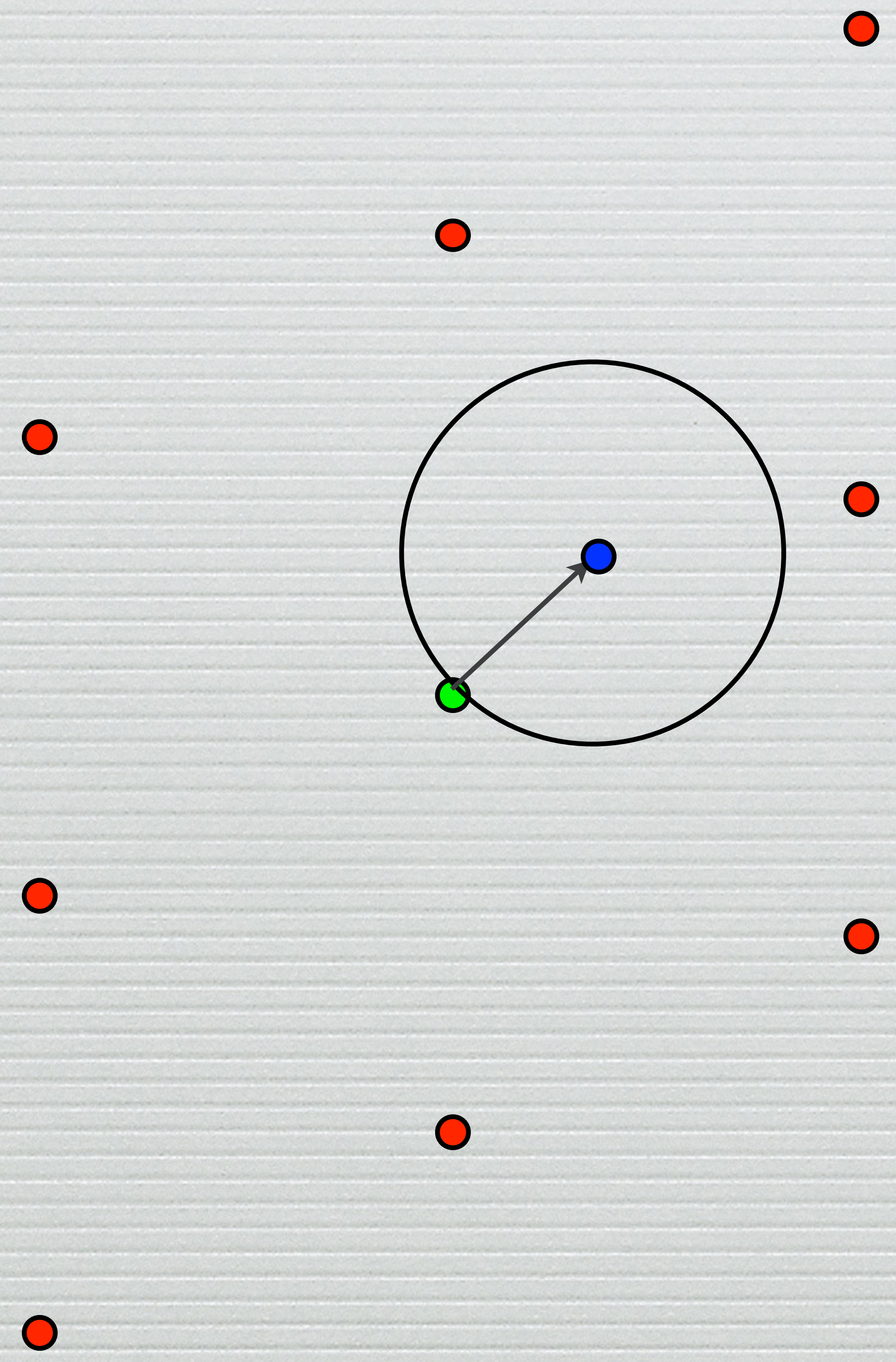


- codewords
- sent codeword

Error correction: decoding

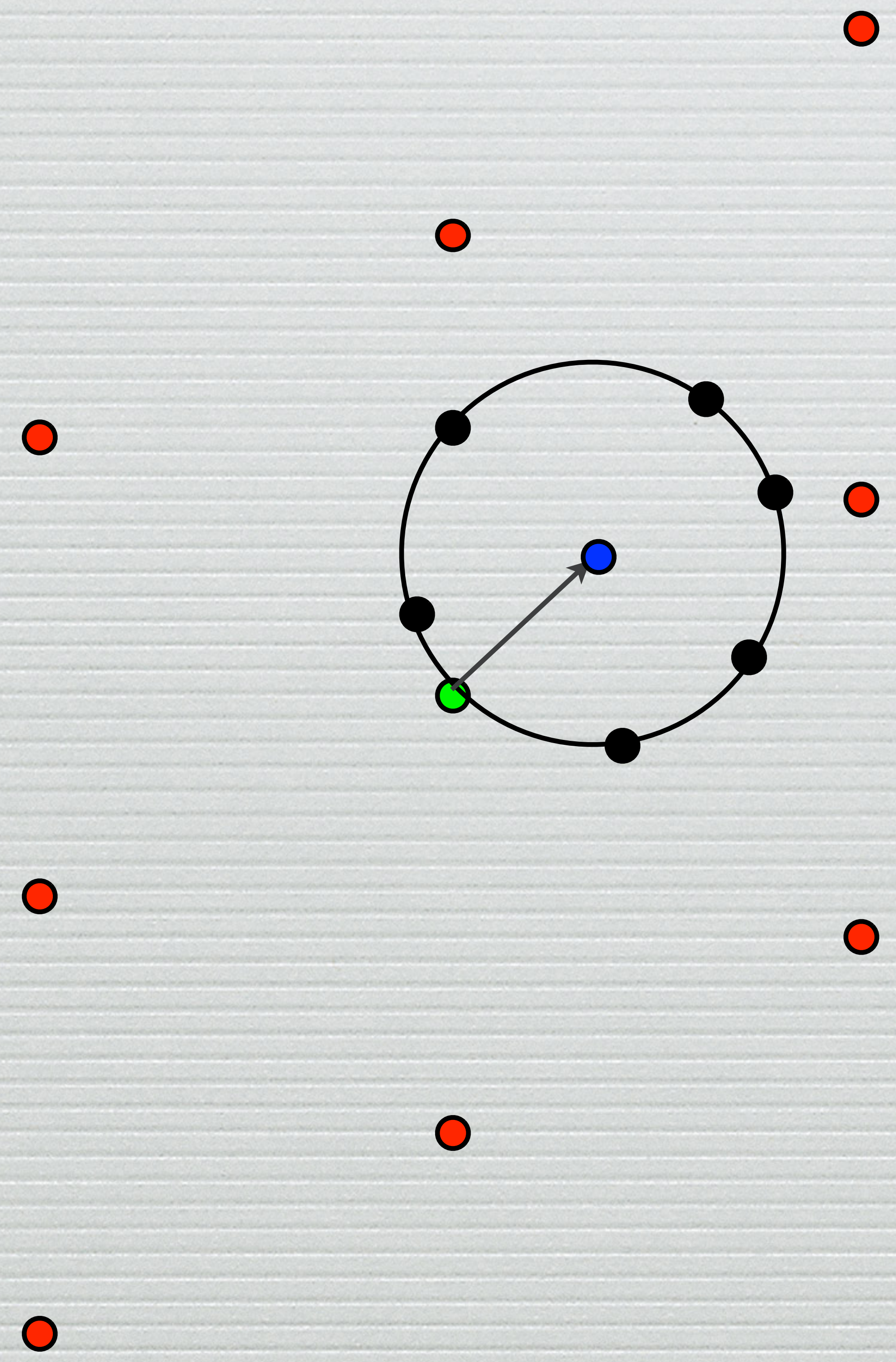


Error correction: decoding



- codewords
- sent codeword
- received word

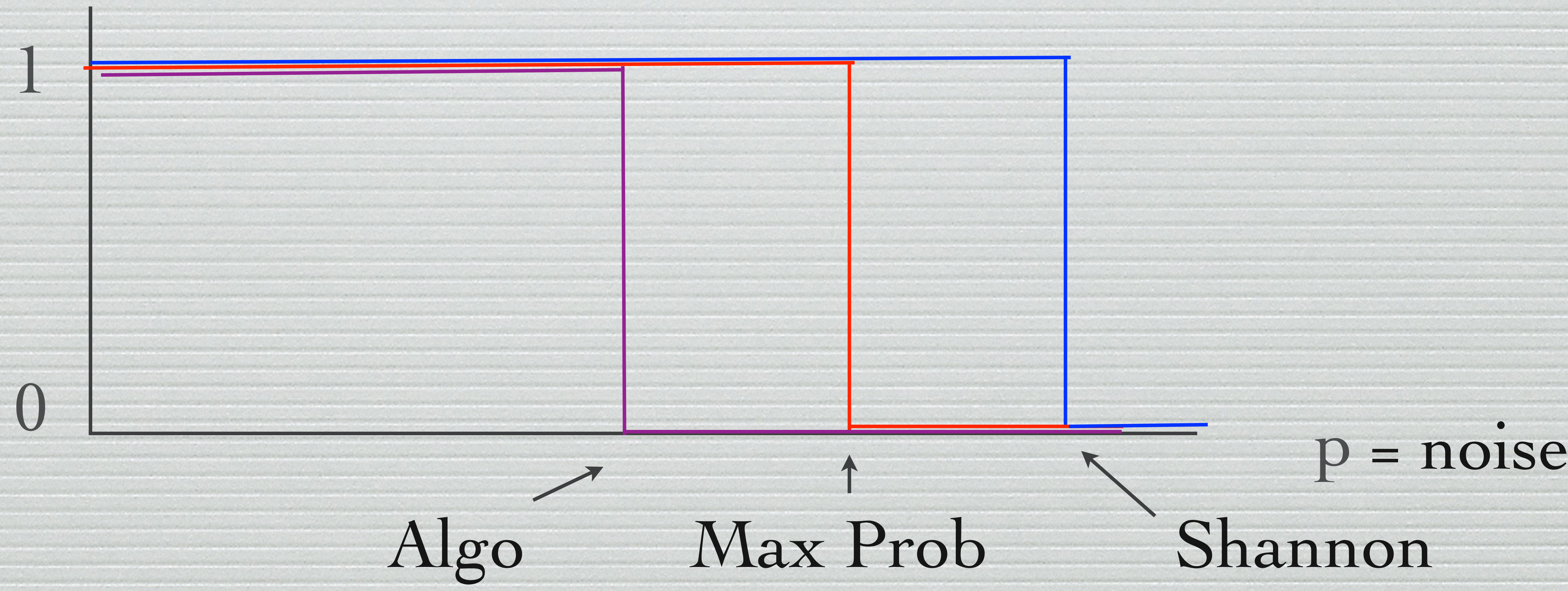
Error correction: decoding



- codewords
- sent codeword
- received word
- metastable states

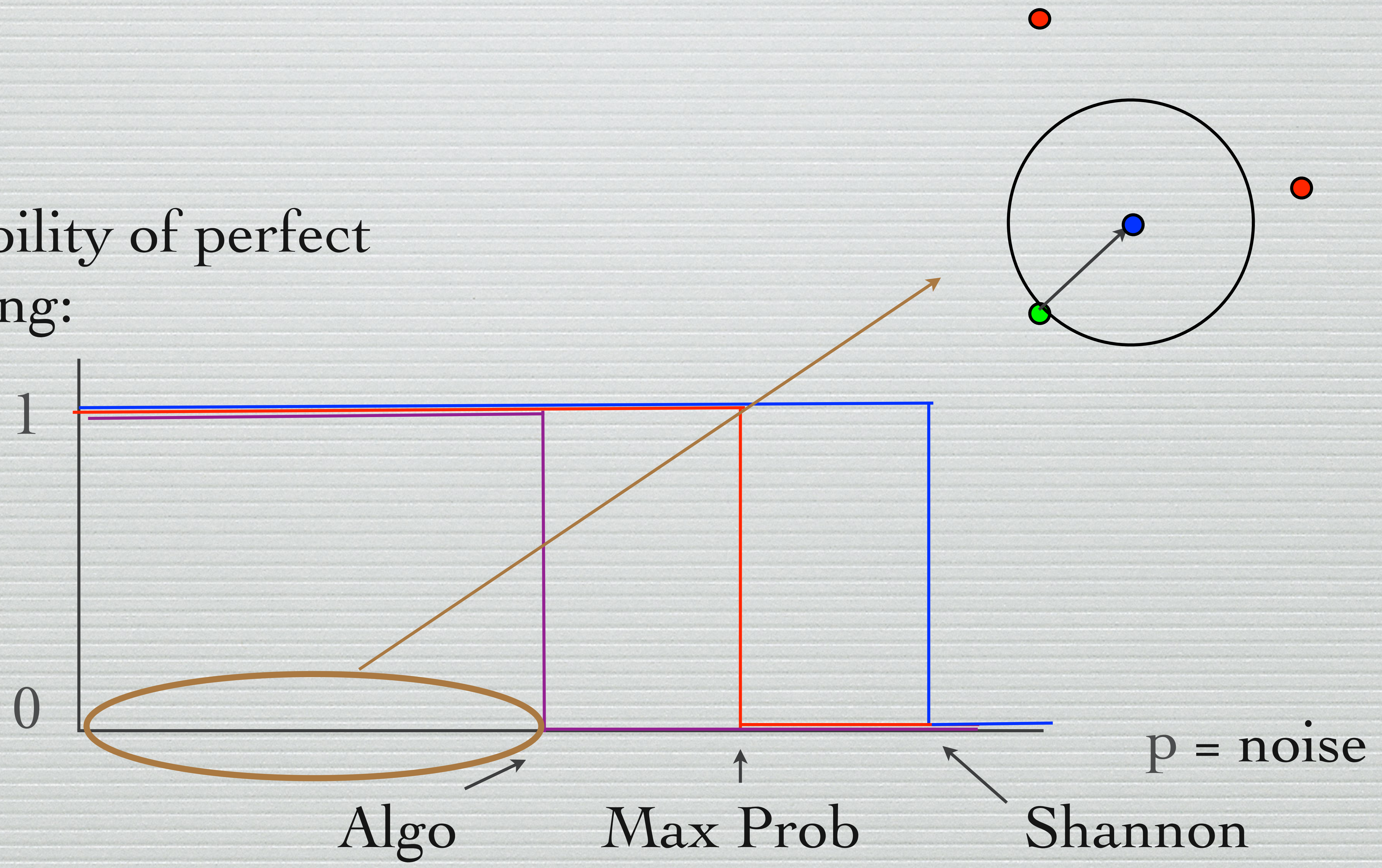
Phase transitions in decoding

Probability of perfect decoding:



Phase transitions in decoding

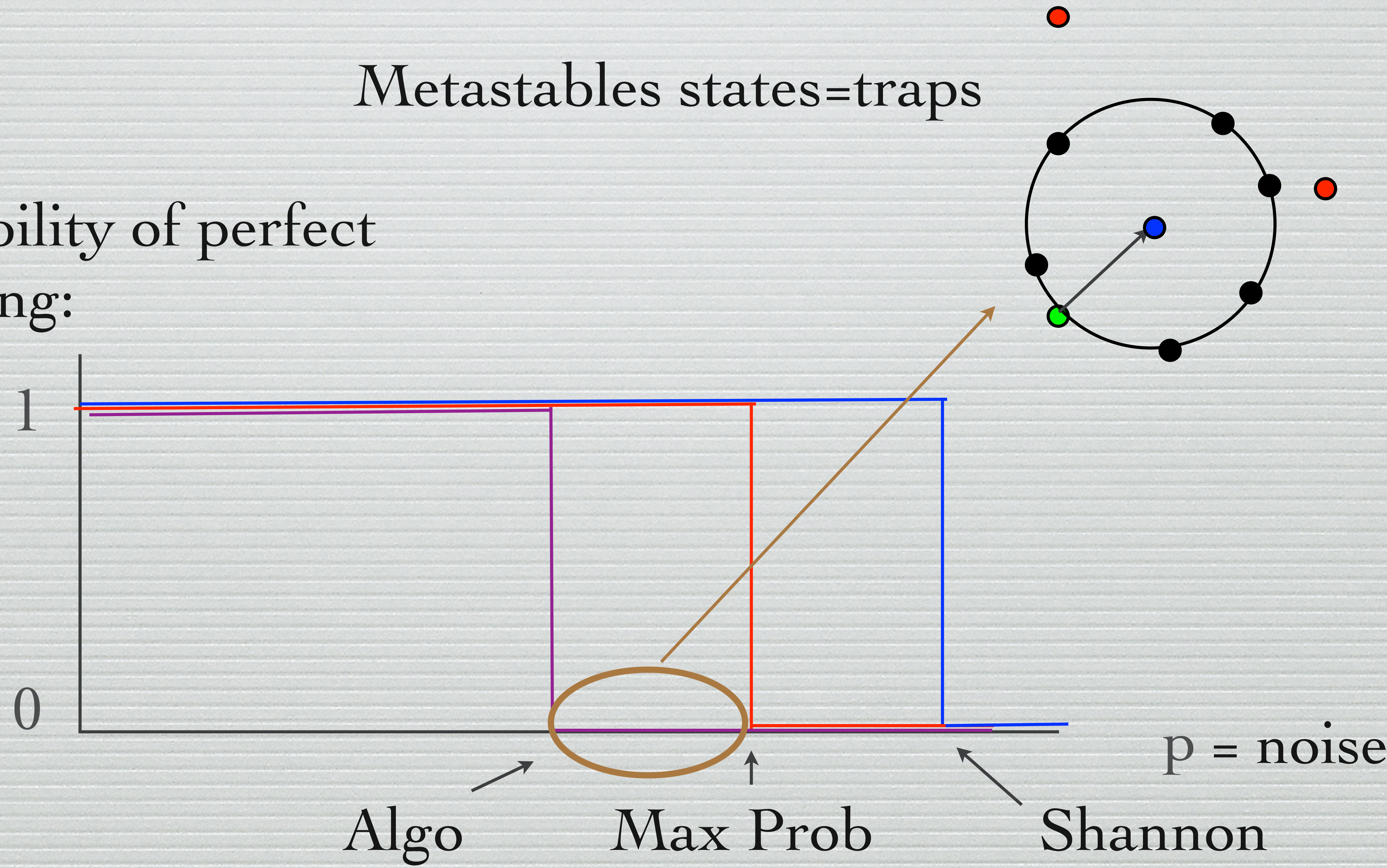
Probability of perfect decoding:



Phase transitions in decoding

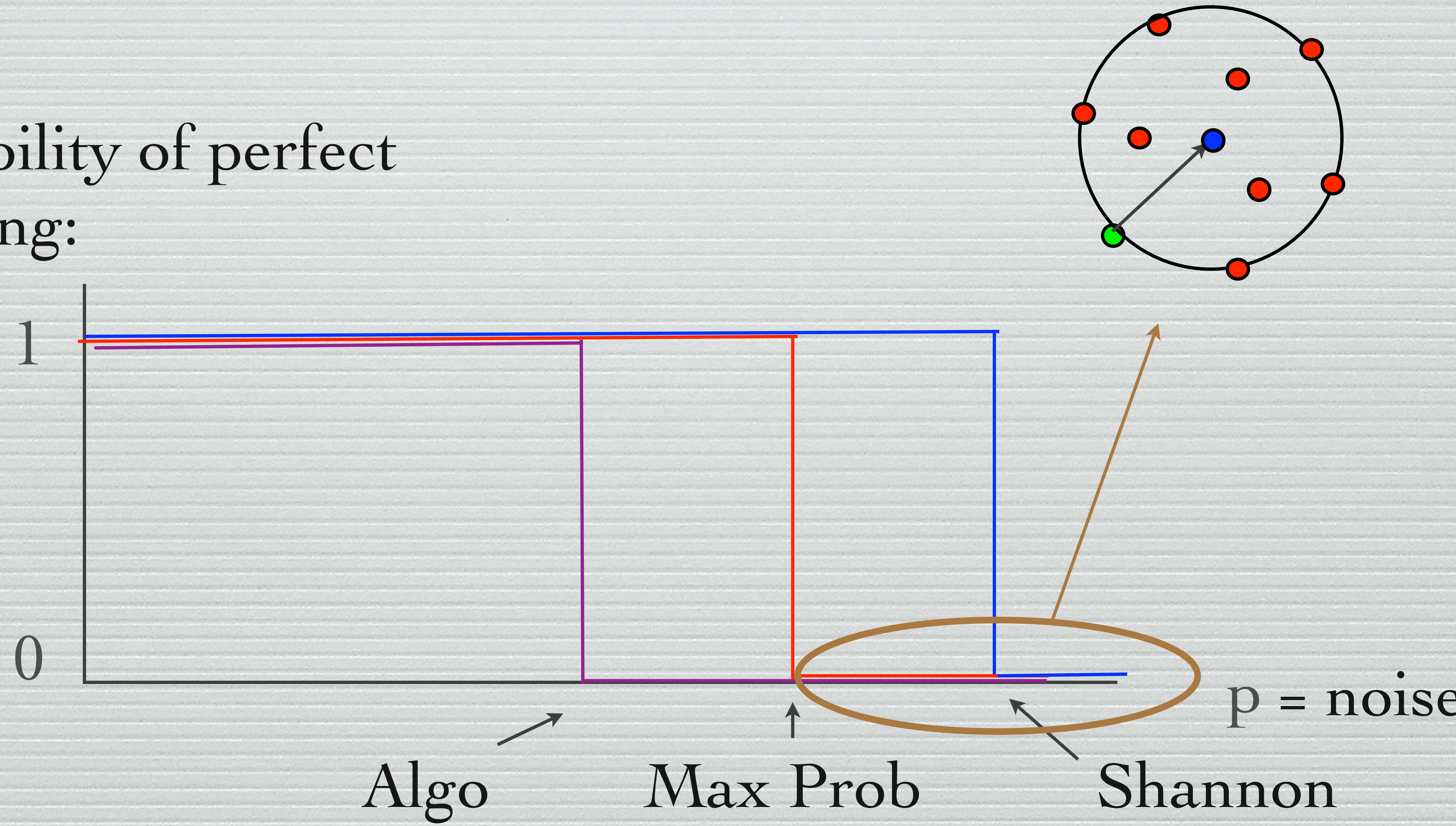
Metastables states=traps

Probability of perfect decoding:

