Compressed sensing using spin-glass concepts

Marc Mézard

Ecole normale supérieure - PSL Research University, Paris
and
CNRS, Université Paris Sud

Cargèse, August 2014

General perspective

Major trend of 21st century scientific research, in all fields (including social sciences and humanities): massive data acquisition, often noisy.

How to search efficiently in a database?
How to make sense of it? What to search?

Models vs statistical analysis
Extract relevant information from data

A kind of introduction: linear regression

Linear regression: Output variable y

N Input variables F_i $i = 1, \ldots, N$

A kind of introduction: linear regression

Linear regression: Output variable y

N Input variables F_i $\quad i = 1, \ldots, N$

Seek a linear correlation between y and the inputs F_i

$$y \simeq \sum_{i=1}^{N} x_i F_i + \eta$$

$x =$ parameter vector

A kind of introduction: linear regression

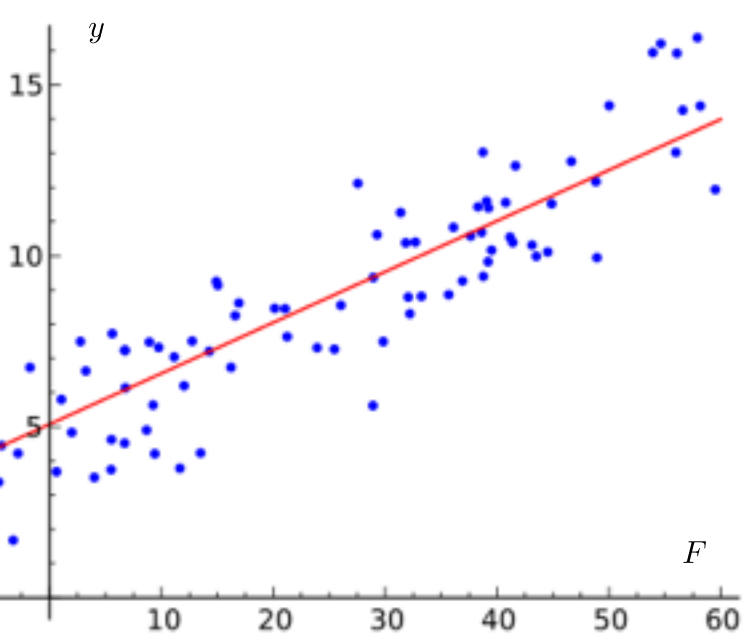
Linear regression: Output variable y

N Input variables F_i $i = 1, \ldots, N$

Seek a linear correlation between y and the inputs F_i

$$y \simeq \sum_{i=1}^{N} x_i F_i + \eta$$

$x =$ parameter vector



A kind of introduction: linear regression

Linear regression: Output variable y

N Input variables F_i $i = 1, \ldots, N$

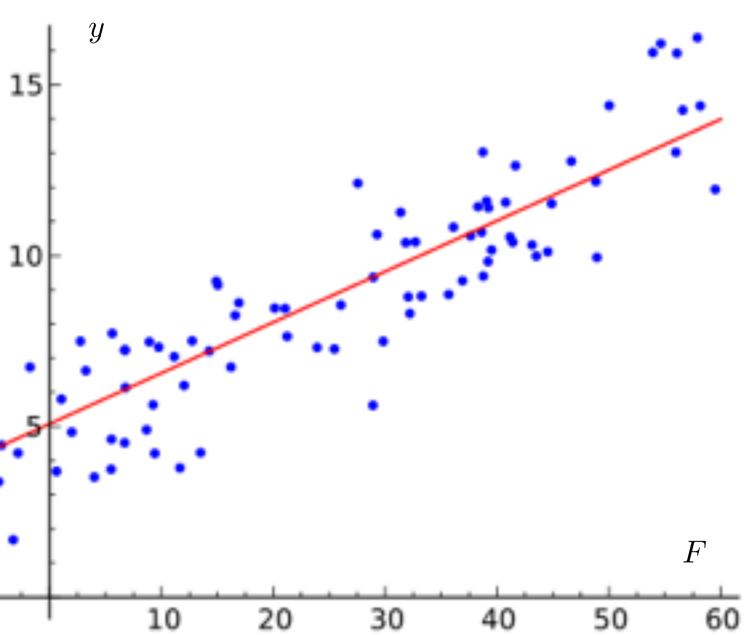
Seek a linear correlation between y and the inputs F_i