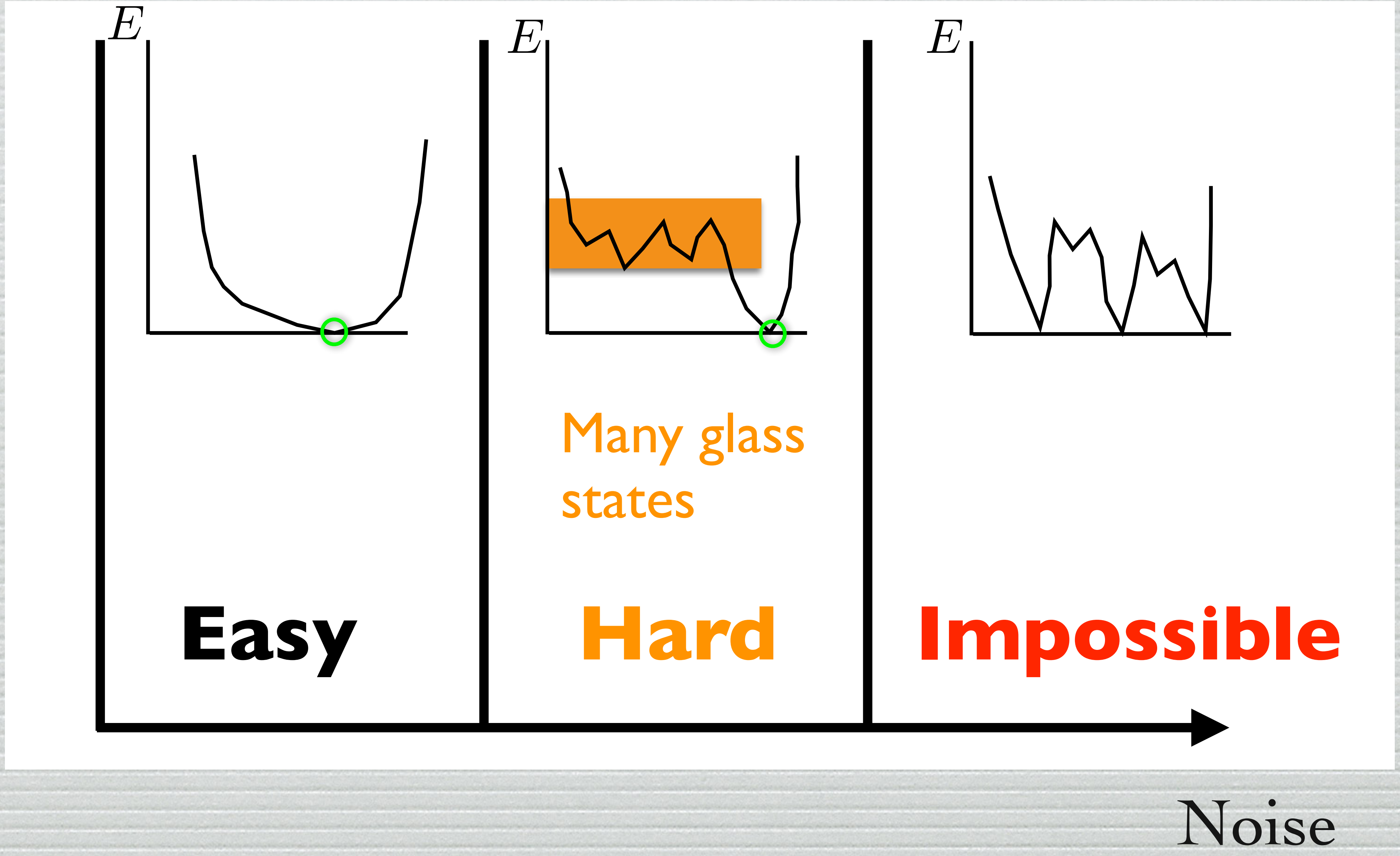


Phase transitions in decoding

Probability of perfect decoding:



Phase transitions in decoding



Statistical inference: general scheme

Challenge = rules with **many hidden parameters.** eg : machine learning with large machine and big data, decoding in communication,…

$x = (x_1, \ldots, x_N) \qquad N \gg 1$

Many measurements $\qquad y = (y_1, \ldots, y_M) \qquad M \gg 1$

Measure of the amount of data $\qquad \alpha = M/N$

→ **Algorithms**

→ **Prediction on the quality of inference**, on the performance of the algorithms, on the type of situations where they can be applied

Bayesian inference with many unknown and many measurements

Unknown parameters $x = (x_1, \ldots, x_N)$ Prior $P^0(x)$

Measurements $y = (y_1, \ldots, y_M)$ $P(y|x)$

Bayesian inference $$P(x|y) \propto P(y|x) P^0(x)$$

Often (but not necessarily):

Independent measurements $$P(y|x) = \prod_\mu P_\mu(y_\mu|x)$$

Factorized prior $$P^0(x) = \prod_i P_i^0(x_i)$$

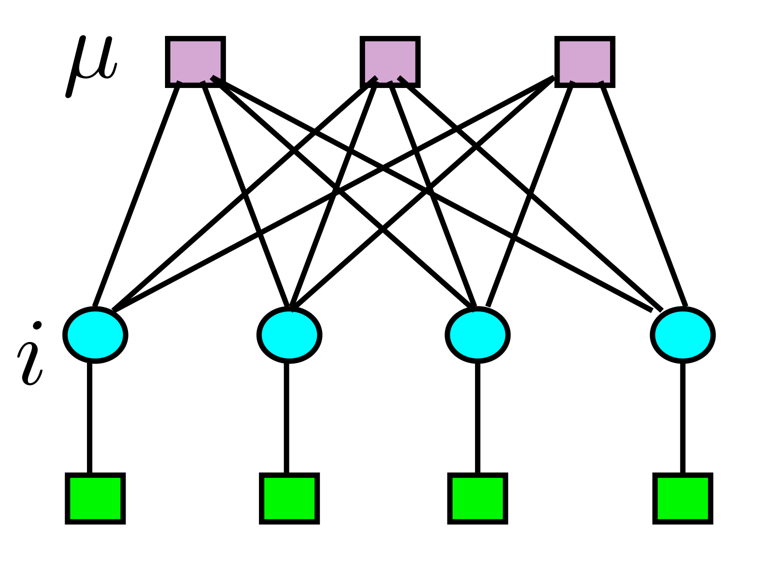
Posterior $$P(x) = \frac{1}{Z(y)} \left(\prod_i P_i^0(x_i) \right) \exp\left[- \sum_\mu E_\mu(x, y_\mu) \right]$$

$$E_\mu(x, y_\mu) = -\log P_\mu(y_\mu|x)$$

Bayesian inference with many unknown and many measurements

$$P(x) = \frac{1}{Z(y)} \left(\prod_i P_i^0(x_i) \right) \exp\left[- \sum_\mu E_\mu(x, y_\mu) \right]$$

$$E_\mu(x, y_\mu) = -\log P_\mu(y_\mu | x)$$



Statistical mechanics (inverse problem).

Example: Inverse Ising

$y_\mu = \{s_1^\mu, \ldots, s_N^\mu\}$ = configuration N Ising spins

$x = \{x_{rs}\}$ =coupling constants between spins r and s

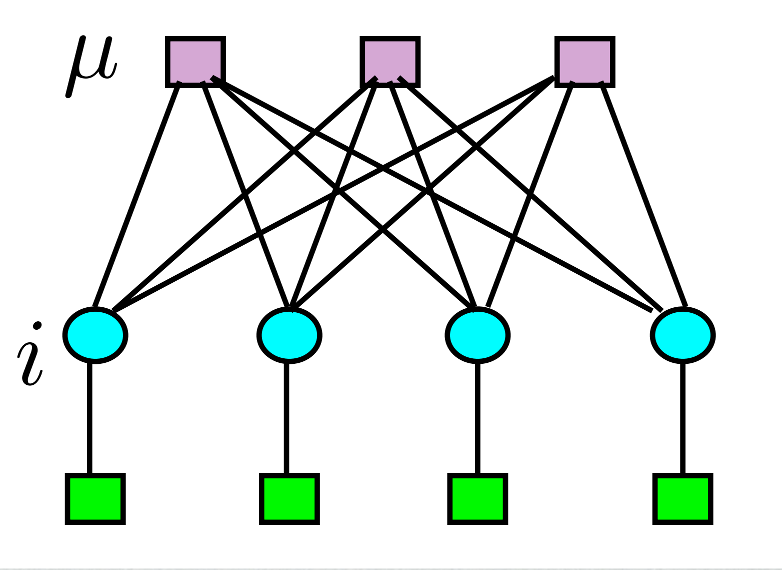
Boltzmann distribution: $P(y|x) = \frac{1}{Z(x)} e^{-\sum_{r,s} x_{rs} s_r s_s}$

Reconstruct the couplings from typical spin configuration

Bayesian inference with many unknown and many measurements

$$P(x) = \frac{1}{Z(y)} \left(\prod_i P_i^0(x_i) \right) \exp\left[-\sum_\mu E_\mu(x, y_\mu) \right]$$

$$E_\mu(x, y_\mu) = -\log P_\mu(y_\mu | x)$$



Statistical mechanics (inverse problem).

✦Discrete or continuous variables x_i

✦Interactions through $e^{-E_\mu(x, y_\mu)}$ can be

- pairwise : $E_\mu = J_\mu x_{i(\mu)} x_{j(\mu)}$
- multibody

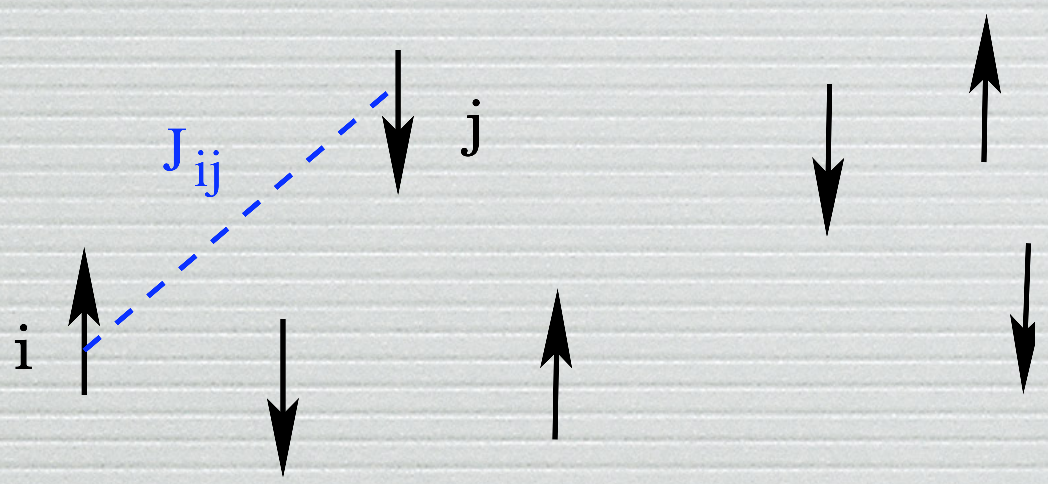
✦Disordered system, ensemble

✦Thermodynamic limit, phase transitions

« Spin glass »

Spin glasses

- Disordered magnetic systems e.g.: CuMn



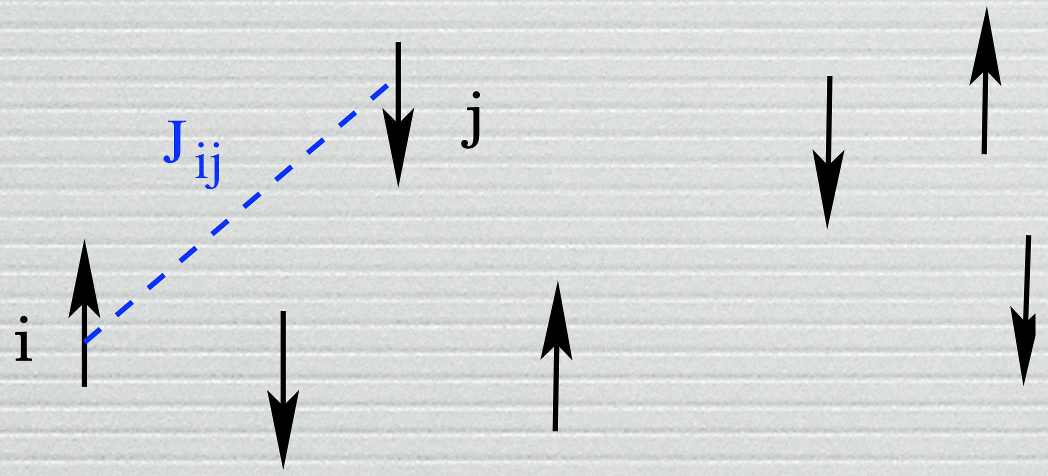
$$s_i = \pm 1$$

$$E = -\sum_{i,j} J_{ij} s_i s_j$$

$$P(s_1, \ldots, s_N) = \frac{1}{Z} e^{-E/T}$$

Spin glasses

- Disordered magnetic systems e.g.: CuMn



$$s_i = \pm 1$$

$$E = -\sum_{i,j} J_{ij} s_i s_j$$

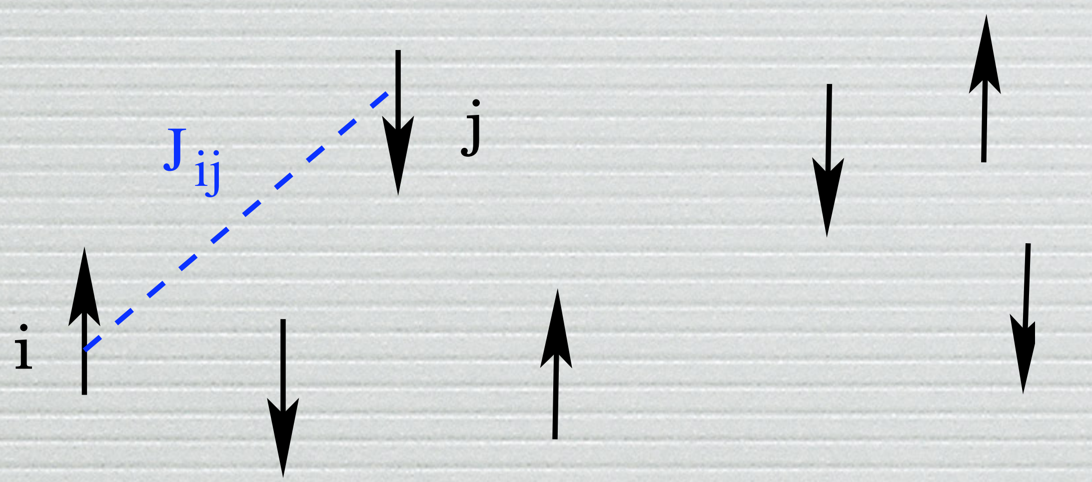
$$P(s_1, \ldots, s_N) = \frac{1}{Z} e^{-E/T}$$

➡ Each spin 'sees' a different local field

Phase transition with many states: spin glasses

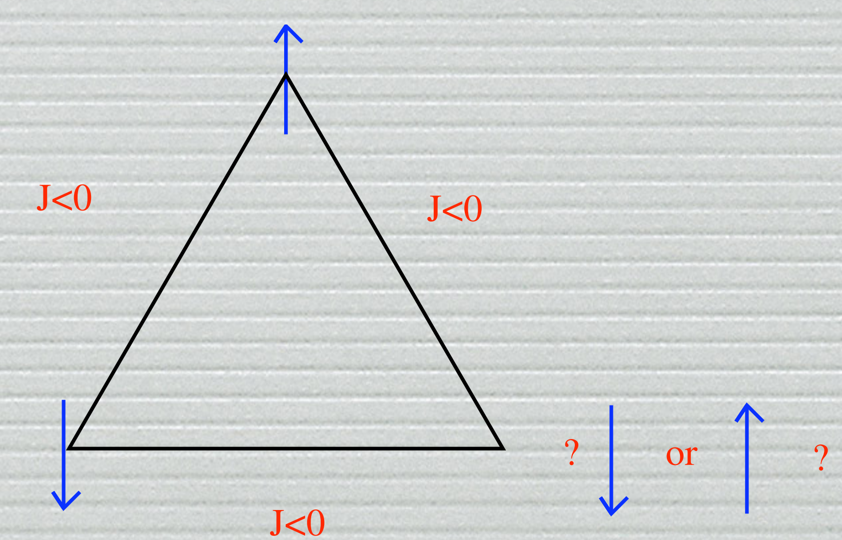
- Many atoms, microscopic interactions are known, "disordered systems"

e.g.: CuMn



➡ Each spin 'sees' a different local field

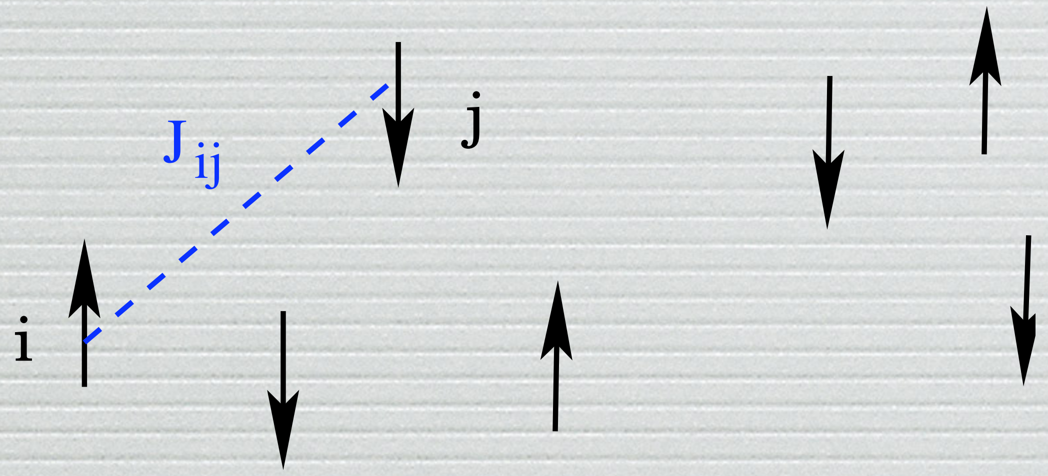
➡ Low temperature: frustration



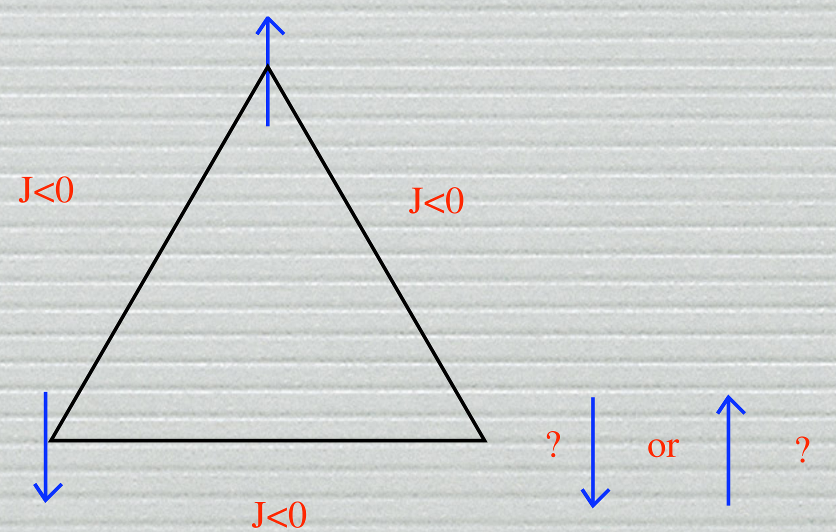
Phase transition with many states: spin glasses

- Many atoms, microscopic interactions are known, "disordered systems"

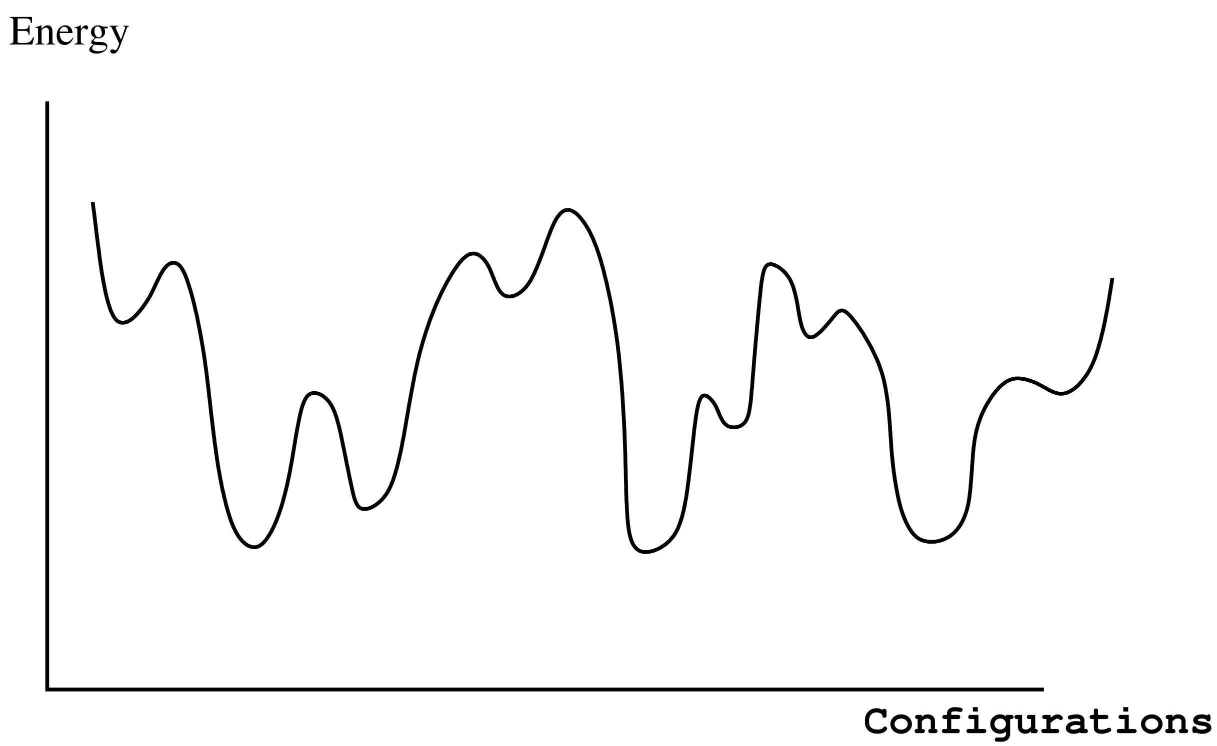
e.g.: CuMn



➡ Each spin 'sees' a different local field
➡ Low temperature: frustration
➡ Spins freeze in random directions
➡ Difficult to find min. of E



Phase transition with many states: spin glasses

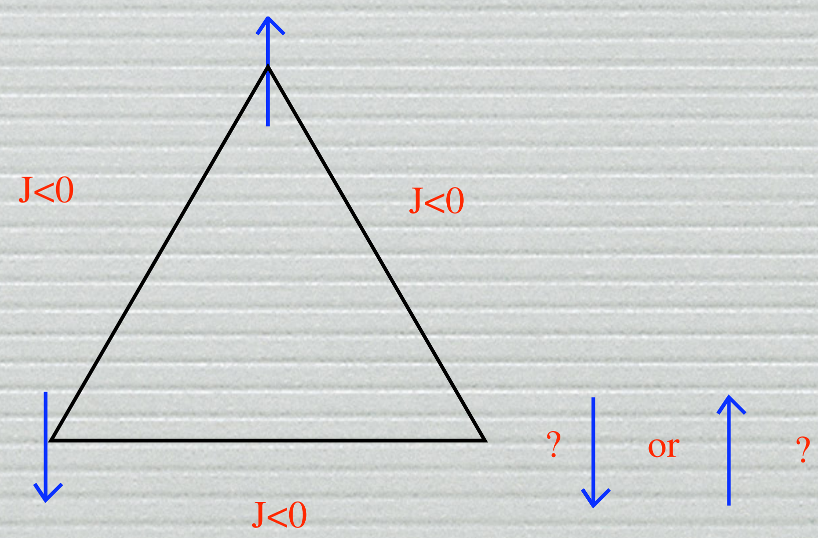


Many quasi-ground states unrelated by symmetries, many metastable states

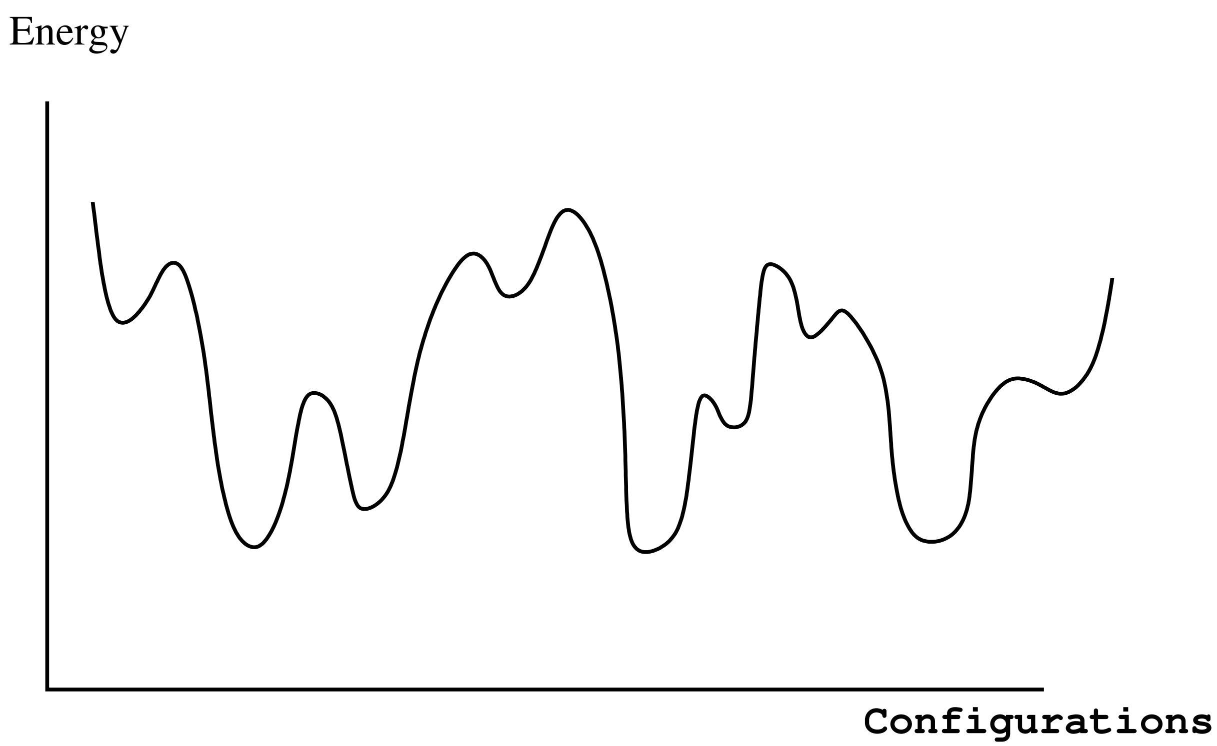
Slow dynamics, aging

Spin glass

- Each spin 'sees' a different local field
- Low temperature: frustration
- Spins freeze in random directions
- Difficult to find min. of E



Phase transition with many states: spin glasses

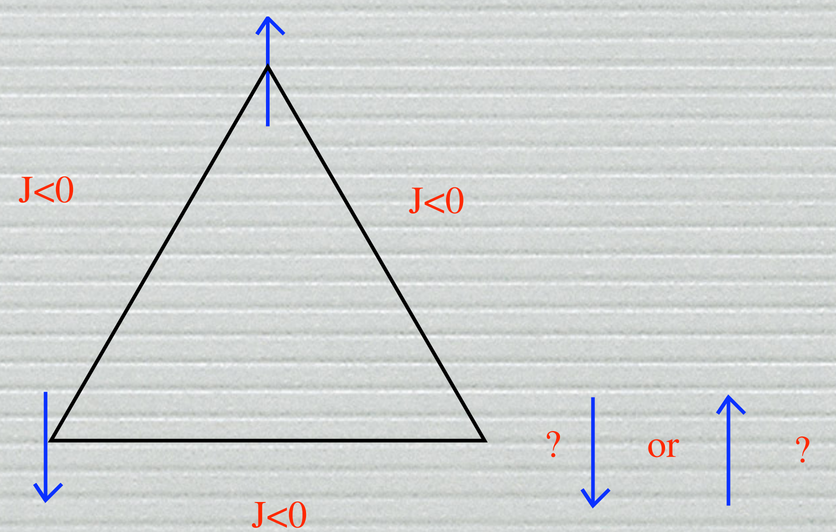


Spin glass

Many quasi-ground states unrelated by symmetries, many metastable states

Slow dynamics, aging

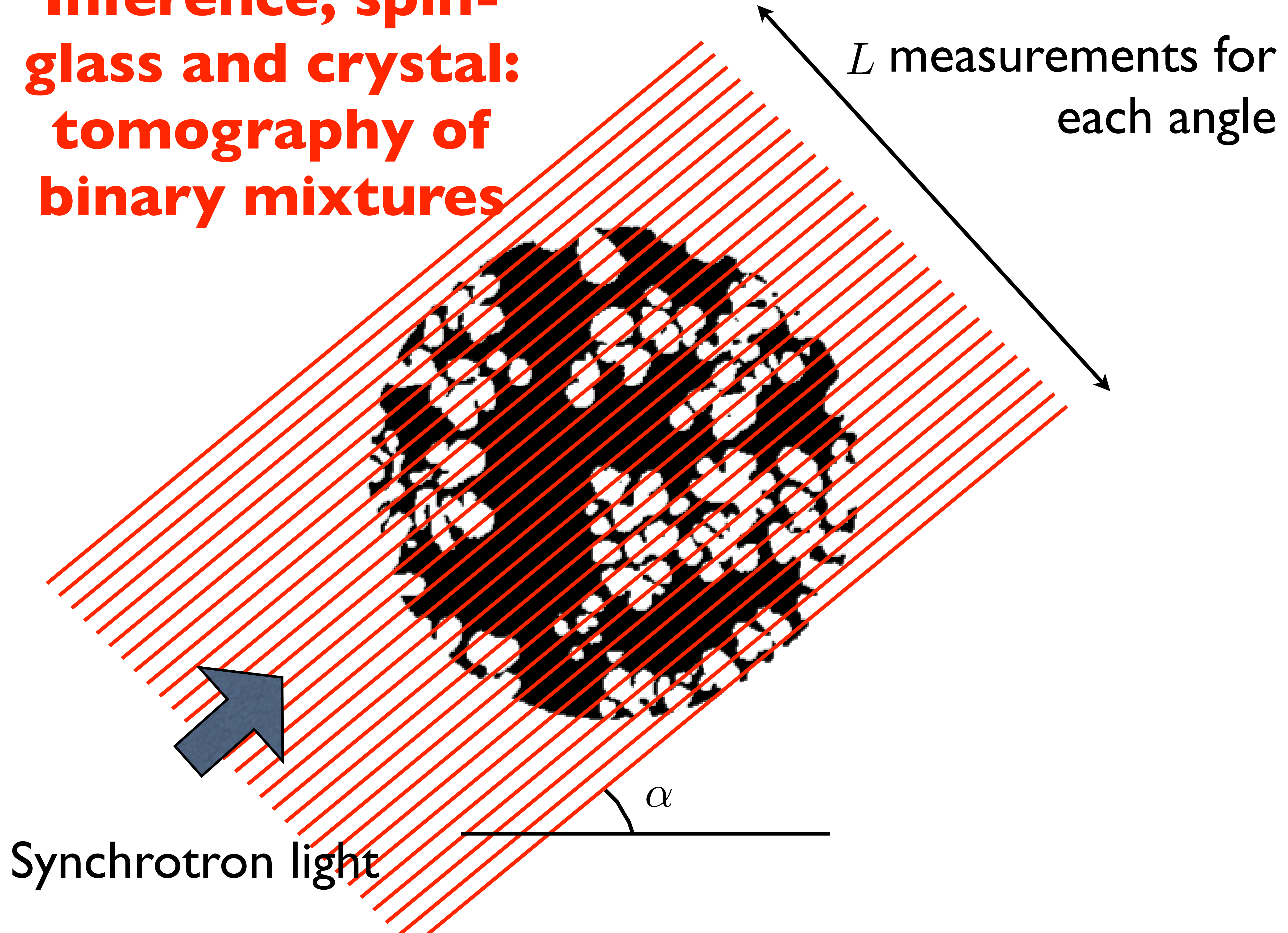
- Each spin 'sees' a different local field
- Low temperature: frustration
- Spins freeze in random directions
- Difficult to find min. of E



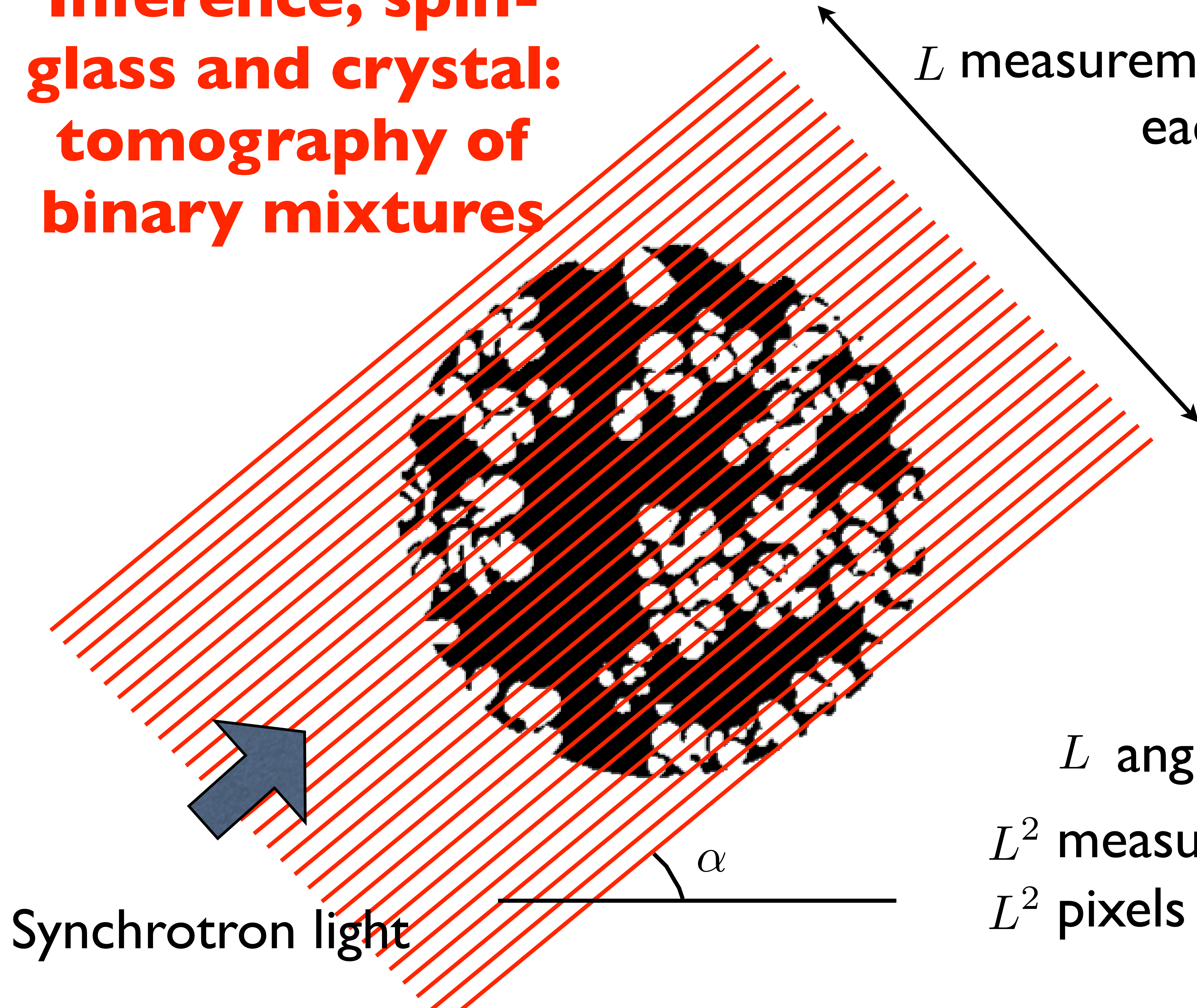
Useless, but thousands of papers...

Inference, spin-glass and crystal: tomography of binary mixtures

Inference, spin-glass and crystal: tomography of binary mixtures



Inference, spin-glass and crystal: tomography of binary mixtures



L angles:

L^2 measurements

L^2 pixels

Tomography of binary mixtures

L angles:

L^2 measurements

L^2 pixels

sample size L

pixel size 1

domain size ξ

If the size of domains is $\gg$ pixel: possible to reconstruct with $\ll L^2$ measurements

$\xi \gg 1$

Tomography of binary mixtures

sample size L

pixel size 1

domain size ξ

If the size of domains is $\gg$ pixel: possible to reconstruct with $\ll L^2$ measurements

$\xi \gg 1$

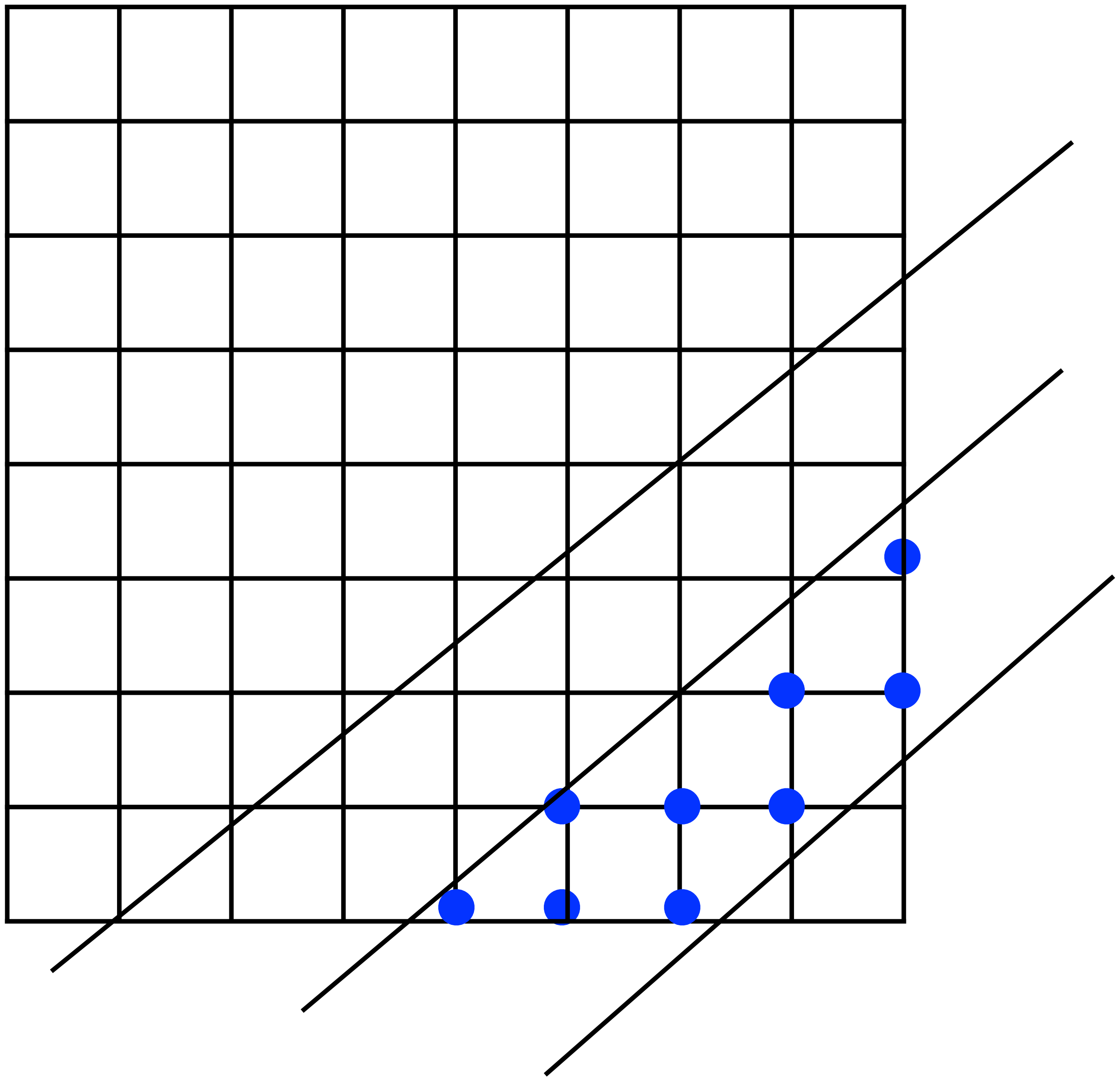
Tomography of binary mixtures

This picture, digitalized on 1000×1000 grid, can be reconstructed fom measurements with 16 angles

Compressed sensing

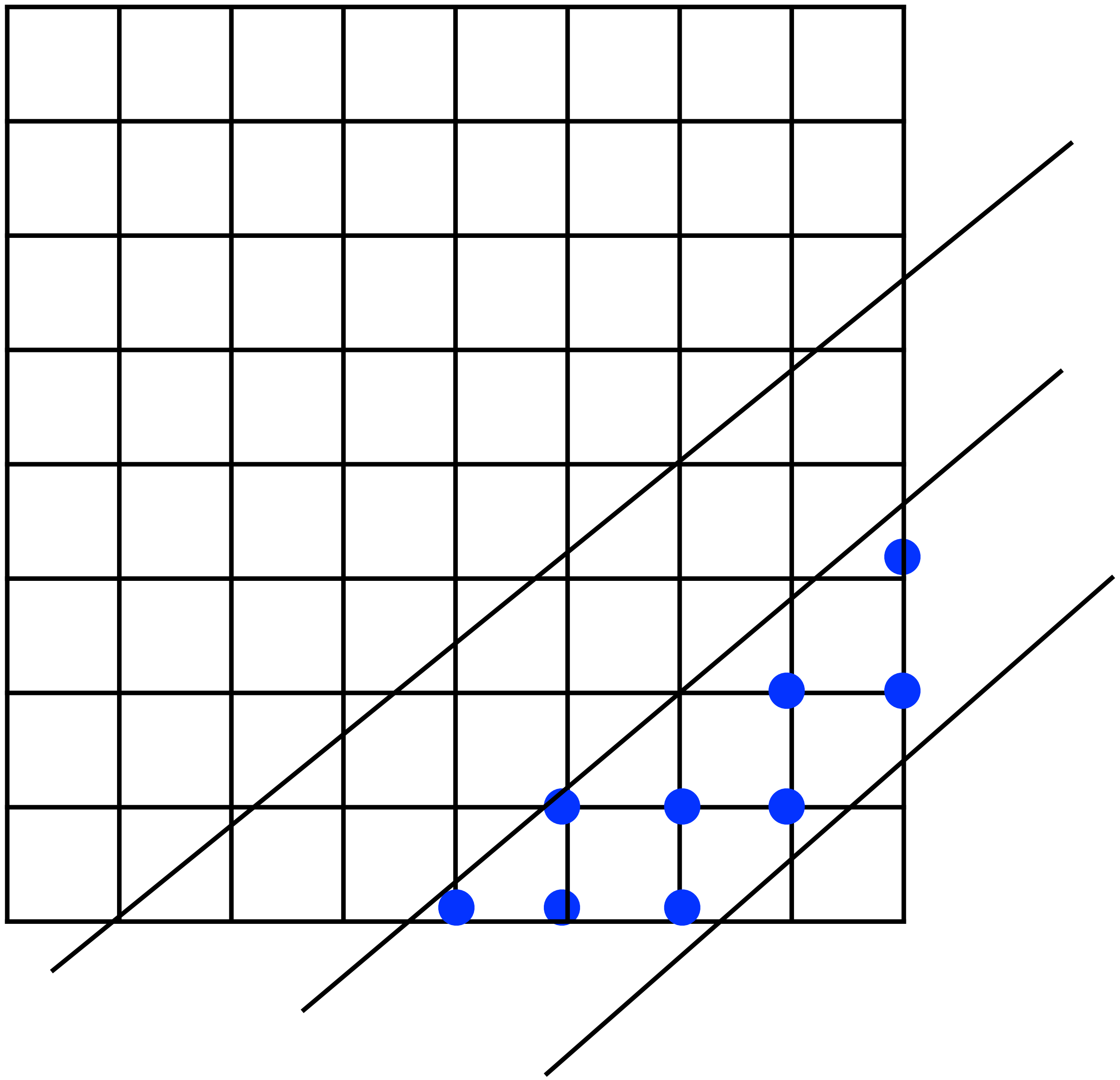
Gouillart et al., Inverse problems 2013

If the size of domains is $\gg$ pixel: possible to reconstruct with $\ll L^2$ measurements



μ $\qquad y_\mu = \sum_{i \in \partial\mu} s_i$

Prior knowledge on $\{s_i\}$: neighboring pixels more likely to be equal



μ $$y_\mu = \sum_{i \in \partial\mu} s_i$$

Prior knowledge on $\{s_i\}$: neighboring pixels more likely to be equal

$$P(S) = \prod_{ij \in \text{grid}} e^{J s_i s_j} \prod_\mu \delta\left(y_\mu, \sum_{i \in \partial\mu} s_i\right)$$

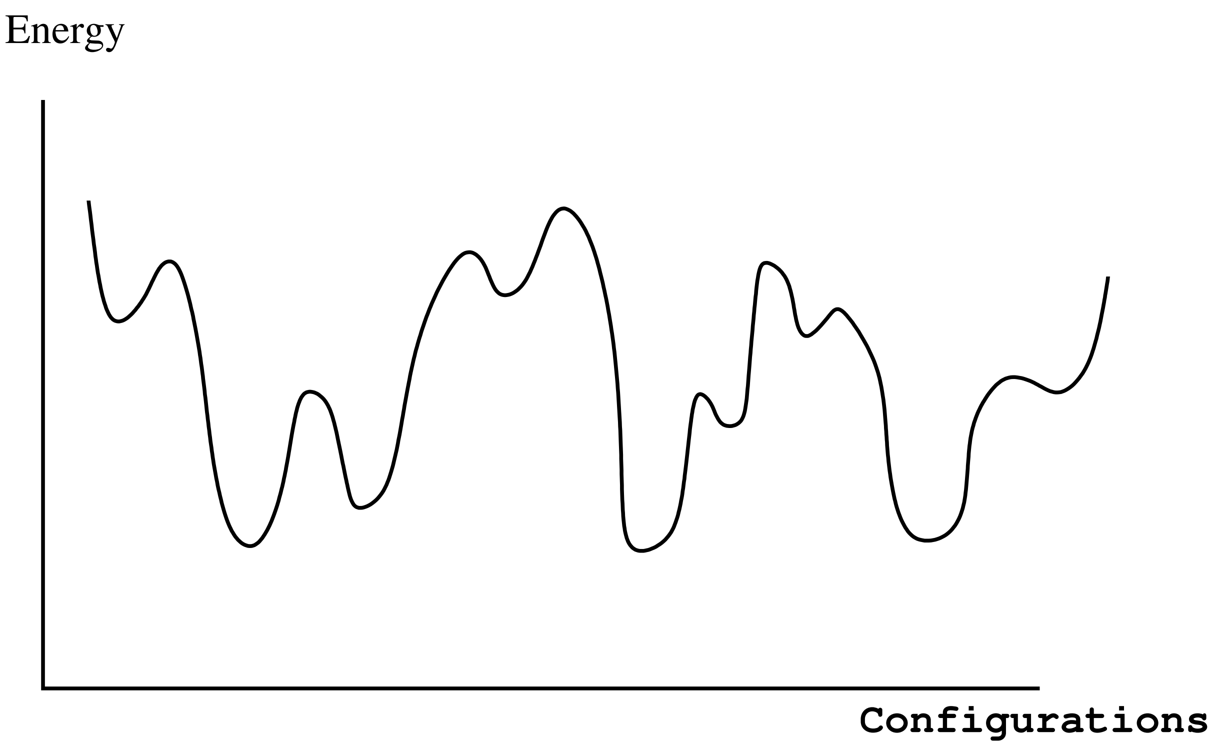
prior measurement

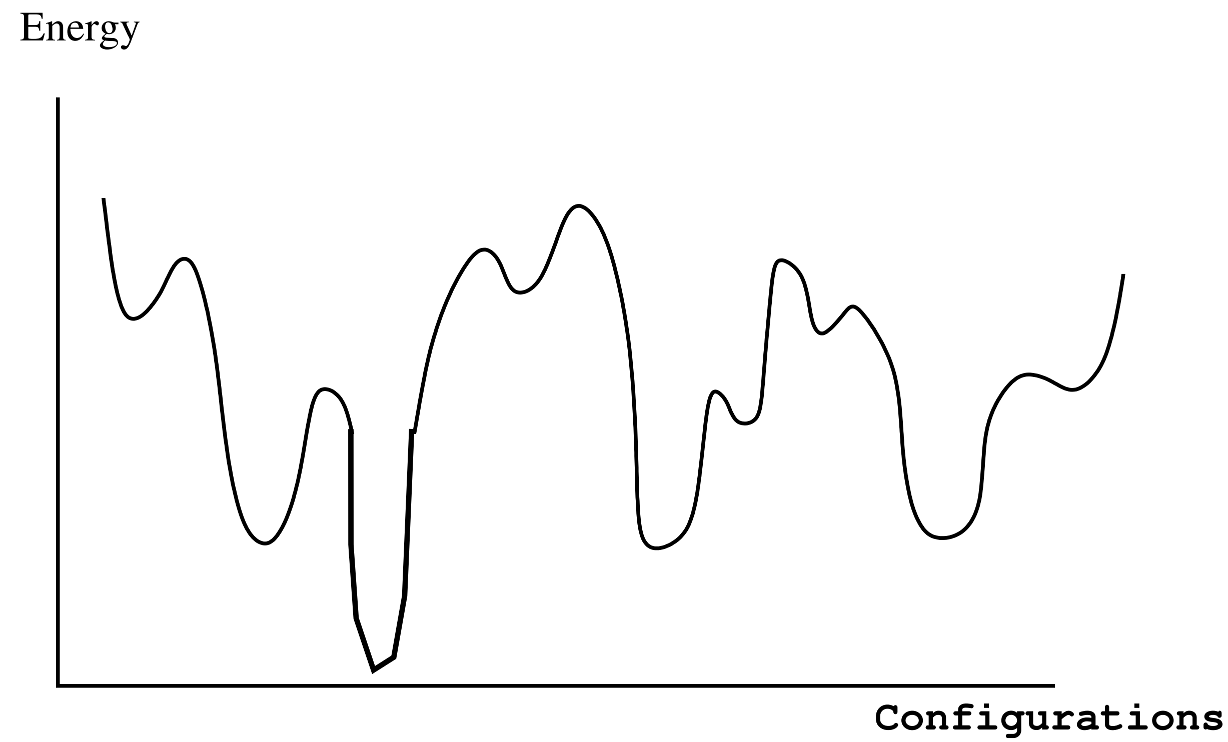
Studied with mean-field

$$P(S) = \prod_{ij \in \text{grid}} e^{J s_i s_j} \prod_{\mu} \delta\left(y_\mu, \sum_{i \in \partial\mu} s_i \right)$$