$$y \simeq \sum_{i=1}^{N} x_i F_i + \eta$$

$x =$ parameter vector

M measurements = (input,output pairs)

$$(y^\mu, F_1^\mu, \ldots, F_N^\mu) \qquad \mu = 1, \ldots, M$$



Find $\{x_i\}$ such that $\sum_\mu \left(y^\mu - \sum_i F_i^\mu x_i \right)^2$

is minimal

Usual method

M measurements = (input,output pairs)

Find $\{x_i\}$ such that $\sum_\mu \left(y^\mu - \sum_i F_i^\mu x_i \right)^2$ is minimal

Least squares method (Legendre 1805, Gauss1795-1809)

Usual method

M measurements = (input,output pairs)

Find $\{x_i\}$ such that $\sum_\mu \left(y^\mu - \sum_i F_i^\mu x_i \right)^2$ is minimal

Least squares method (Legendre 1805, Gauss1795-1809)

New strategy : include many input variables (large N)

but ask that many of the regression coefficients x_i be zero

Usual method

M measurements = (input,output pairs)

Find $\{x_i\}$ such that $\sum_\mu \left(y^\mu - \sum_i F_i^\mu x_i \right)^2$ is minimal

Least squares method (Legendre 1805, Gauss1795-1809)

New strategy : include many input variables (large N)

but ask that many of the regression coefficients x_i be zero

-LASSO= solve linear regression with L_1 constraint

Tibshirani 1996

$$\sum_i |x_i| < s$$

Usual method

M measurements = (input,output pairs)

Find $\{x_i\}$ such that $\sum_\mu \left(y^\mu - \sum_i F_i^\mu x_i \right)^2$ is minimal

Least squares method (Legendre 1805, Gauss1795-1809)

New strategy : include many input variables (large N)

but ask that many of the regression coefficients x_i be zero

-LASSO= solve linear regression with L_1 constraint

Tibshirani 1996

$$\sum_i |x_i| < s$$

-Compressed sensing: ask that only R regression coefficients be non-zero, …

See YK's last slide

Compressed sensing and beyond

- Explain data by decomposing it into unknown factors
- Include many factors in the analysis
- Ask that most factors do not contribute
- Useful if data has some structure (Science!)

Compressed sensing and beyond

- Explain data by decomposing it into unknown factors
- Include many factors in the analysis
- Ask that most factors do not contribute
- Useful if data has some structure (Science!)

A new way of doing science? Discover subtle hidden « factors », then look for an explanation (or not…) e.g. personalized medicine.

Compressed sensing and beyond

- Explain data by decomposing it into unknown factors
- Include many factors in the analysis
- Ask that most factors do not contribute
- Useful if data has some structure (Science!)

A new way of doing science? Discover subtle hidden « factors », then look for an explanation (or not…) e.g. personalized medicine.

Not limited to linear problems. e.g.:

- Discrete tomography
- Group testing
- Infer regulatory interactions in gene expression network

Compressed sensing and beyond

- Explain data by decomposing it into unknown factors
- Include many factors in the analysis
- Ask that most factors do not contribute
- Useful if data has some structure (Science!)