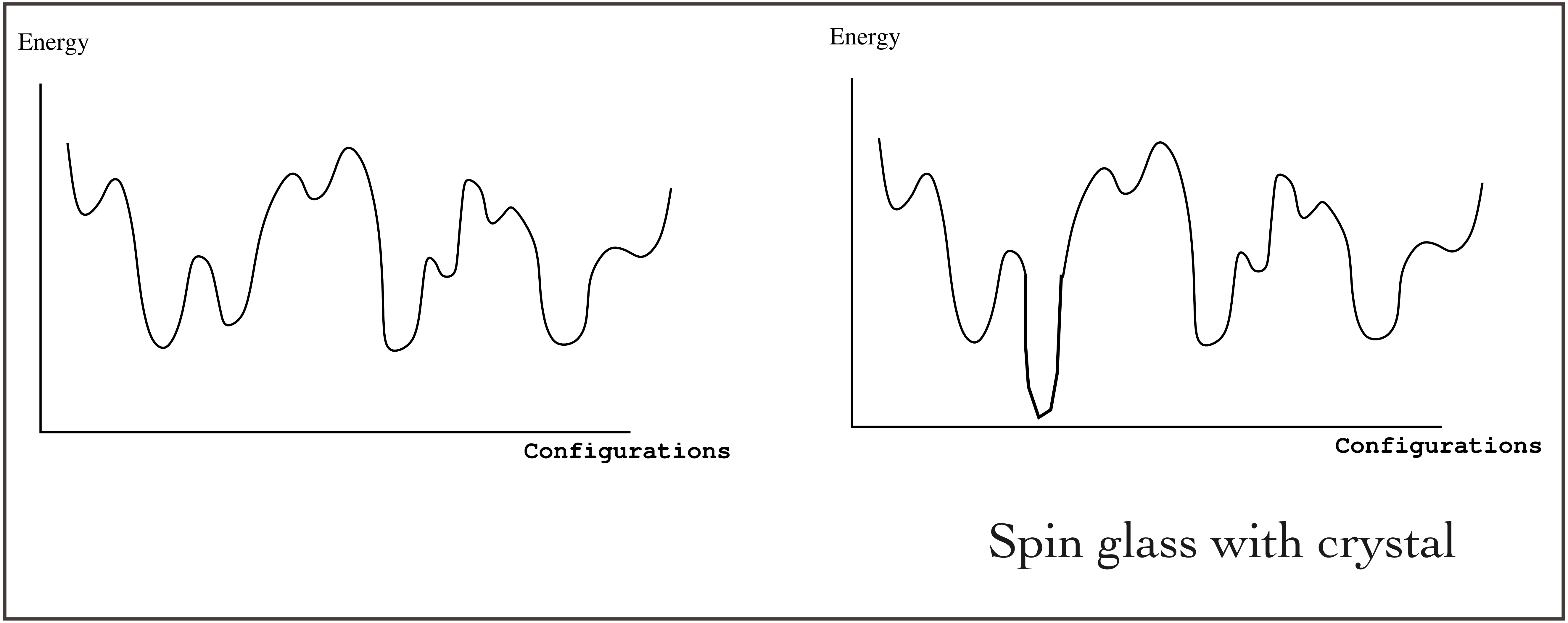
If enough measurements: The most probable S (the ground state) gives the perfect composition of the sample.

« Crystal » : much more probable





Spin glass with crystal



« Crystal » : much more probable

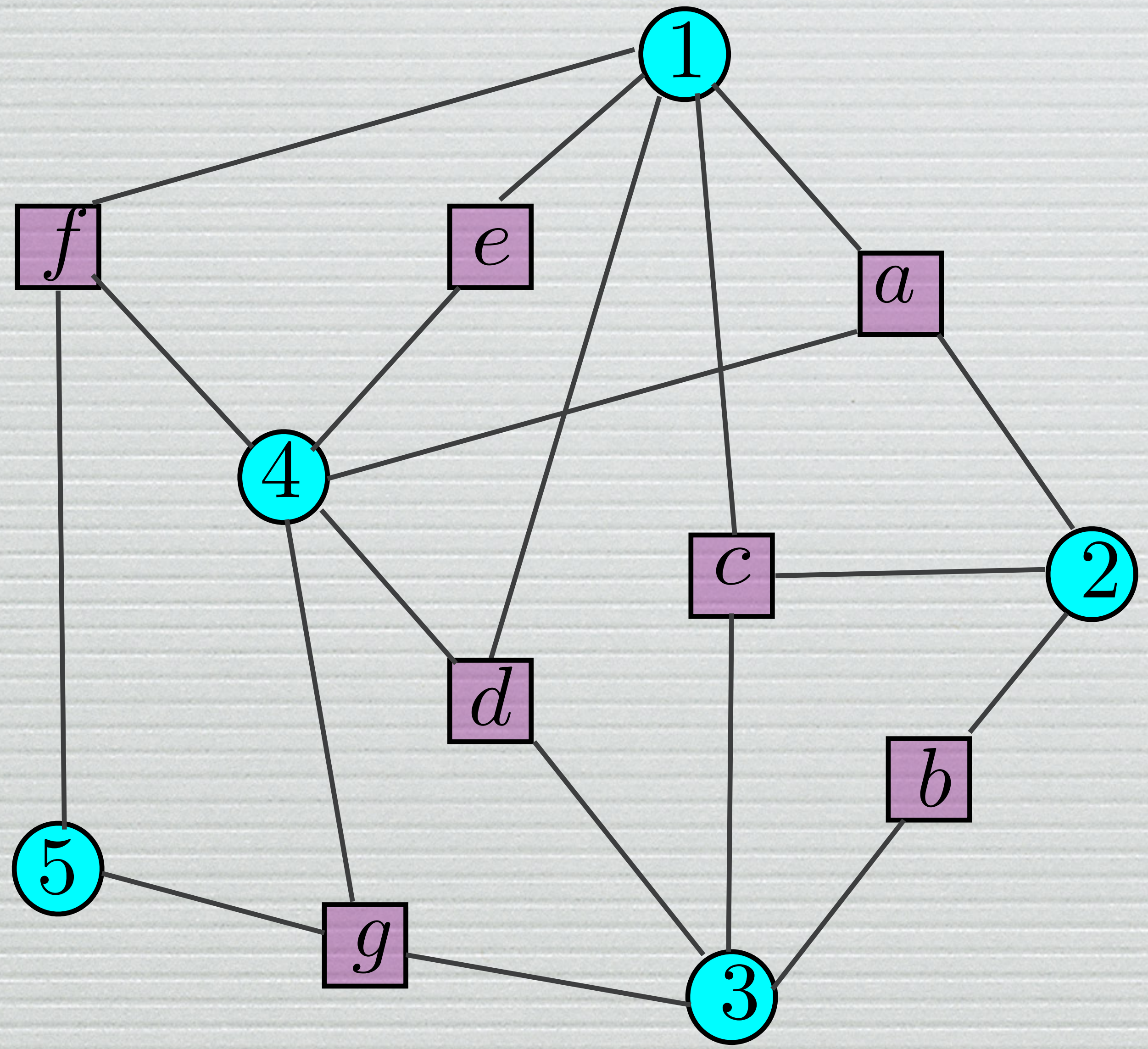
But in some cases « crystal hunting » may be computationally very hard !

Inference with many unknowns : « crystal hunting » with mean-field based algorithms

Historical development of mean field equations

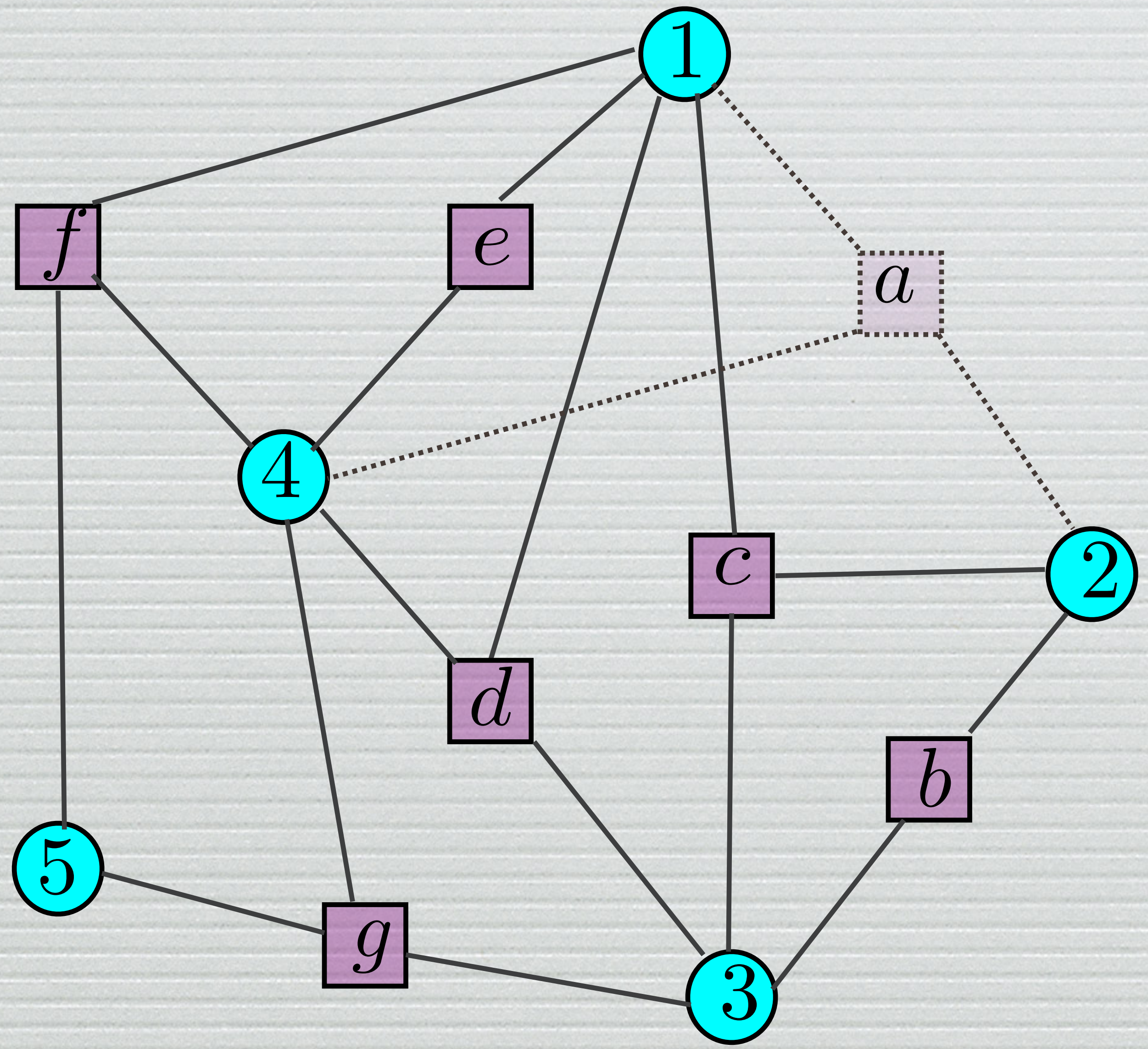
- In homogeneous ferromagnets:
  - Weiss (infinite range, 1907)
  - Bethe Peierls (finite connectivity, 1935)
- In glassy systems:
  - Thouless Anderson Palmer 1977,
  - MM Parisi Virasoro 1986 (infinite range)
  - Kabashima Saad 1998 (finite connectivity)
  - MM Parisi 2001 (finite connectivity)
- As an algorithm:
  - Gallager 1963
  - Pearl 1986
  - MM Parisi Zecchina 2002
  - Kabashima 2003, 2008
  - Donoho Bayati Montanari 2009
  - Rangan 2010
  - Krzakala MM Zdeborova 2012 …

BP = Bethe-Peierls = Belief Propagation



$$P(x_1, \cdots, x_5) = \psi_a(x_1, x_2, x_4)\psi_b(x_2, x_3) \cdots$$

BP equations

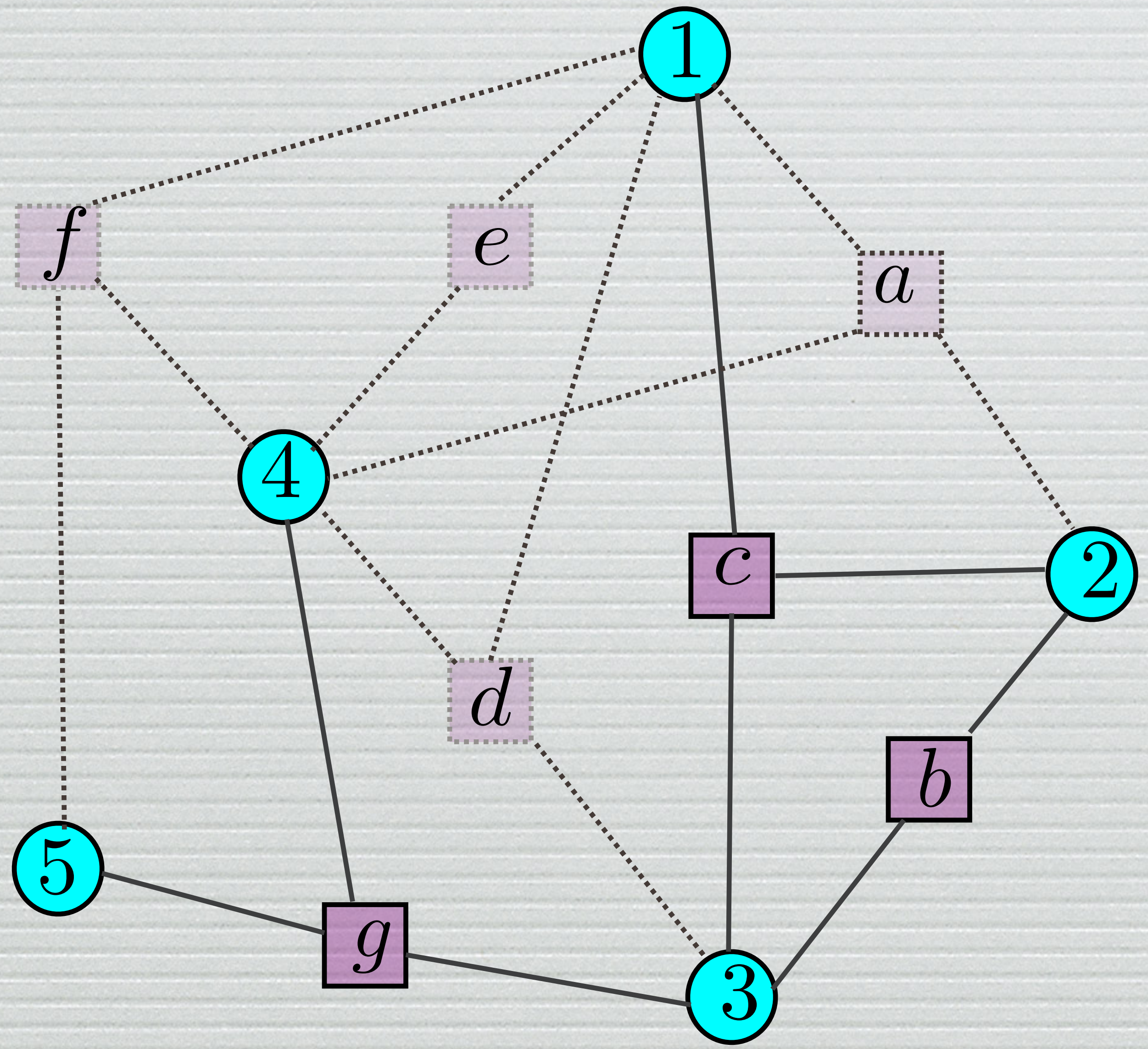


First type of messages:

Probability of x_1 in the absence of a:

$$m_{1\to a}(x_1)$$

BP equations

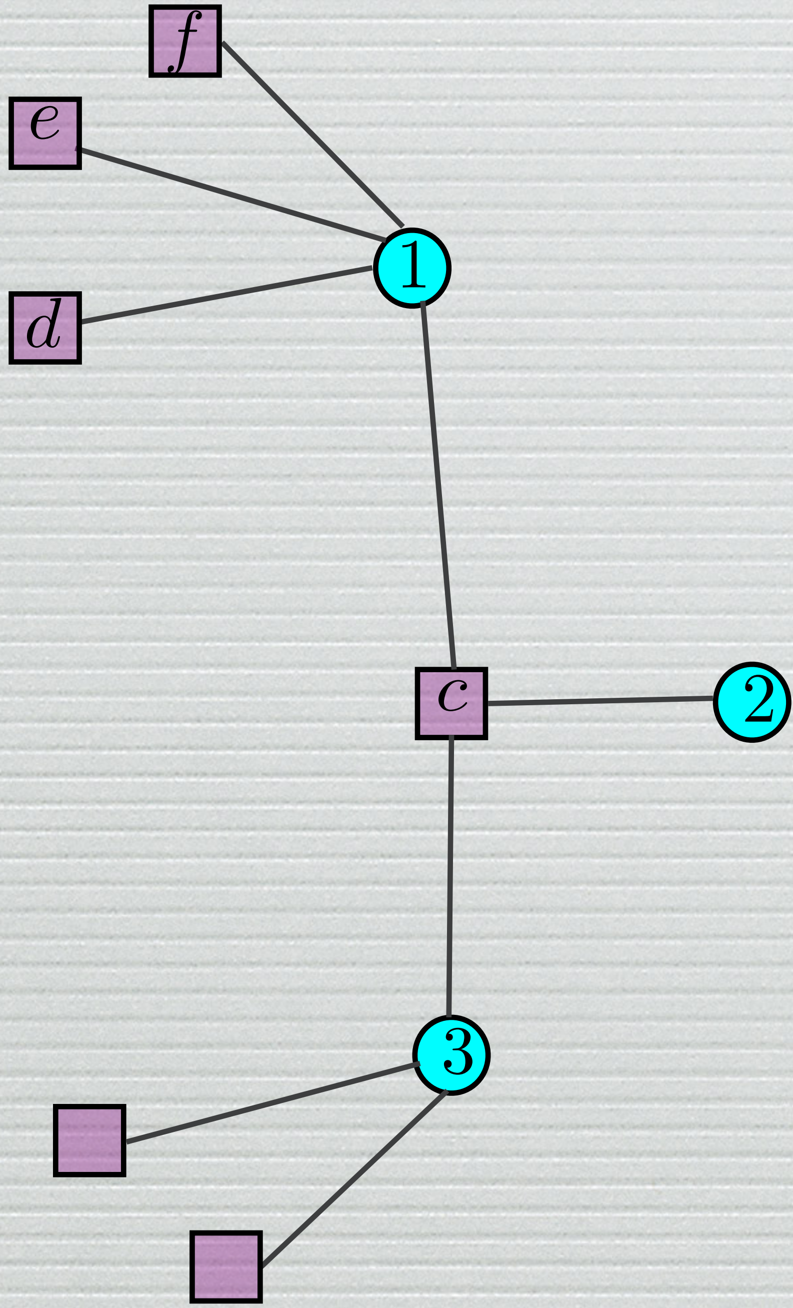


Second type of messages:

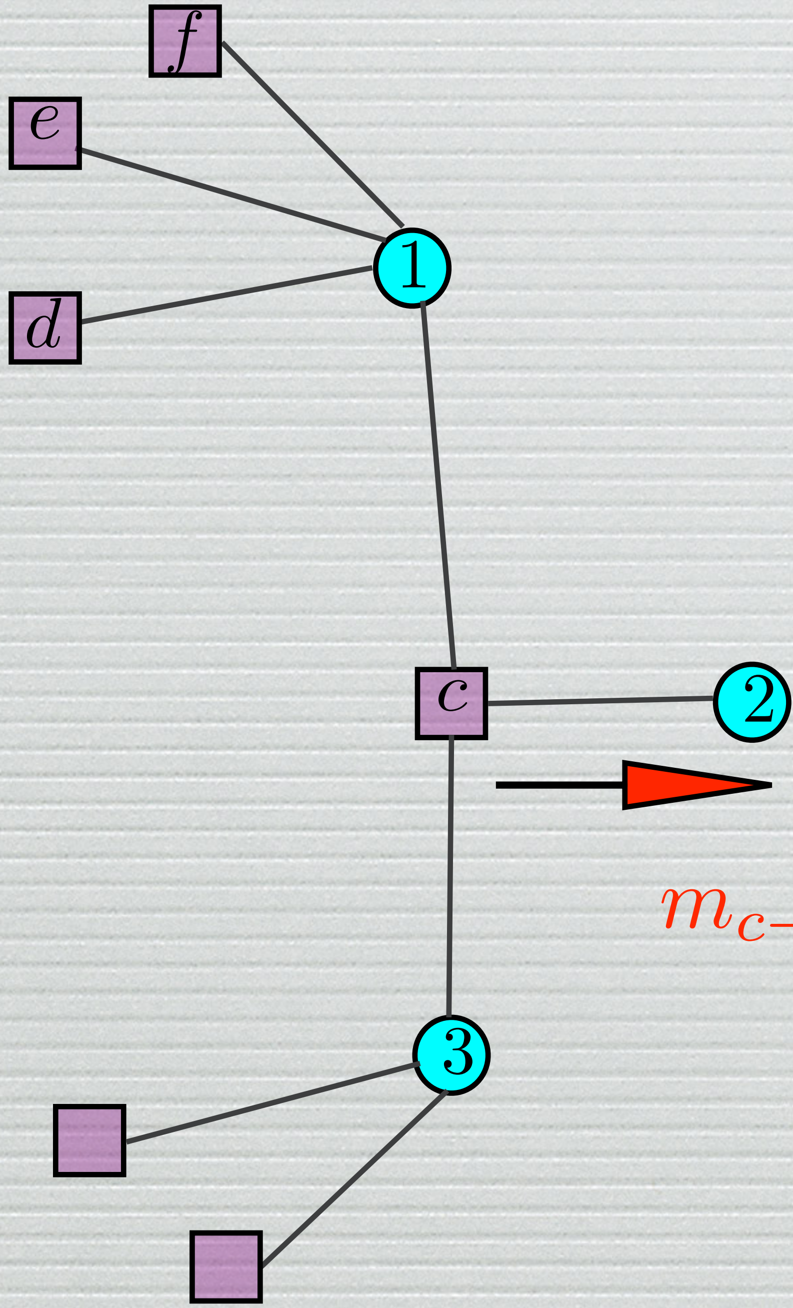
Probability of x_1 when it is connected only to c :

$$m_{c \to 1}(x_1)$$

BP equations

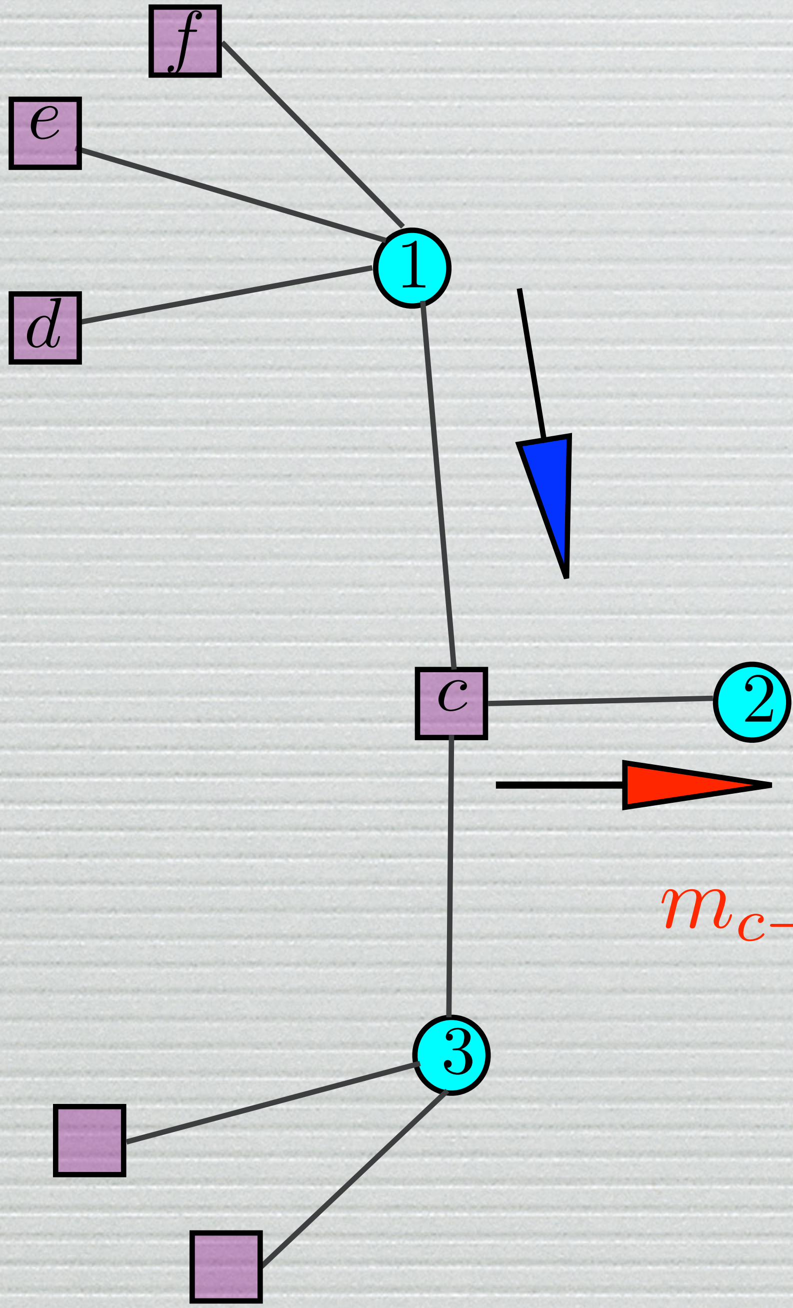


BP equations



$$m_{c\to 2}(x_2) = \sum_{x_1, x_3} \psi_c(x_1, x_2, x_3) m_{1\to c}(x_1) m_{3\to c}(x_3)$$

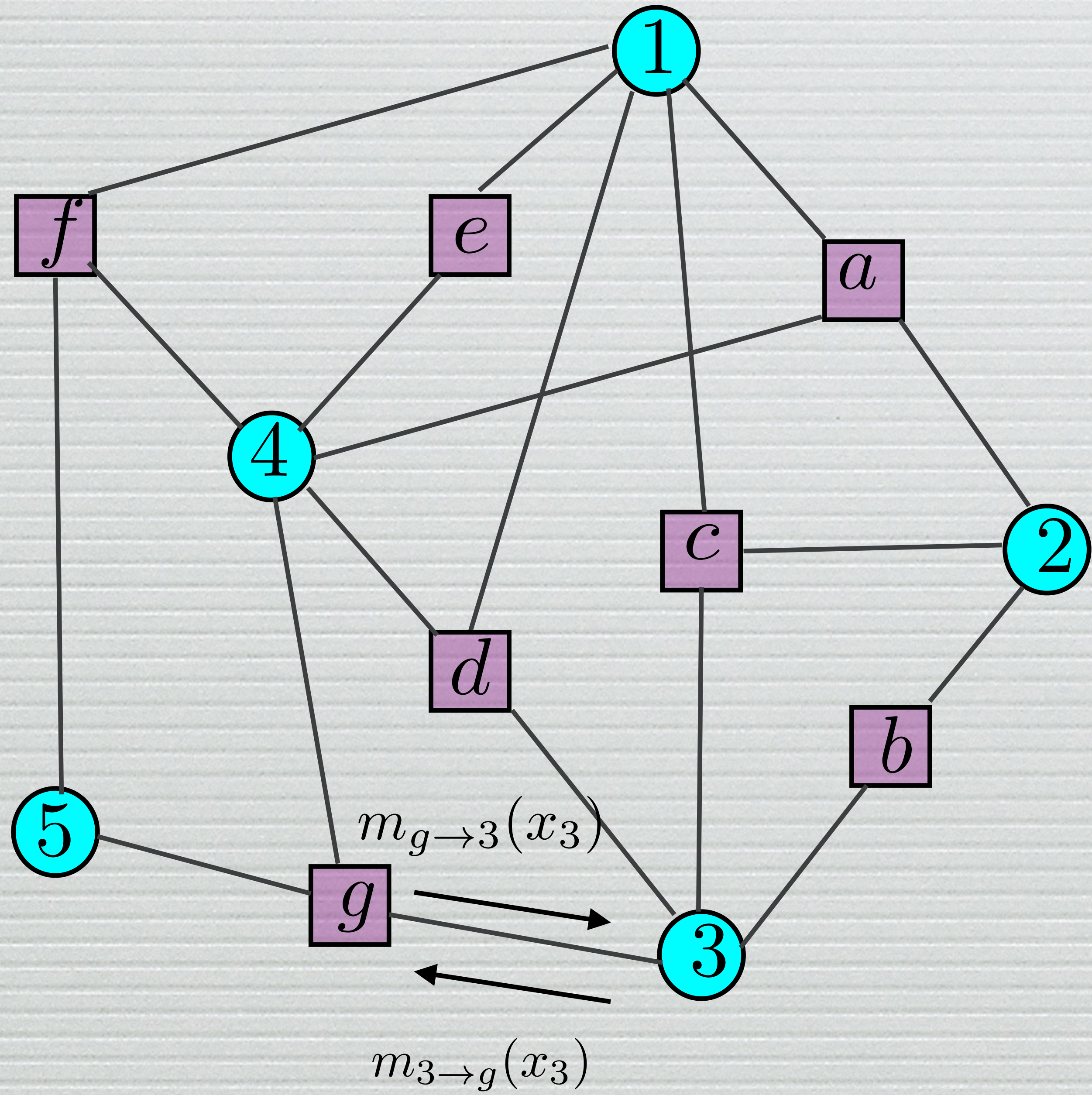
BP equations



$$m_{1\to c}(x_1) = C m_{d\to 1}(x_1) m_{e\to 1}(x_1) m_{f\to 1}(x_1)$$

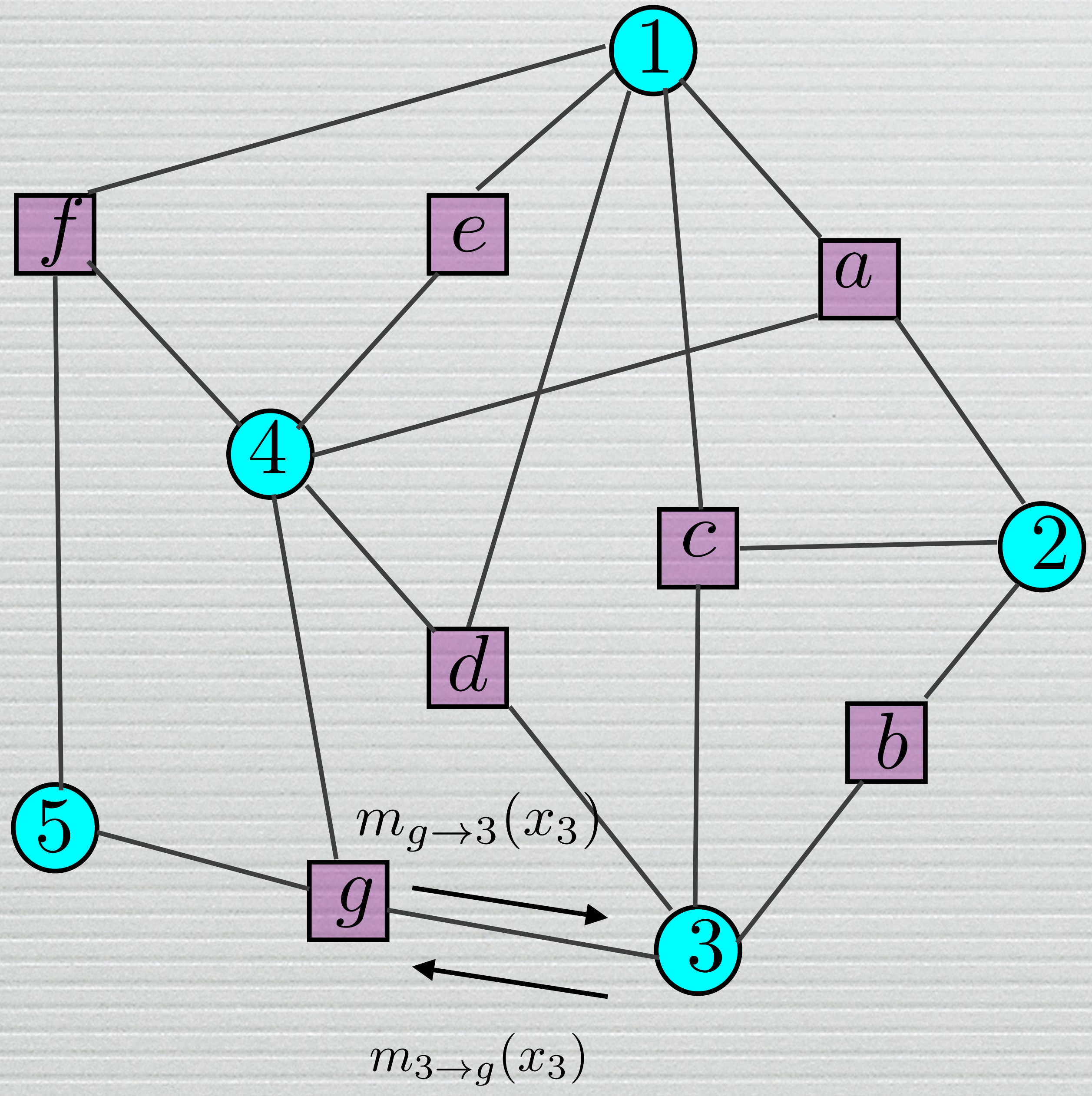
$$m_{c\to 2}(x_2) = \sum_{x_1, x_3} \psi_c(x_1, x_2, x_3) m_{1\to c}(x_1) m_{3\to c}(x_3)$$

BP equations



Propagate messages along the edges, update messages at vertices, using elementary local probabilistic rules

BP equations



Propagate messages along the edges, update messages at vertices, using elementary local probabilistic rules

Closed set of equations: two messages "propagate" on each edge of the factor graph.

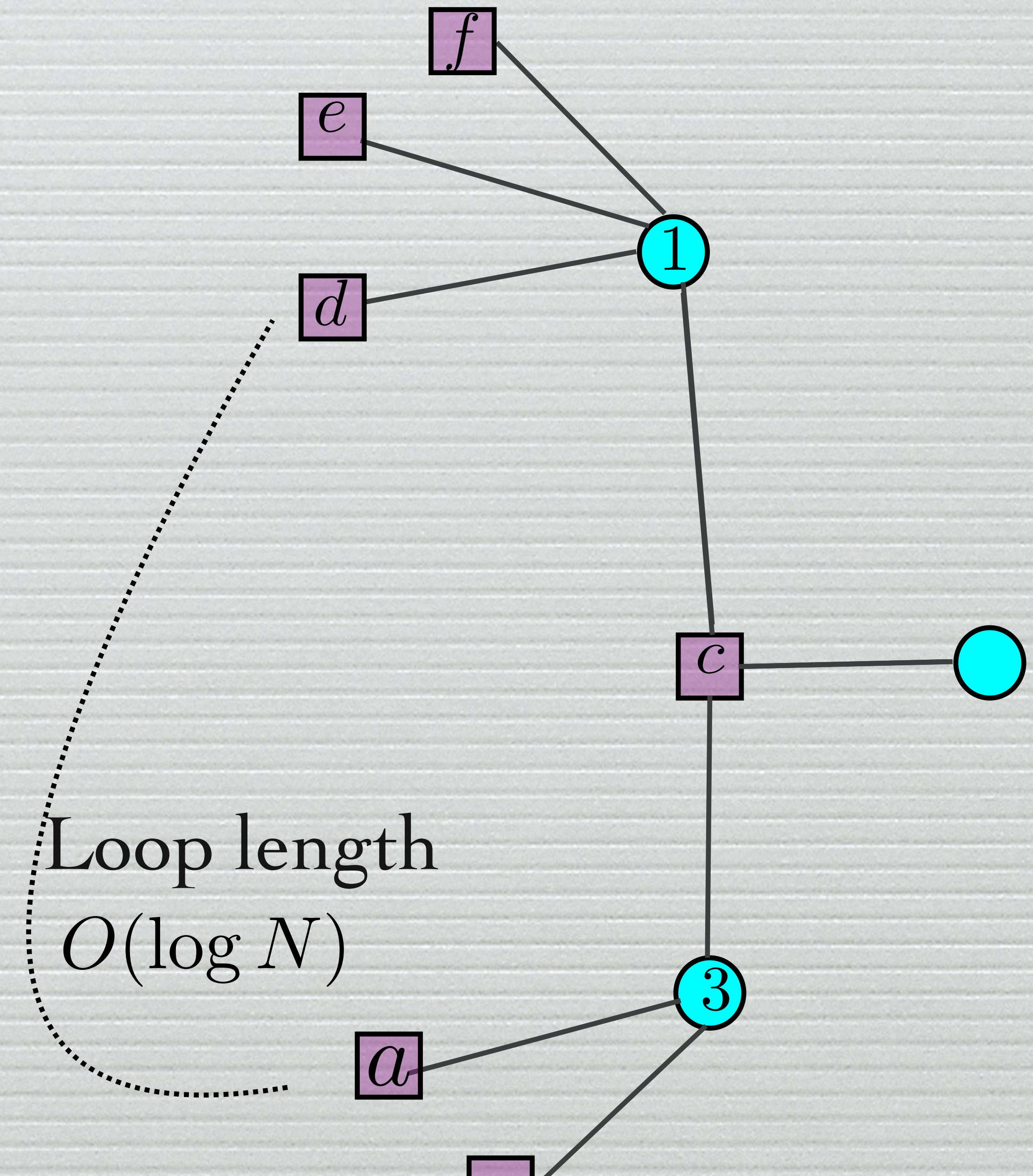
When is BP exact?

$$m_{1\to c}(x_1) = C m_{d\to 1}(x_1) m_{e\to 1}(x_1) m_{f\to 1}(x_1)$$

$$m_{c\to 2}(x_2) = \sum_{x_1, x_3} \psi_c(x_1, x_2, x_3) m_{1\to c}(x_1) m_{3\to c}(x_3)$$

Fluctuations are handled correctly, but beware of correlations

- Exact in one dimension (transfer matrix = dynamic programming)
- Exact on a tree (uncorrelated b.c)
- Exact on locally tree-like graphs (Erdös Renyi etc.) if correlations decay fast enough (single pure state) and uncorrelated disorder
- Exact in infinite range problems if correlations decay fast enough (single pure state) and uncorrelated disorder

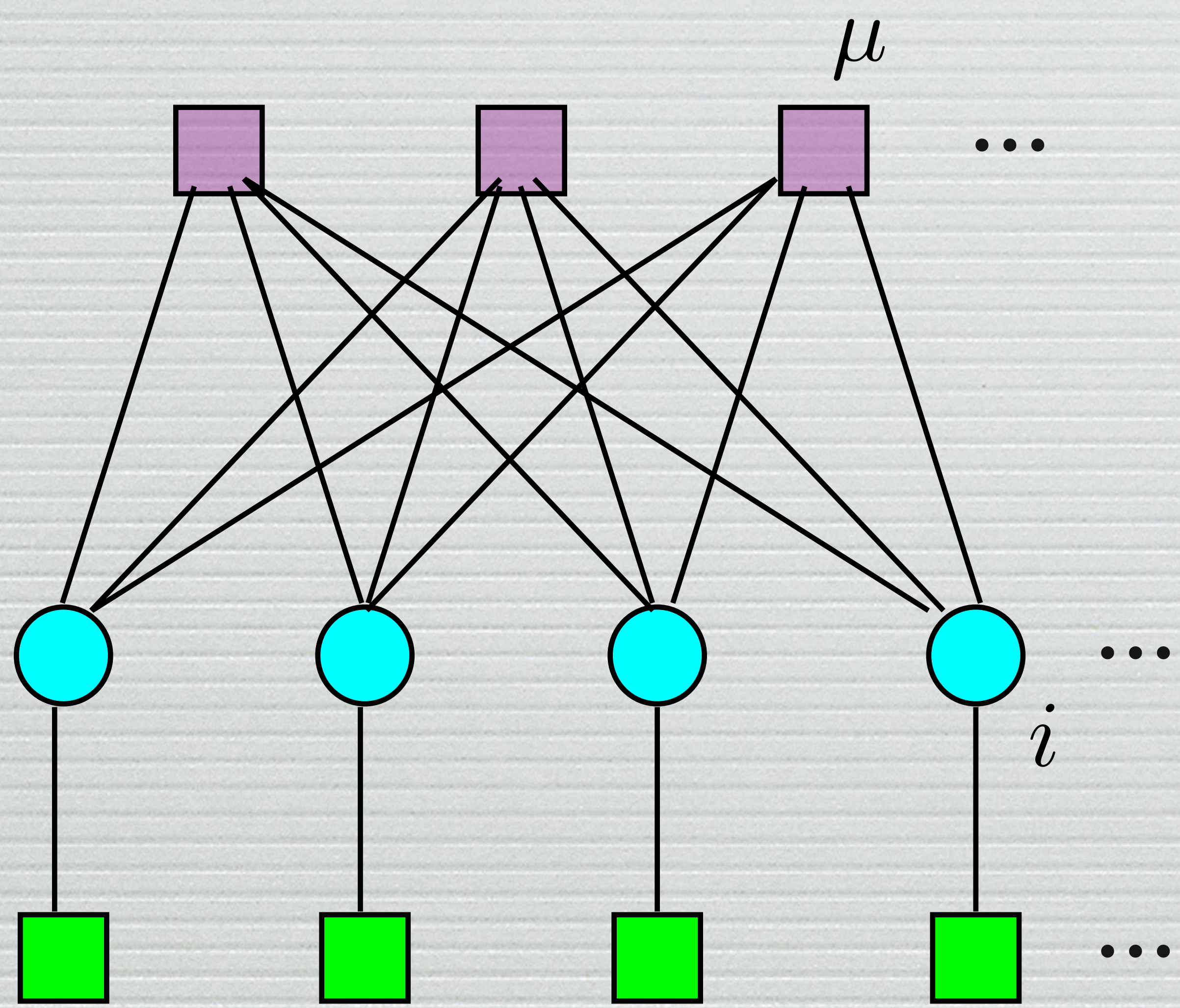


Two important developments

1) The special case of infinite-range models

2) What happens in a glass phase, when there are many pure states, and therefore many solutions ?

Infinite range models : from N^2 messages on the edges to N distributions on the vertices



$$m_{i\to\mu}(x_i) = \prod_{\nu(\neq\mu)} m_{\nu\to i}(x_i)$$

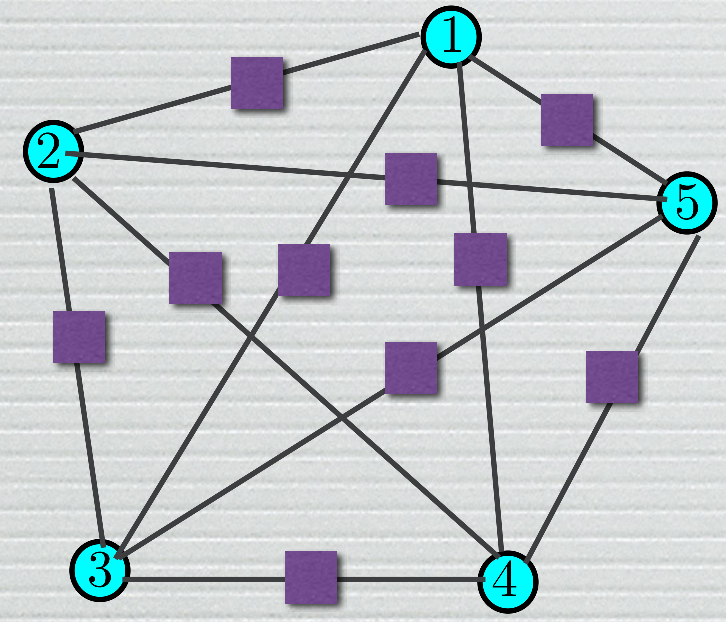
$$M_i(x_i) = \prod_{\nu} m_{\nu\to i}(x_i)$$

Small difference, treated perturbatively

Mean-field equations can be written only in terms of site pdfs: $M_i(x_i)$. TAP, AMP…

Example: SK model

SK model Pairwise interactions $J_{ij} = O\left(\frac{1}{\sqrt{N}}\right)$



$$m_{i\to(ij)}(s_i) \propto e^{h_{i\to j} s_i}$$

$h_{i\to j}$: local field on i in absence of j

BP equations: $$h_{i\backslash j} = \frac{1}{\beta} \sum_{k(\neq i)} \operatorname{atanh}[\tanh(\beta J_{ki}) \tanh(\beta h_{k\backslash i})] \simeq \sum_{k(\neq i)} J_{ki} \tanh(\beta h_{k\backslash i})$$

Full local field: $$H_i = \frac{1}{\beta} \sum_{k} \operatorname{atanh}[\tanh(\beta J_{ki}) \tanh(\beta h_{k\backslash i})] \simeq \sum_{k} J_{ki} \tanh(\beta h_{k\backslash i})$$

$h_{i\backslash j} \simeq H_i - O\left(\frac{1}{\sqrt{N}}\right)$: expand in the difference

SK model, TAP equations

SK model Pairwise interactions $J_{ij} = O\left(\frac{1}{\sqrt{N}}\right)$

Corrections can be handled to first order in perturbation theory, and all the equations close on the N variables H_i → **TAP equations**

$$H_i = \sum_k J_{ki} \tanh(\beta H_k) - \beta \tanh(\beta H_i) \sum_k J_{ki}^2 [1 - \tanh^2(\beta H_k)]$$

SK model, TAP equations

SK model Pairwise interactions $J_{ij} = O\left(\frac{1}{\sqrt{N}}\right)$

Corrections can be handled to first order in perturbation theory, and all the equations close on the N variables H_i → **TAP equations**

$$\overset{t+1}{H_i} = \sum_k J_{ki} \tanh(\beta \overset{t}{H_k}) - \beta \tanh(\beta \overset{t-1}{H_i}) \sum_k J_{ki}^2 [1 - \tanh^2(\beta \overset{t-1}{H_k})]$$

Time iteration (Bolthausen): AMP algorithm in information theory

Two important developments

1) The special case of infinite-range models

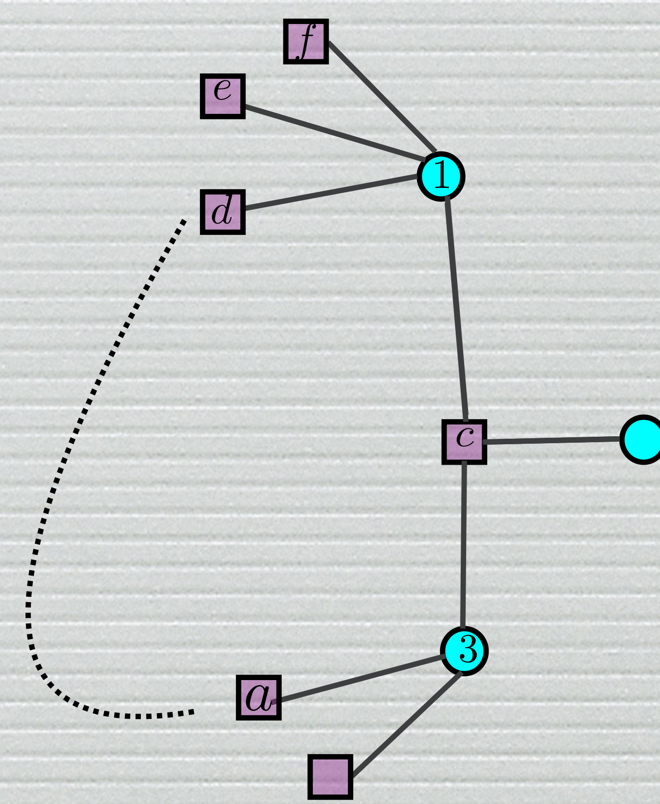
2) What happens in a glass phase, when there are many pure states, and therefore many solutions ?

2) What happens in a glass phase, when there are many pure states, and therefore many solutions ?

BP equations

$$m_{i\to\mu}(x_i) = \prod_{\nu(\neq\mu)} m_{\nu\to i}(x_i)$$

Correct if, in absence of the i-j interaction, the correlations between k and ℓ can be neglected.



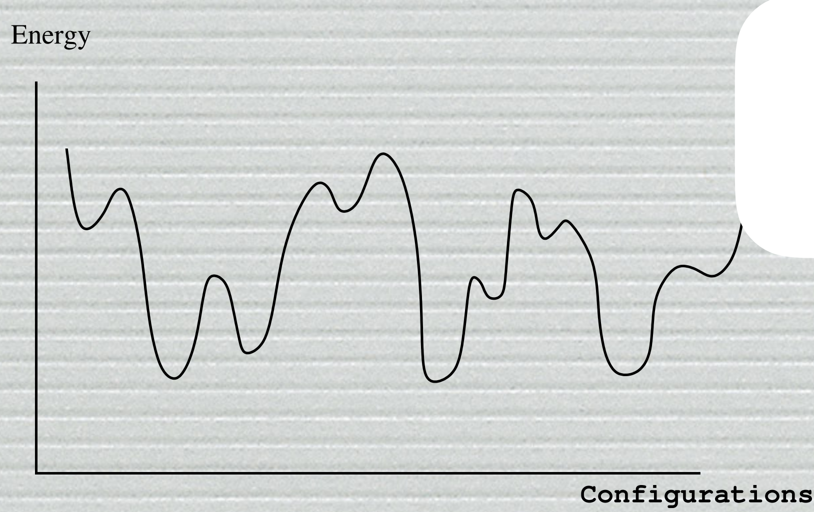
Loop length $O(\log N)$

2) What happens in a glass phase, when there are many pure states, and therefore many solutions ?

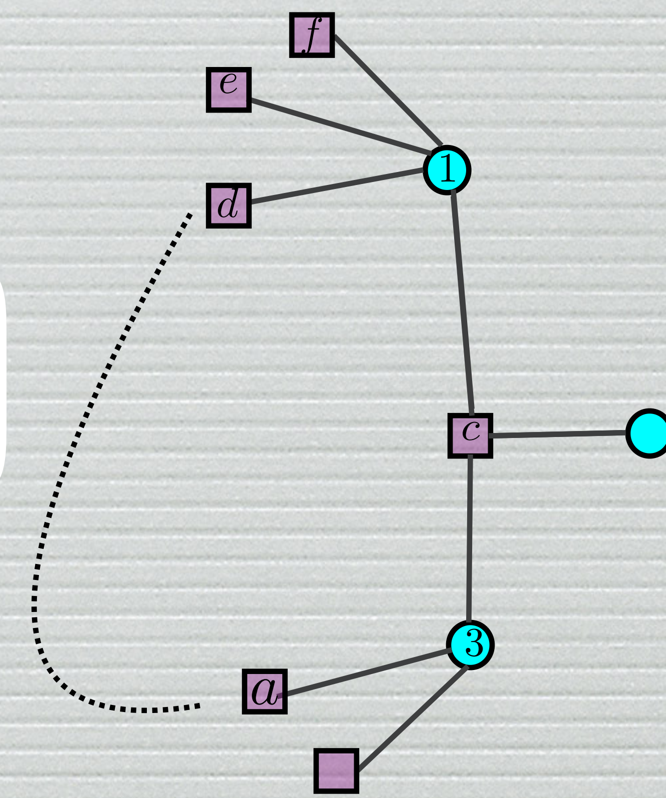
BP equations

$$m_{i\to\mu}(x_i) = \prod_{\nu(\neq\mu)} m_{\nu\to i}(x_i)$$

Correct if, in absence of the i-j interaction, the correlations between k and ℓ can be neglected.



$$m^{\alpha}_{i\to\mu}(x_i) = \prod_{\nu(\neq\mu)} m^{\alpha}_{\nu\to i}(x_i)$$



Loop length $O(\log N)$

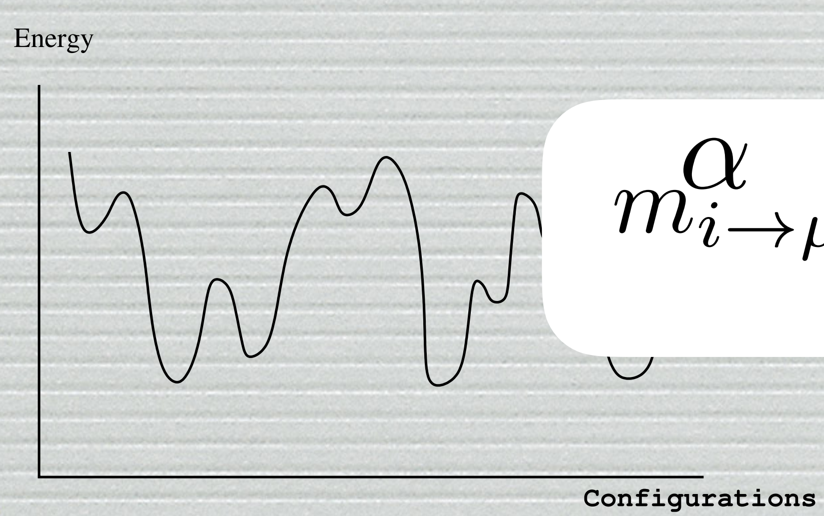
Glassy phase: many states, many solutions of BP

2) What happens in a glass phase, when there are many pure states, and therefore many solutions ?

BP equations

$$m_{i\to\mu}(x_i) = \prod_{\nu(\neq\mu)} m_{\nu\to i}(x_i)$$

Correct if, in absence of the i-j interaction, the correlations between k and ℓ can be neglected.



$$m^{\alpha}_{i\to\mu}(x_i) = \prod_{\nu(\neq\mu)} m^{\alpha}_{\nu\to i}(x_i)$$

Glassy phase: many states, many solutions of BP

Statistics of $m^{\alpha}_{i\to\mu}(x_i)$ over the many states α

$$P_{i\to\mu}(m)$$

related to

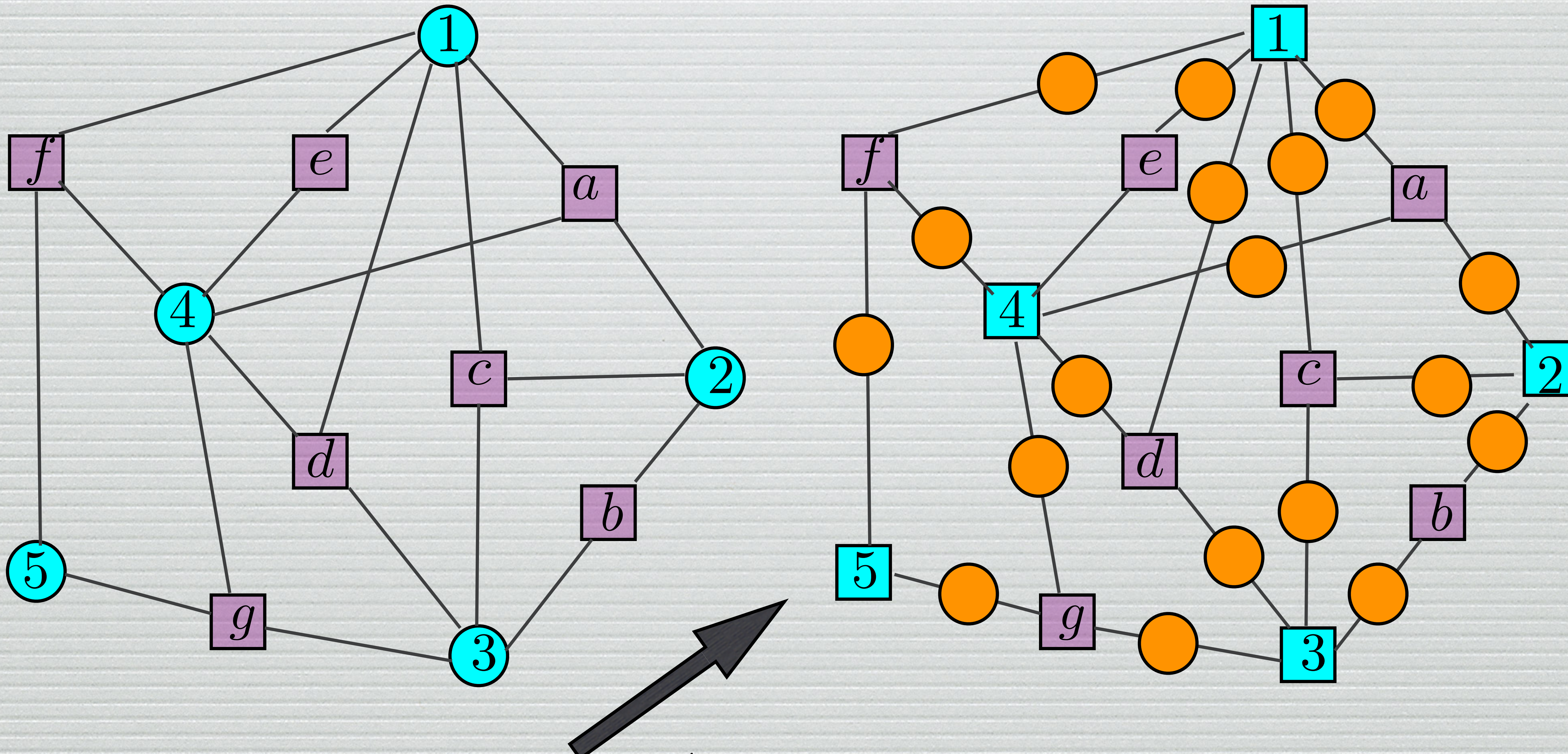
$$P_{\nu\to i}(m)$$

Survey propagation

M Parisi Zecchina 2002

SP=BP^2

Auxiliary problem: statistics over solutions:



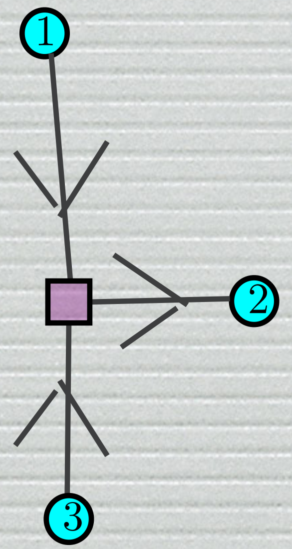
$$P(\{m_{a\to i}\}\{m_{j\to b}\}) = \frac{1}{Z}\prod_{ai}\mathbb{I}(m_{a\to i} = f(\{m_{j\to a}\})$$

$$\prod_{jb}\mathbb{I}(m_{j\to b} = f(\{m_{c\to j}\})$$

Power of message passing algorithms

Approximate solution of very hard, and very large constraint satisfaction problems, ...FAST! (typically linear time)

- BP: Best decoders for LDPC error correcting codes
- SP: Best solver of random satisfiability problems
- BP: Best algorithm for learning patterns in neural networks (e.g. binary perceptron)
- Data clustering, graph coloring, Steiner trees, etc…
- Fully connected networks : TAP (=AMP). Compressed sensing, linear estimation, etc.

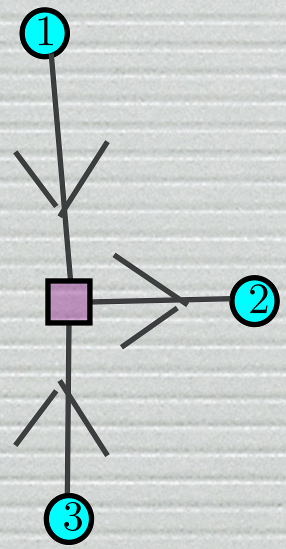


Local, simple update equations:
Each message is updated using information from incoming messages on the same node.
Distributed, solves hard global pb

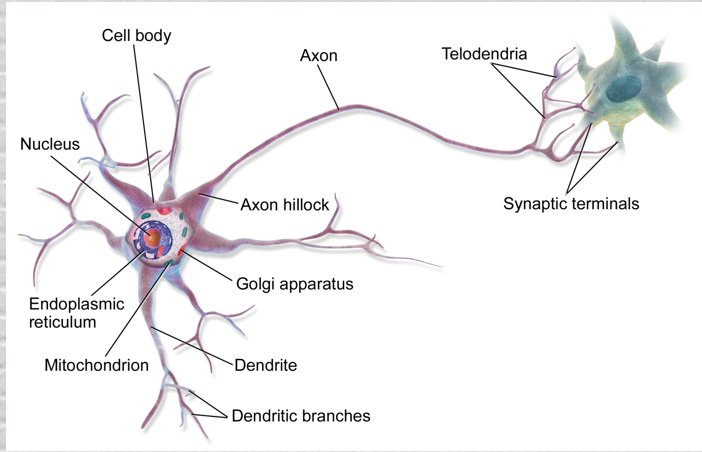
Power of message passing algorithms

Approximate solution of very hard, and very large constraint satisfaction problems, ...FAST! (typically linear time)

- BP: Best decoders for LDPC error correcting codes
- SP: Best solver of random satisfiability problems
- BP: Best algorithm for learning patterns in neural networks (e.g. binary perceptron)
- Data clustering, graph coloring, Steiner trees, etc…
- Fully connected networks : TAP (=AMP). Compressed sensing, linear estimation, etc.



Local, simple update equations:
Each message is updated using information from incoming messages on the same node.
Distributed, solves hard global pb



An example of fully connected model: Generalized Linear Regression

Unknowns: $x_i \qquad i = 1, \ldots, N$

Linear combinations: $z_\mu = \sum_i F_{\mu i} x_i \qquad \mu = 1, \ldots, M$

Outputs y_μ generated from $P_{\text{out}}(y_\mu | z_\mu)$

Prior factorized $\prod_i P(x_i)$

Bayes

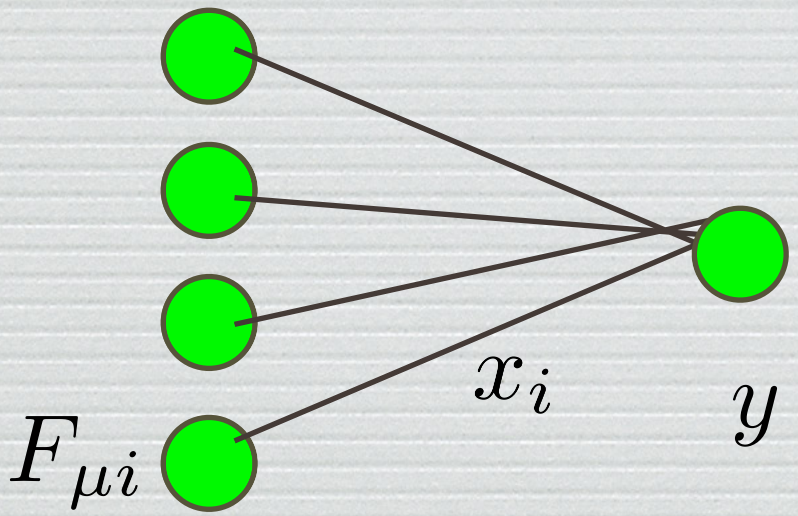
$$P(x|y) = \frac{1}{Z(y)} \prod_i P(x_i) \prod_\mu P_{\text{out}}(y_\mu | \sum_i F_{\mu i} x_i)$$

Examples: tomography, linear regression, perceptron learning, compressed sensing…

$$P(x|y) = \frac{1}{Z(y)} \prod_i P(x_i) \prod_\mu P_{\text{out}}(y_\mu | \sum_i F_{\mu i} x_i)$$

Perceptron learning

Pattern μ



$$P_{out} = \delta(y_\mu, Sign(z_\mu))$$

Linear regression:

Individual μ : expression of disease y_μ

Value of factor i for individual μ : $F_{\mu i}$

Find the best weights of factors x_i

Minimize mean square error with regularization

$$\frac{1}{2} \sum_\mu (y_\mu - \sum_i F_{\mu i} x_i)^2 + \sum_i ||x_i||$$

Lasso

$$P_{out} = e^{-(y_\mu - z_\mu)^2/(2\Delta)}$$

$$P(x_i) = e^{-||x_i||/\Delta}$$

$$P(x|y) = \frac{1}{Z(y)} \prod_i P(x_i) \prod_\mu P_{\rm out}(y_\mu | \sum_i F_{\mu i} x_i)$$

Compressed sensing

Unknown variables x_i

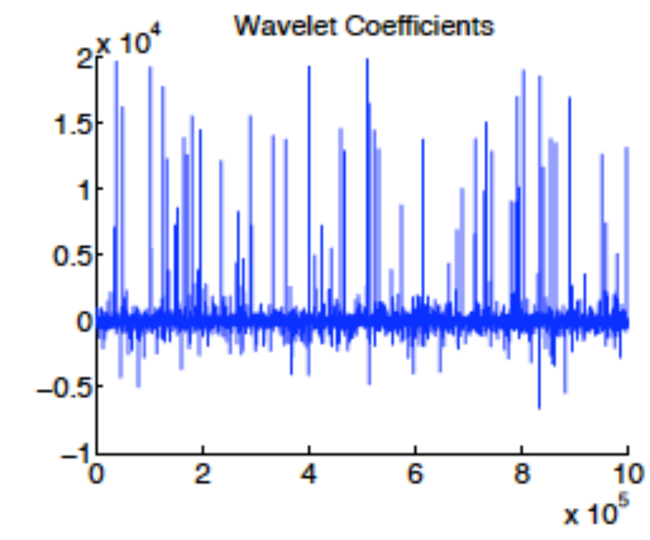
Linear measurements $y_\mu = \sum_i F_{\mu i} x_i + \eta_\mu$

$$P_{out} = e^{-(y_\mu - z_\mu)^2/(2\Delta)}$$

Compressed sensing regime : $M < N$

sparse prior (in appropriate basis)

$$P(x_i) = (1-\rho)\delta(x_i) + \rho\phi(x_i)$$

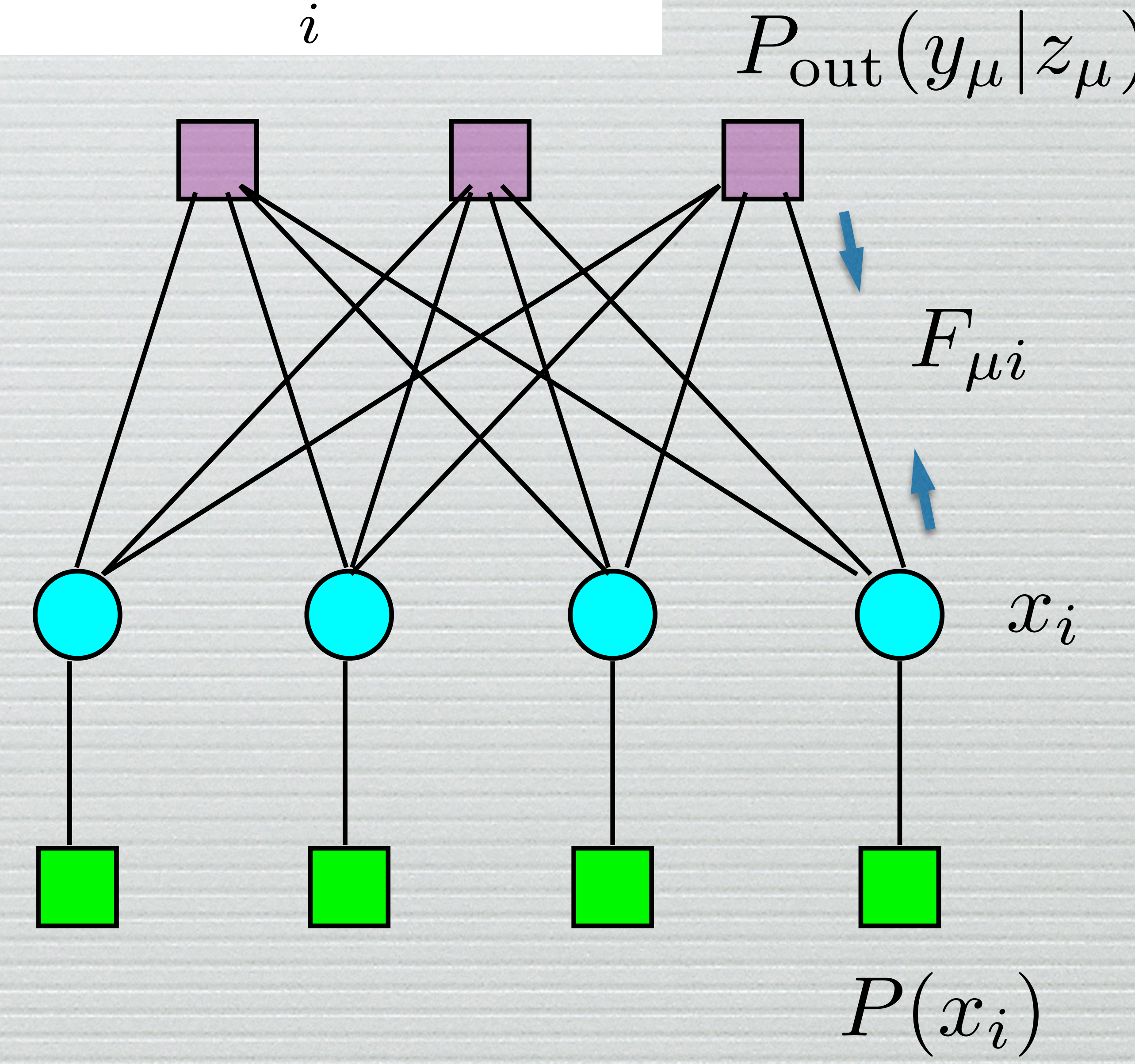


$$P(x|y) = \frac{1}{Z(y)} \prod_i P(x_i) \prod_\mu P_{\text{out}}(y_\mu | \sum_i F_{\mu i} x_i)$$

$F_{\mu i}$: iid, known

Spin glass with multispin interactions, infinite range: write mean field equations.