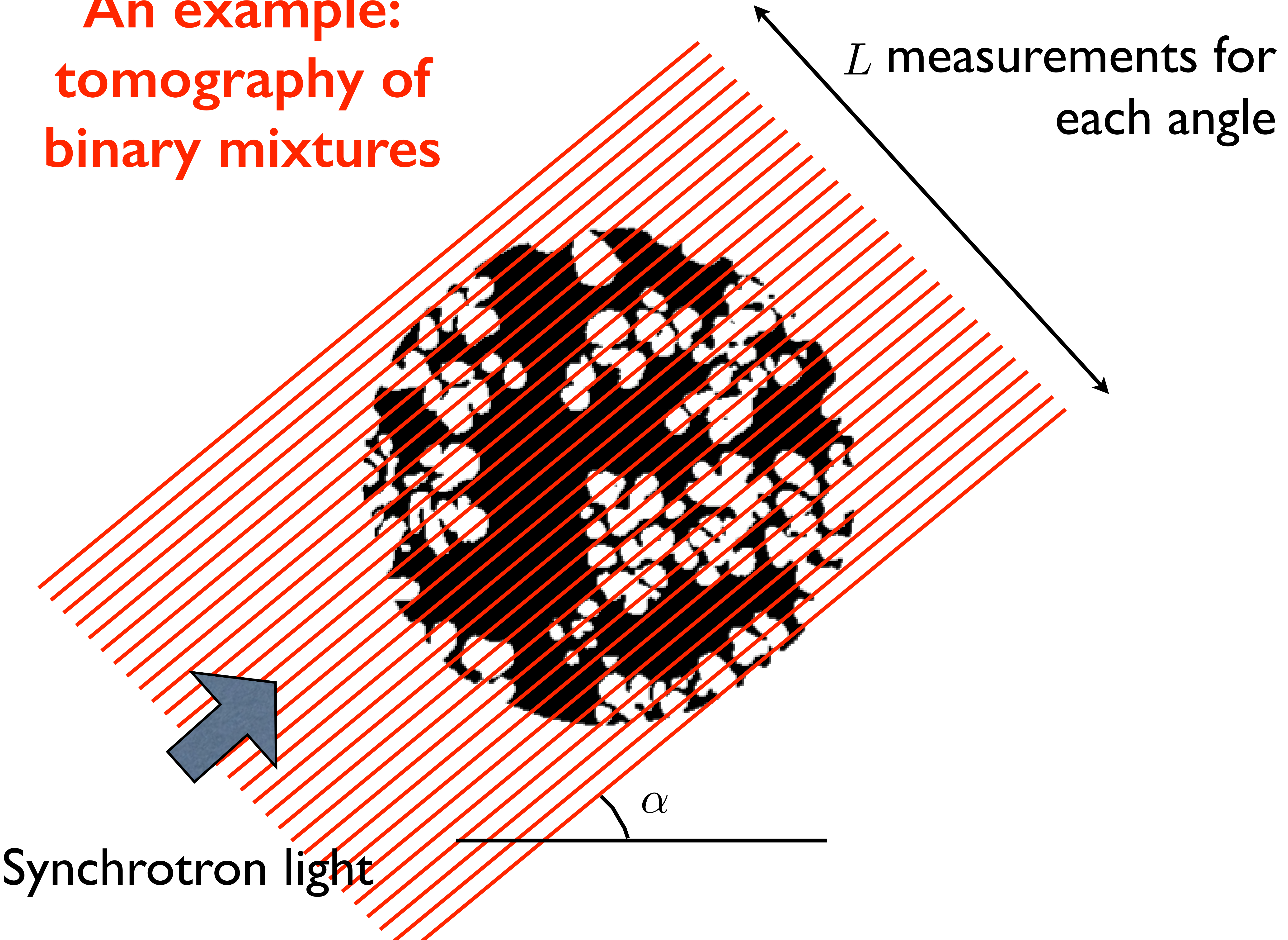
A new way of doing science? Discover subtle hidden « factors », then look for an explanation (or not…) e.g. personalized medicine.

Not limited to linear problems. e.g.:

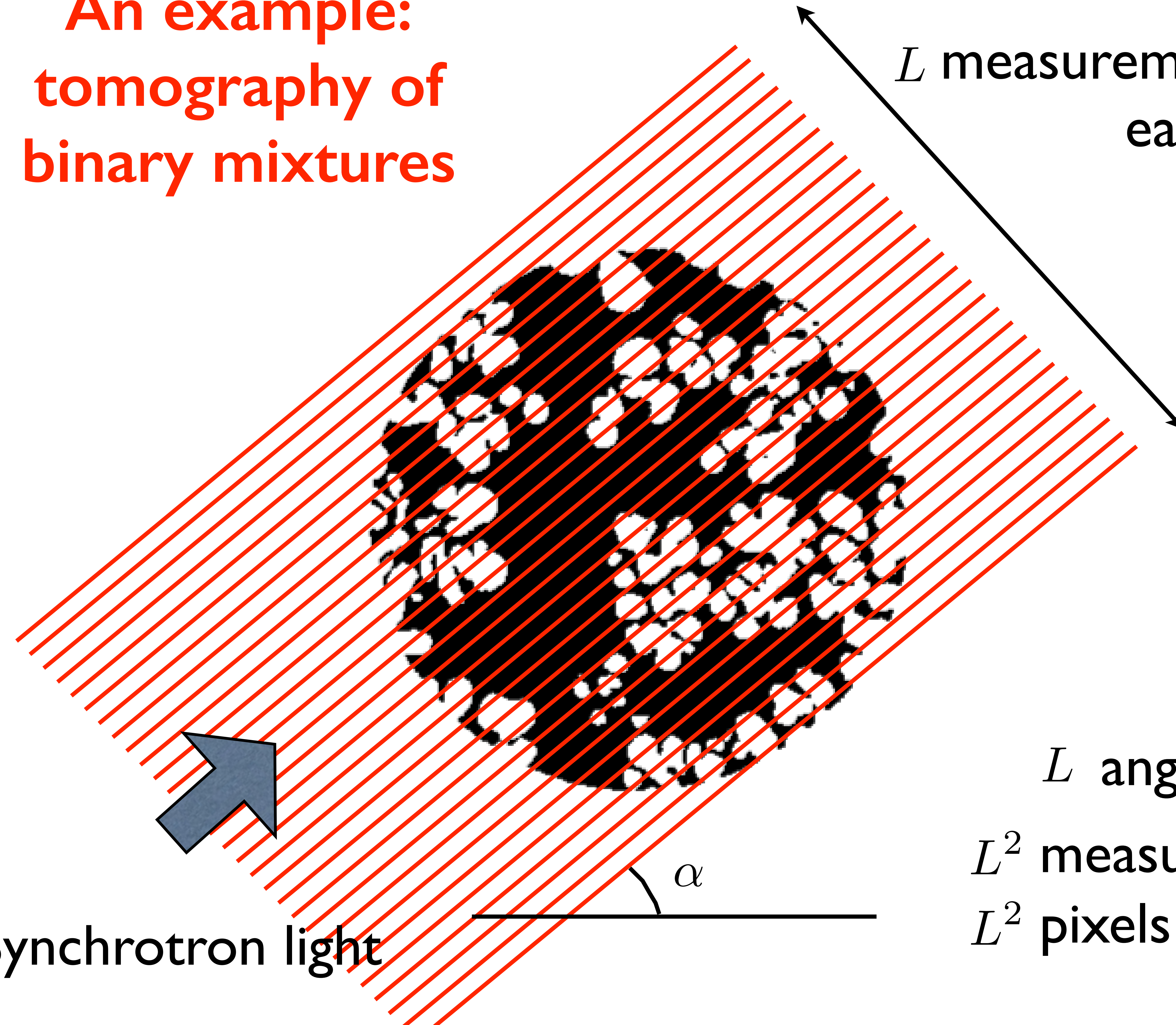
- Discrete tomography
- Group testing
- Infer regulatory interactions in gene expression network
- How does the brain work?

An example: tomography of binary mixtures

An example: tomography of binary mixtures



An example: tomography of binary mixtures



L angles:

L^2 measurements

L^2 pixels

An example: 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

$\xi \gg 1$

An example: tomography of binary mixtures

sample size L

pixel size 1

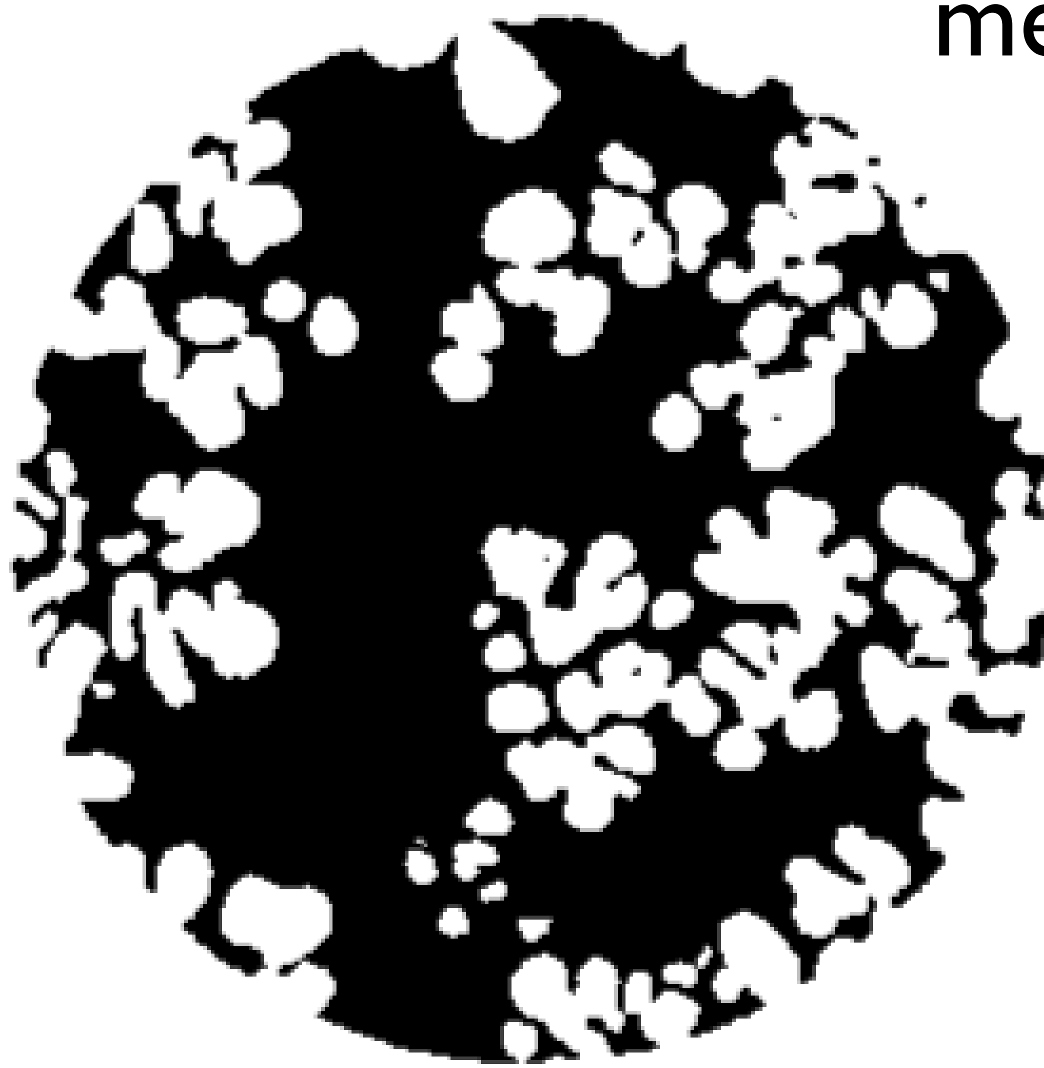
domain size ξ

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

$\xi \gg 1$

An example: tomography of binary mixtures

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



Gouillart et al., Inverse problems 2013

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

Back to the simplest problem: getting a signal from some measurement= **linear transforms**

Consider a system of linear measurements

Measurements → $$y = F\, s$$ ← Unknown signal

Known measurement matrix

$$y = \begin{pmatrix} y^1 \\ . \\ . \\ y^M \end{pmatrix} \qquad s = \begin{pmatrix} s_1 \\ . \\ . \\ . \\ s_N \end{pmatrix}$$