Messages: $m_{i \to \mu}(x_i)$
$m_{\mu \to i}(x_i)$



Becomes Gaussian in the thermodynamic limit

Mézard 1989, Opper Winther 96, Kabashima 2003, 2008 , Donoho Maleki Montanari 2009, Rangan+ 2011, Krzakala+ 2012, …

BP equations

$$a_{i\to\mu} = \int \mathrm{d}x_i \, x_i \, m_{i\to\mu}(x_i)$$

$$v_{i\to\mu} = \int \mathrm{d}x_i \, x_i^2 \, m_{i\to\mu}(x_i) - a_{i\to\mu}^2$$

$$m_{\mu\to i}(x_i) = \frac{1}{\tilde{Z}^{\mu\to i}} e^{-\frac{x_i^2}{2} A_{\mu\to i} + B_{\mu\to i} x_i}$$

$$m_{i\to\mu}(x_i) = \frac{1}{\tilde{Z}^{i\to\mu}} \left[(1-\rho)\delta(x_i) + \rho\phi(x_i)\right] e^{-\frac{x_i^2}{2} \sum_{\gamma\neq\mu} A_{\gamma\to i} + x_i \sum_{\gamma\neq\mu} B_{\gamma\to i}}$$

BP equations

$$a_{i\to\mu} = \int \mathrm{d}x_i \, x_i \, m_{i\to\mu}(x_i)$$

$$v_{i\to\mu} = \int \mathrm{d}x_i \, x_i^2 \, m_{i\to\mu}(x_i) - a_{i\to\mu}^2$$

$$m_{\mu\to i}(x_i) = \frac{1}{\tilde{Z}^{\mu\to i}} e^{-\frac{x_i^2}{2} A_{\mu\to i} + B_{\mu\to i} x_i}$$

$$m_{i\to\mu}(x_i) = \frac{1}{\tilde{Z}^{i\to\mu}} \left[(1-\rho)\delta(x_i) + \rho\phi(x_i)\right] e^{-\frac{x_i^2}{2} \sum_{\gamma\neq\mu} A_{\gamma\to i} + x_i \sum_{\gamma\neq\mu} B_{\gamma\to i}}$$

Four «messages» sent along each edge $i-\mu$

($4NM$ numbers) can be simplified to $O(N)$ parameters

From « cavity messages »

TAP equations

$$a_{i\to\mu} = \int \mathrm{d}x_i \, x_i \, m_{i\to\mu}(x_i)$$

$$v_{i\to\mu} = \int \mathrm{d}x_i \, x_i^2 \, m_{i\to\mu}(x_i) - a_{i\to\mu}^2$$

To full local distribution

$$a_i = \int dx_i \; m_i(x_i) x_i = \langle x_i \rangle$$

$$v_i = \int dx_i \; m_i(x_i) x_i^2 - a_i^2 = \langle x_i^2 \rangle - a_i^2$$

TAP = coupled equations between the 2N variables $a_i, \; v_i$

Iteration ⟶ algorithm : GAMP

Statistical study ⟶ phase diagram

Benchmark: noiseless limit of compressed sensing with iid measurements

System of linear measurements

$$y = Fx$$

Measurements

$$y = \begin{pmatrix} y^1 \\ \cdot \\ \cdot \\ y^M \end{pmatrix}$$

Signal $x = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{pmatrix}$

$F = M \times N$ matrix

Random F : «random projections» (incoherent with signal)

Pb: Find x when $M < N$ and x is sparse

Phase diagram

«Thermodynamic limit»

$N \gg 1$	variables
$R = \rho N$	non-zero variables
$M = \alpha N$	equations

- Solvable by enumeration when $\alpha > \rho$ but $O(e^N)$
- ℓ_1 norm approach

  Find a N - component vector x such that the M equations $y = Fx$ are satisfied and $||x||_1$ is minimal
- AMP = Bayesian approach

Planted: $\phi_T(x)$

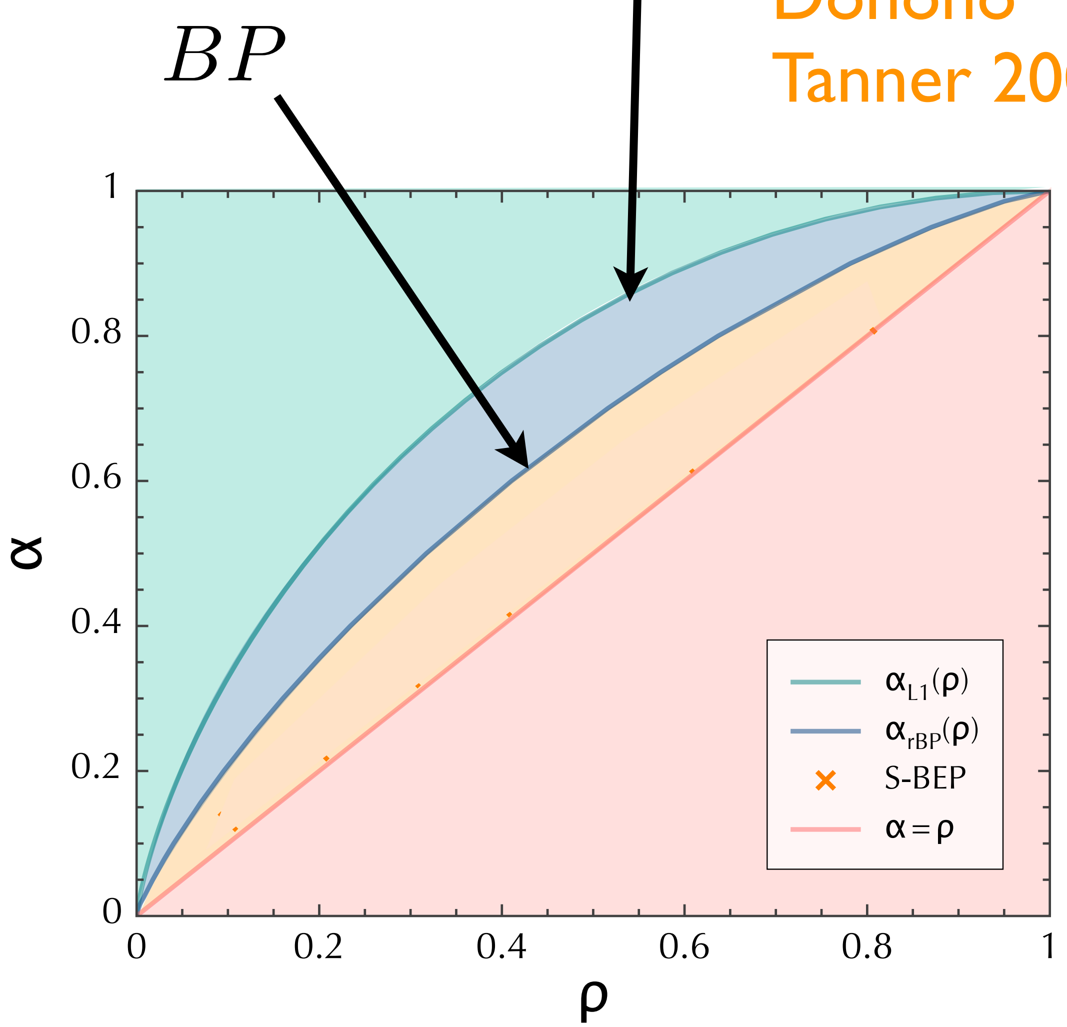
$$P(\mathbf{x}) = \prod_{i=1}^{N} [(1-\rho)\delta(x_i) + \rho\phi(x_i)] \quad \prod_{\mu=1}^{P} \delta\left(y_\mu - \sum_i F_{\mu i} x_i\right)$$

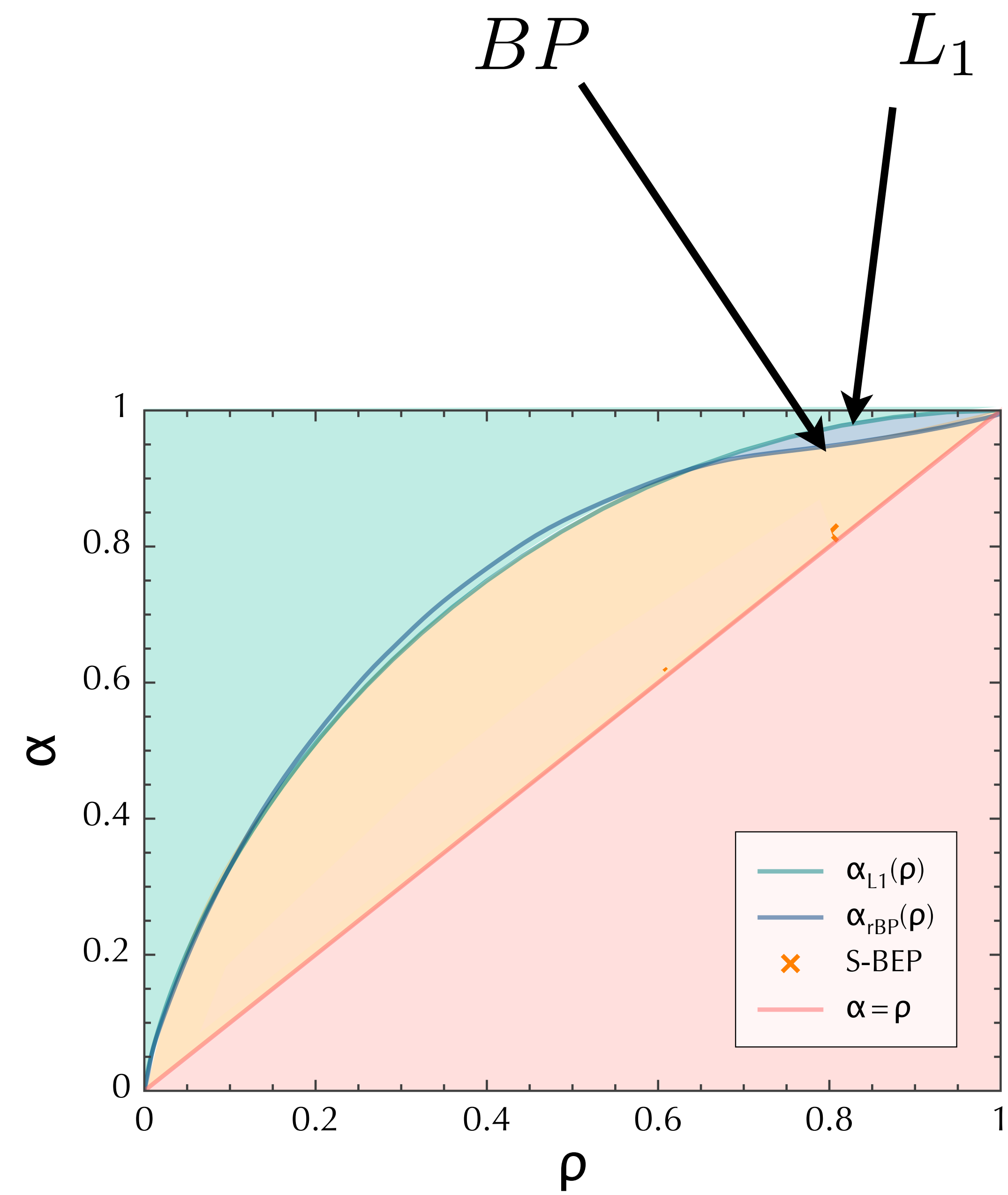
↑

Performance of AMP with Gauss-Bernoulli prior: phase diagram

Krzakala Sausset Mézard Sun Zdeborova 2011

Donoho 2006, Donoho Tanner 2005





Gaussian signal

$$\phi_T(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Binary signal

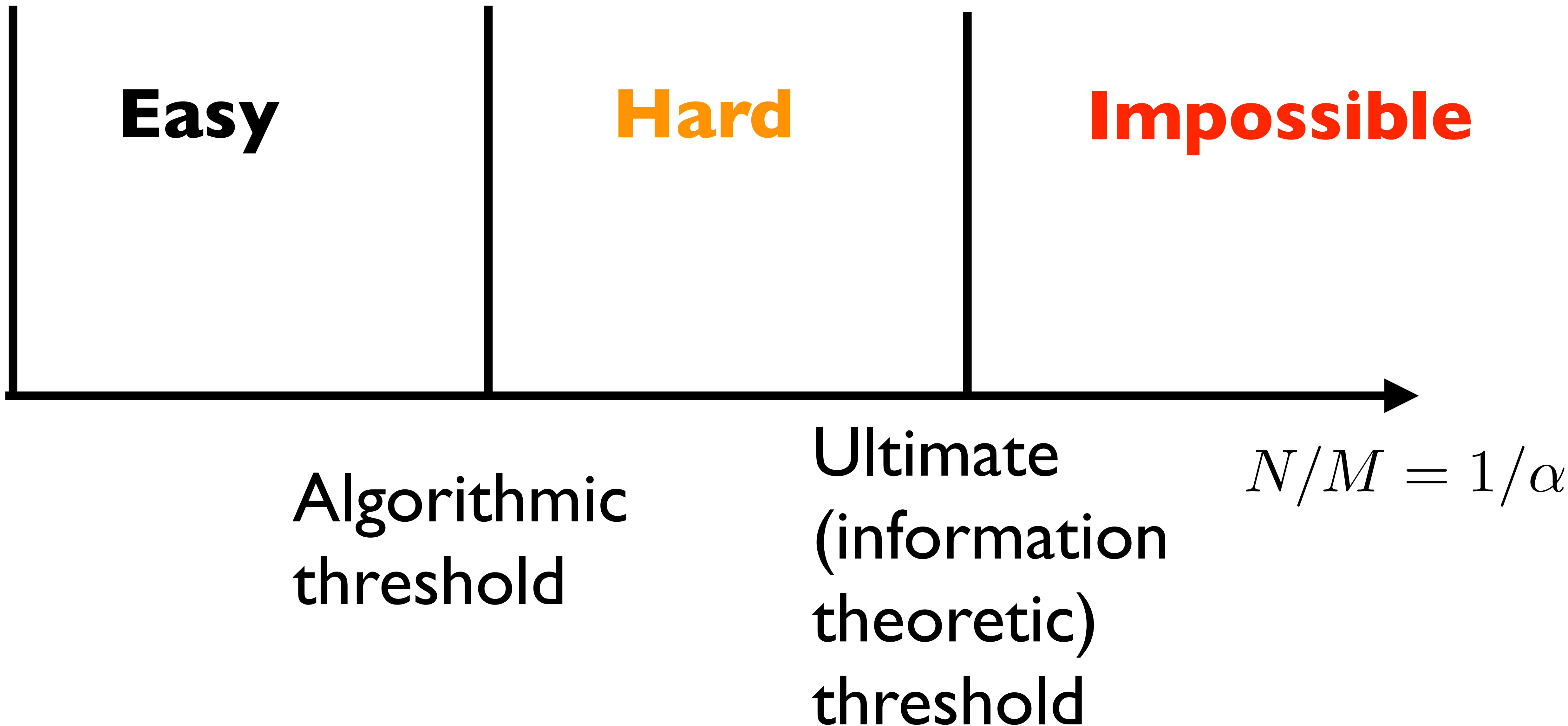
$$\phi_T(x) = \frac{1}{2}\left(\delta_{x,1} + \delta_{x,-1}\right)$$

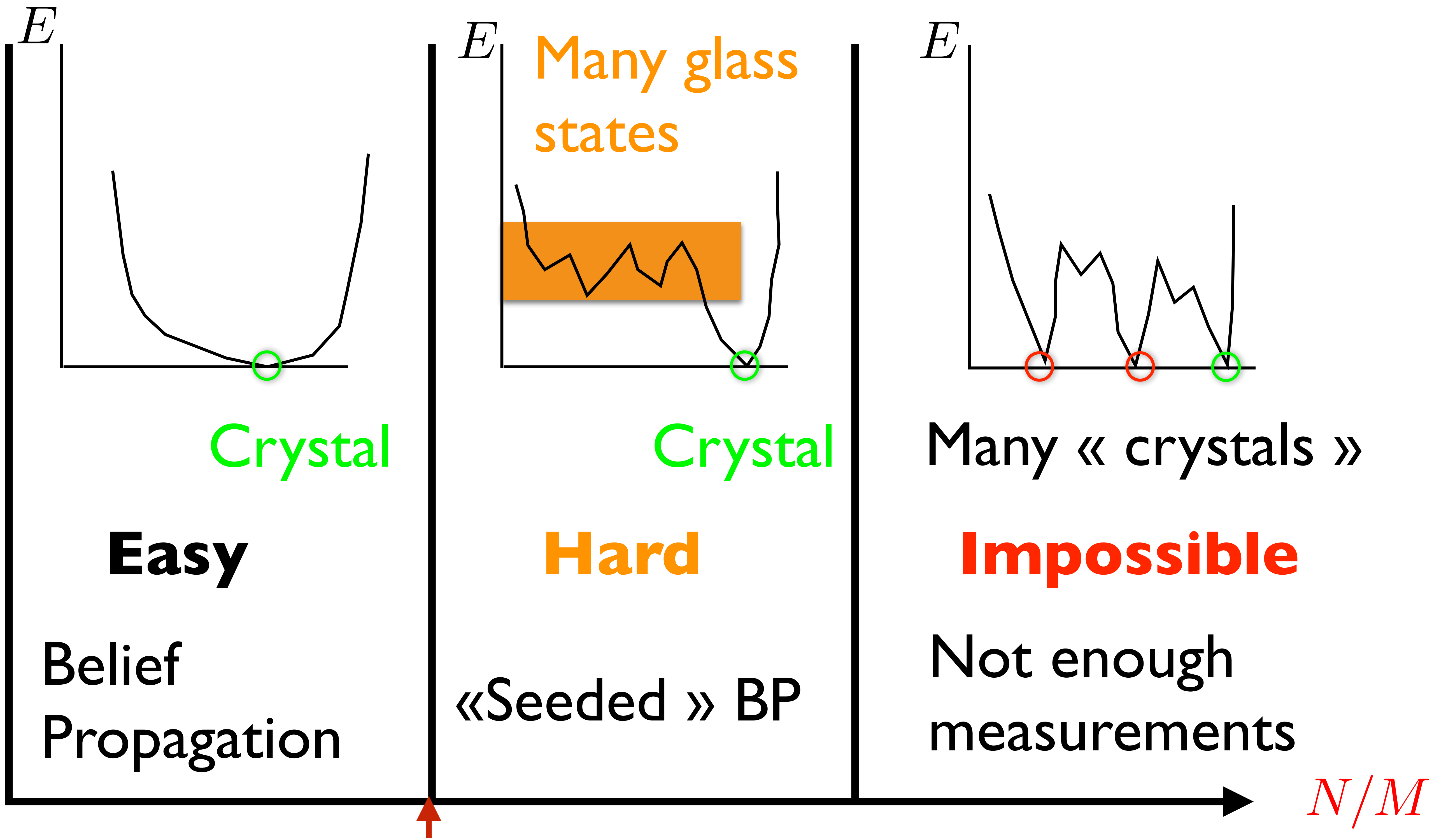
Analysis of random instances : phase transitions

N (real) variables, M measurements (linear functions)

Analysis of random instances : phase transitions

Reconstruction of signal using BP. Fixed ρ , decrease α





Dynamical phase transition. Ubiquitous in statistical inference. Conjecture « All local algorithms freeze »… How universal?

Getting around the glass trap

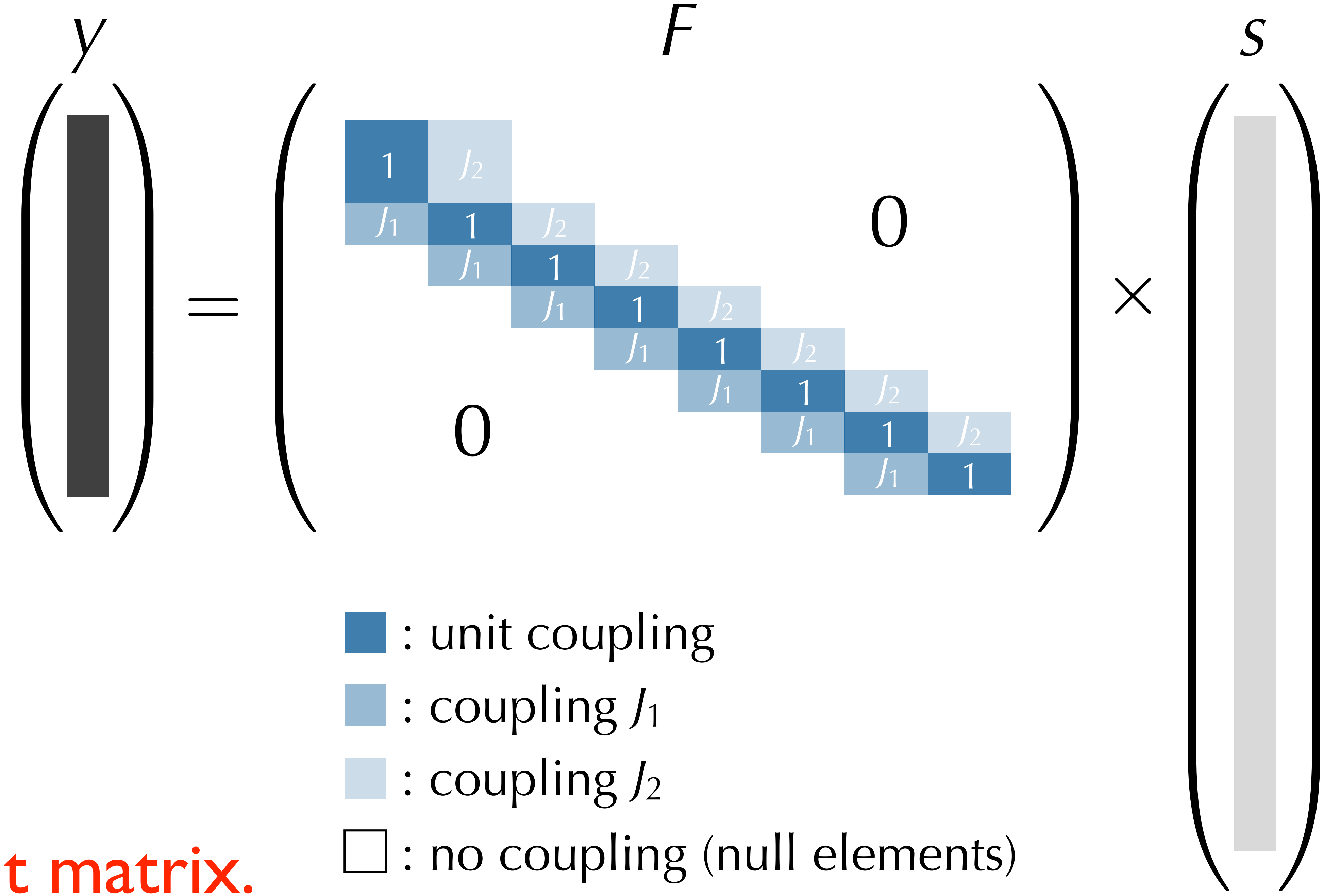
Design the matrix F so that one nucleates the naive state (crystal nucleation idea,
...borrowed from error correcting codes : « spatial coupling »)

Felström-Zigangirov,
Kudekar Richardson Urbanke,
Hassani Macris Urbanke,
...

«Seeded BP»

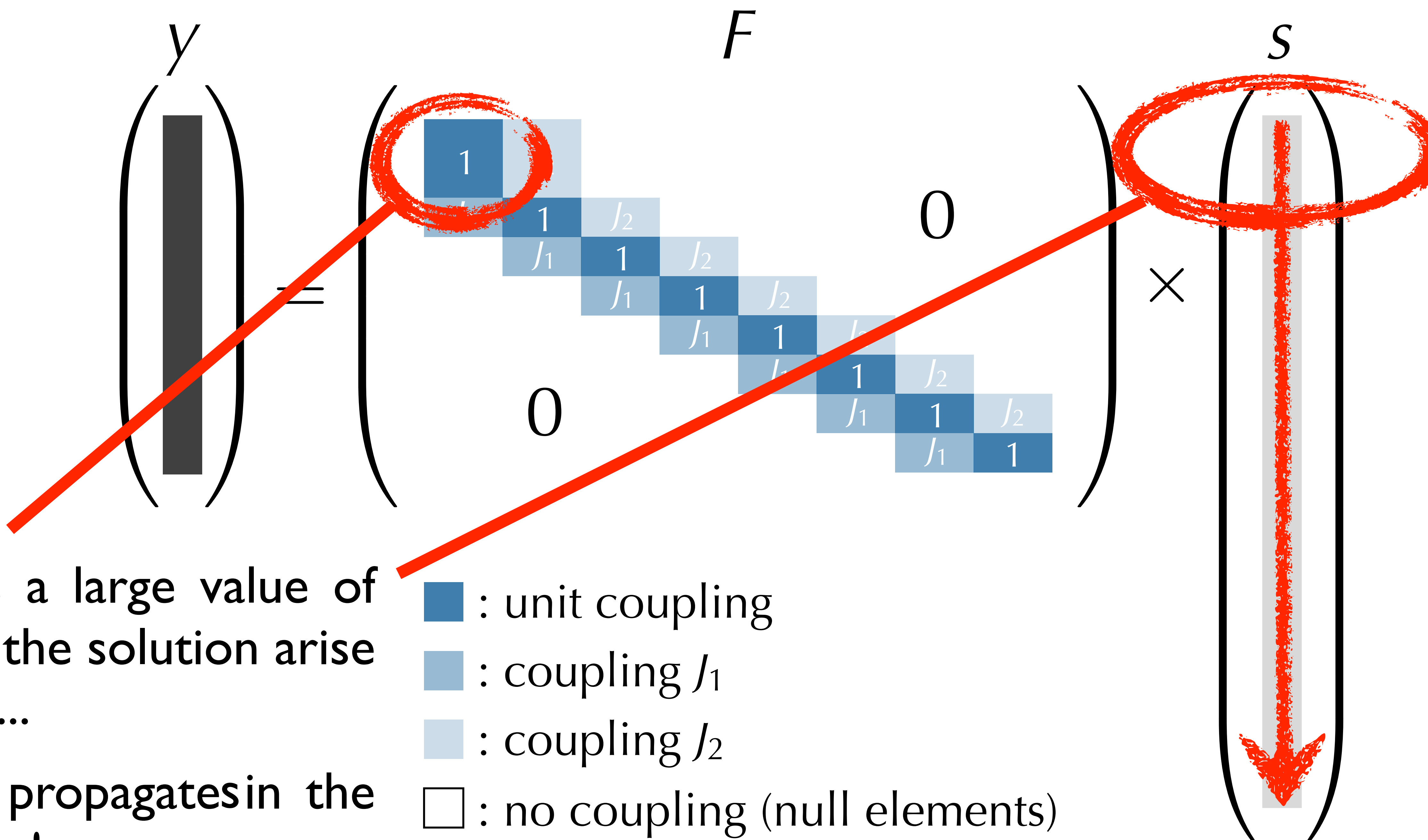
Nucleation and seeding

Nucleation and seeding



Structured measurement matrix. Variances of the matrix elements

$F_{\mu i} =$ independent random Gaussian variables, zero mean and variance $J_{b(\mu)b(i)}/N$



Block 1 has a large value of M such that the solution arise in this block...

... and then propagatesin the whole system!

$$L = 8$$

$$N_i = N/L$$

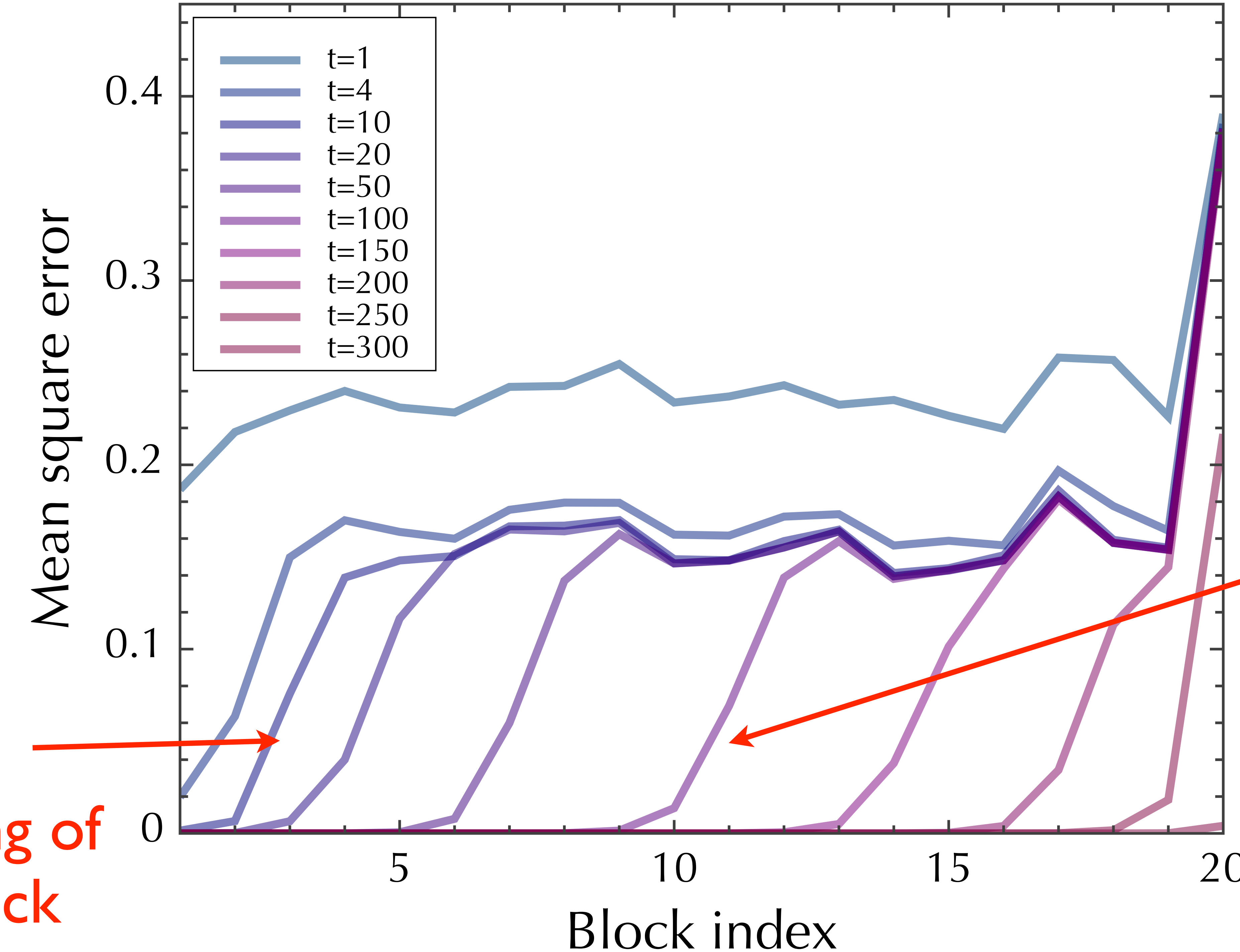
$$M_i = \alpha_i N/L$$

$$\alpha_1 > \alpha_{BP}$$

$$\alpha_j = \alpha' < \alpha_{BP} \qquad j \geq 2$$

$$\alpha = \frac{1}{L}\left(\alpha_1 + (L-1)\alpha'\right)$$

Numerical study



$t = 100$

decoding of blocks 1 to 9

$t = 10$

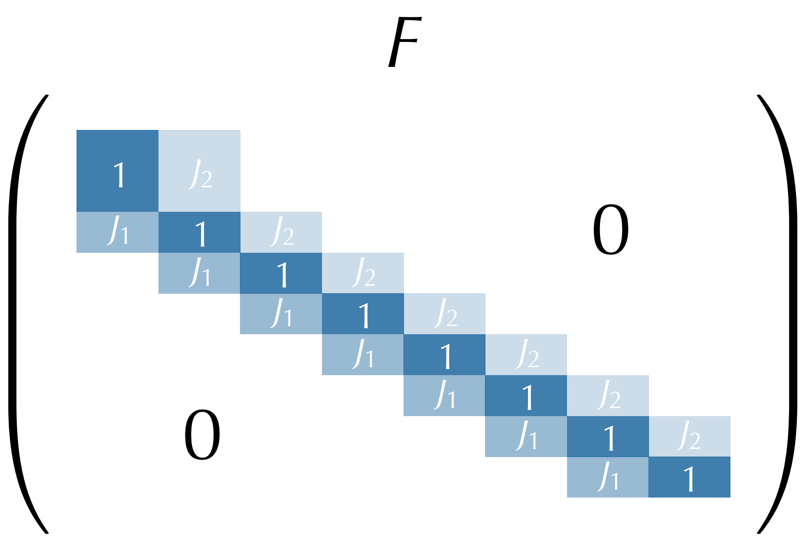
decoding of first block

$L = 20$ $N = 50000$ $\rho = .4$ $J_1 = 20$ $\alpha_1 = 1$

$J_2 = .2$ $\alpha = .5$

Performance of the probabilistic approach + message passing + parameter learning+ seeding matrix

$$Z = \int \prod_{j=1}^{N} \mathrm{d}x_j \prod_{i=1}^{N} [(1-\rho)\delta(x_i) + \rho\phi(x_i)] \prod_{\mu=1}^{M} \delta\left(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i\right)$$



- Simulations
- Analytic approaches (replicas and cavity)

$$\rightarrow \alpha_c = \rho_0$$

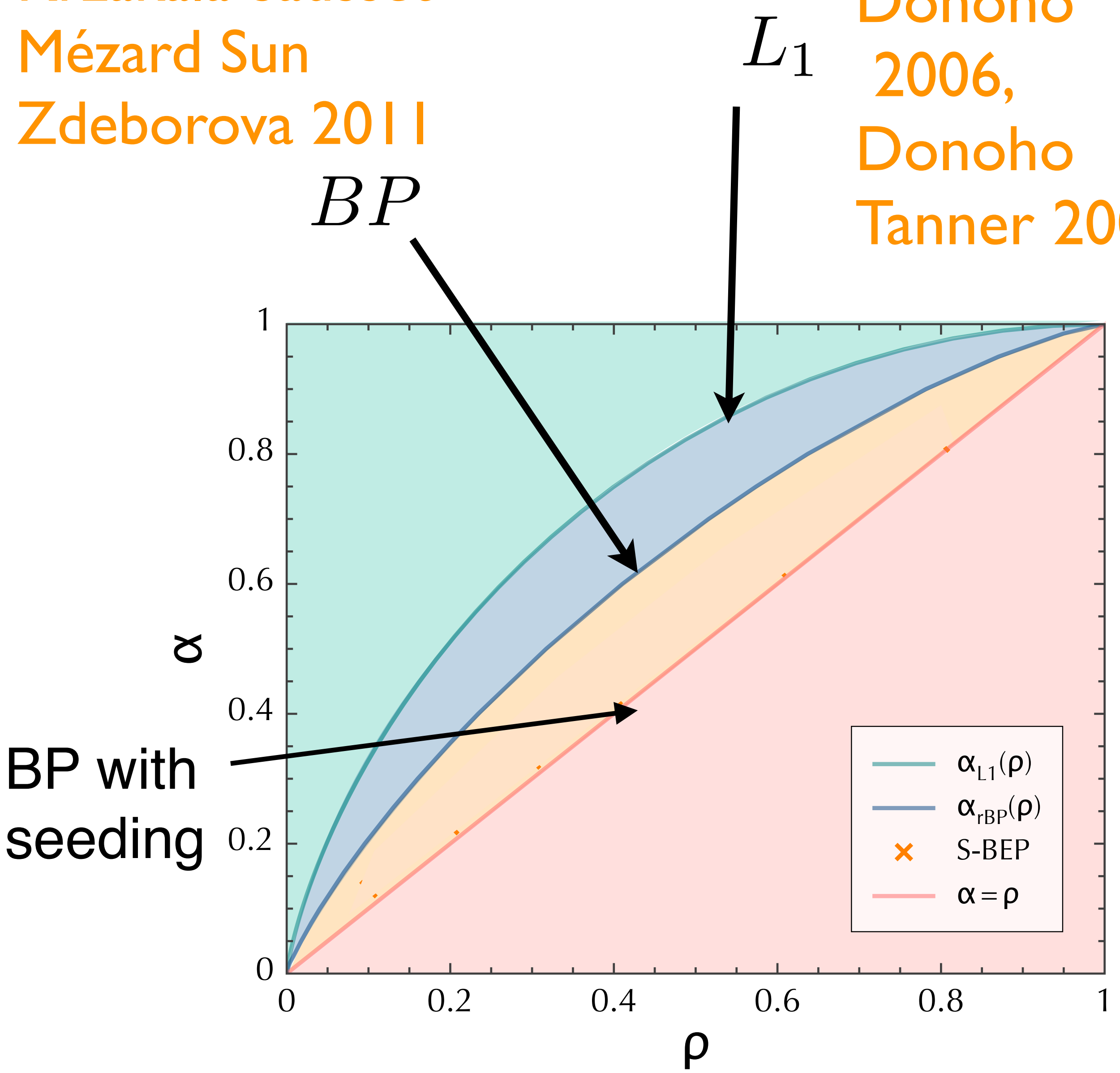
Reaches the ultimate information-theoretic threshold

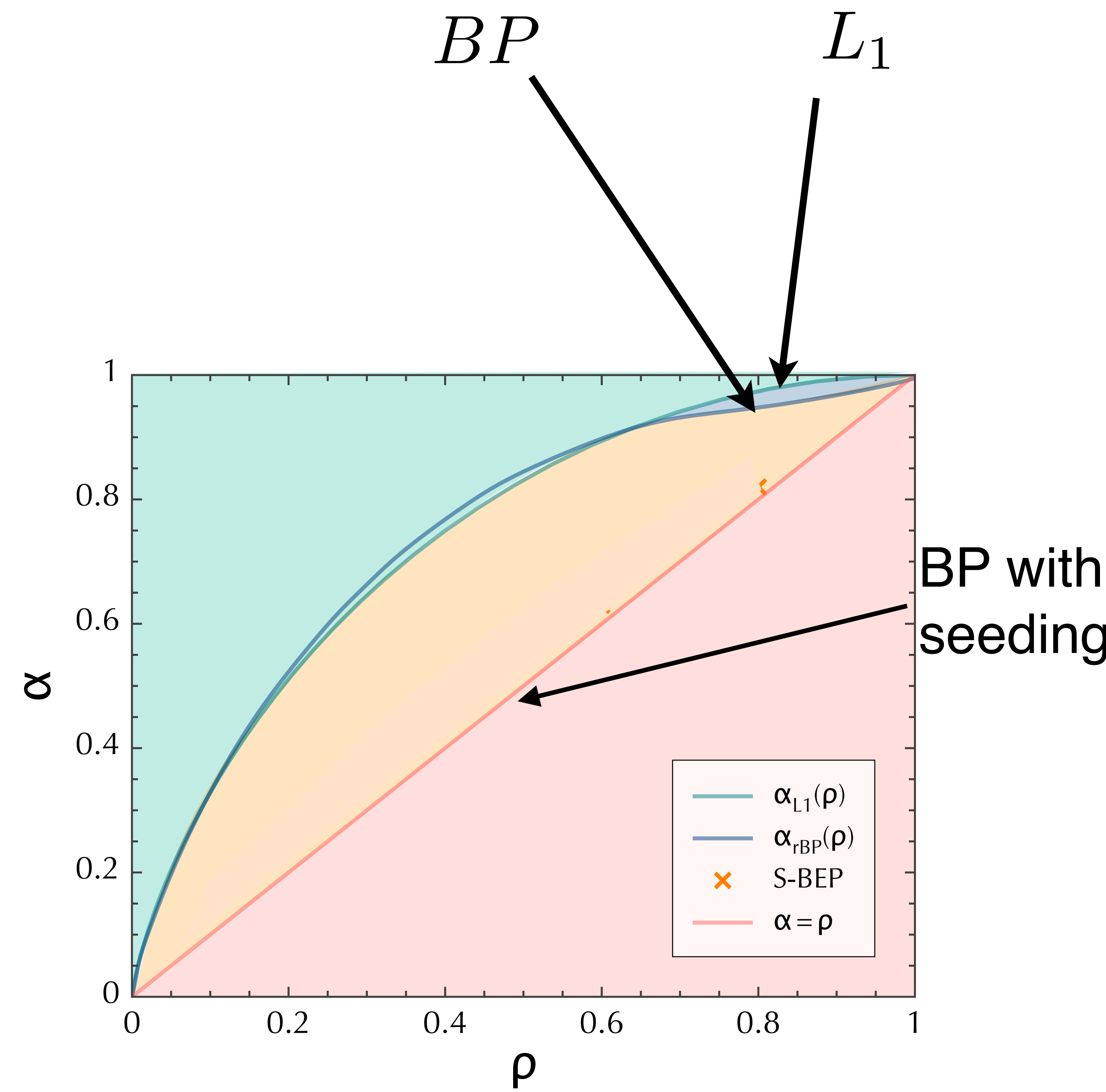
Proof: Donoho Javanmard Montanari

Performance of AMP with Gauss-Bernoulli prior: phase diagram

Krzakala Sausset Mézard Sun Zdeborova 2011

Donoho 2006, Donoho Tanner 2005



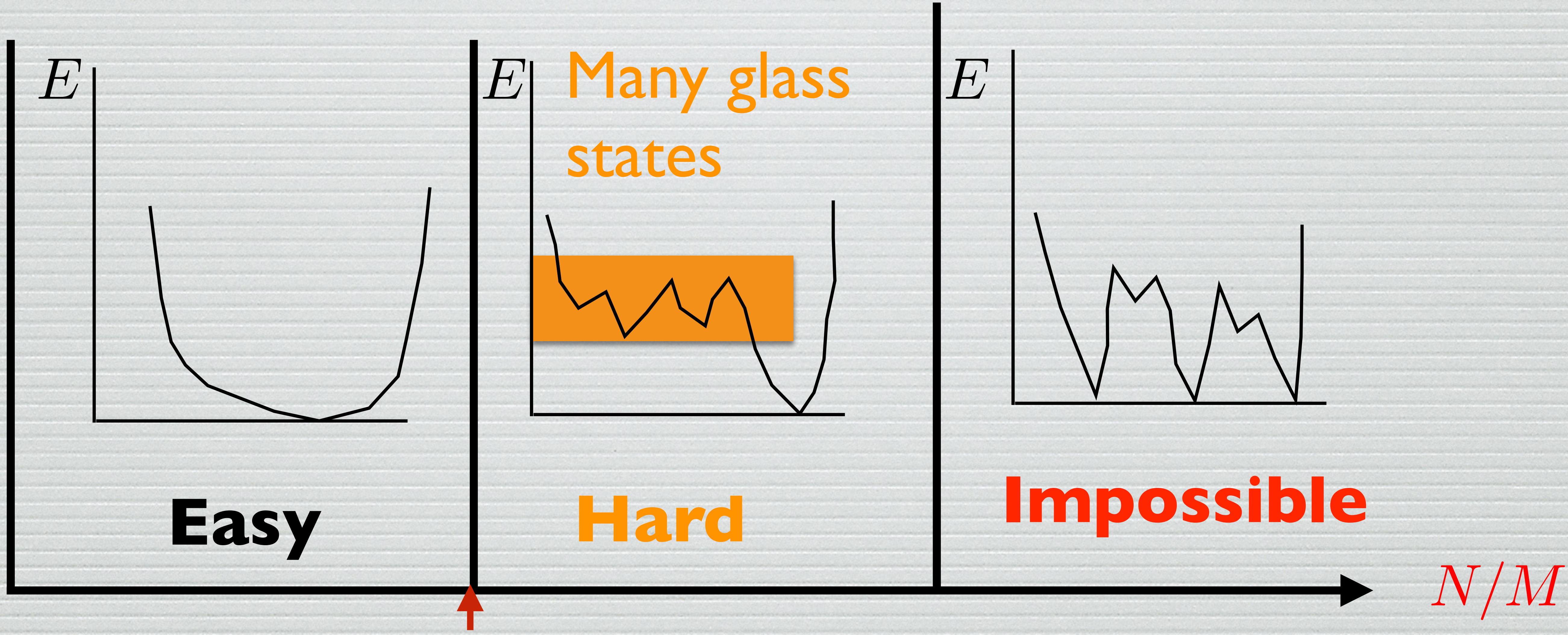


Gaussian signal

$$\phi_T(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Binary signal

$$\phi_T(x) = \frac{1}{2} \left(\delta_{x,1} + \delta_{x,-1}\right)$$



Phase transitions are crucial in large inference problems

Hard-Impossible = absolute limit (Shannon-like)

Easy- Hard = limit for large class of algorithms (local)

The spin glass cornucopia

A very sophisticated and powerful corpus of conceptual and methodological approaches has been developed (replicas, cavity, TAP,....) mostly in the years 1975-2000, and has found applications in many different fields of information theory and computer science

Portrait of Ottavio Strada,

Tintoretto, Venice 1567

Rijk's Museum Amsterdam