$F = M \times N$ matrix

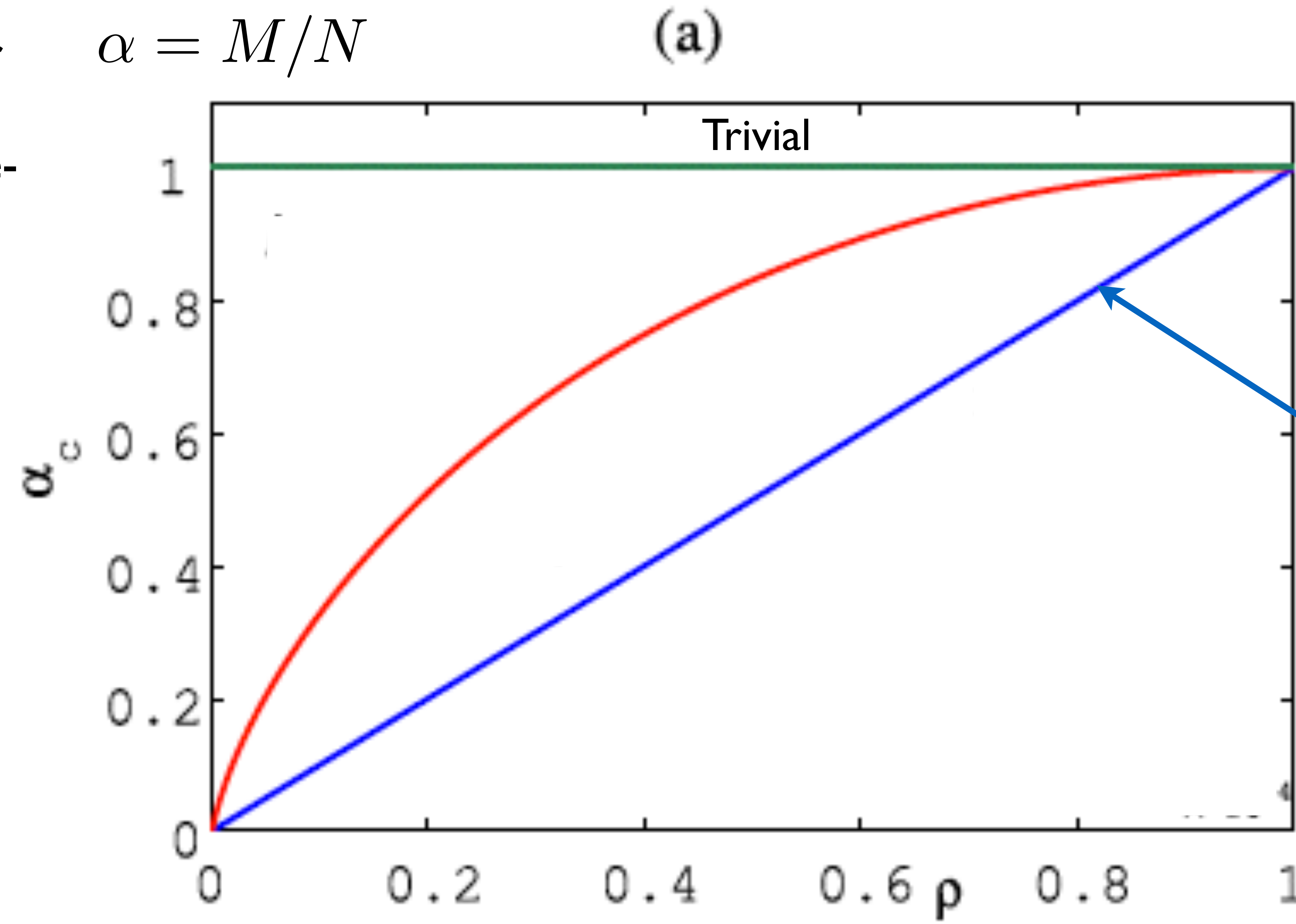
Random F : «random projections» (incoherent with signal)

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

Phase diagram - random Gaussian F

Number of measure-ments per variable

$\alpha = M/N$



Donoho 2006, Donoho Tanner 2005

Kabashima, Wadayama and Tanaka, JSTAT 2009

Optimal rate

$\alpha = \rho$

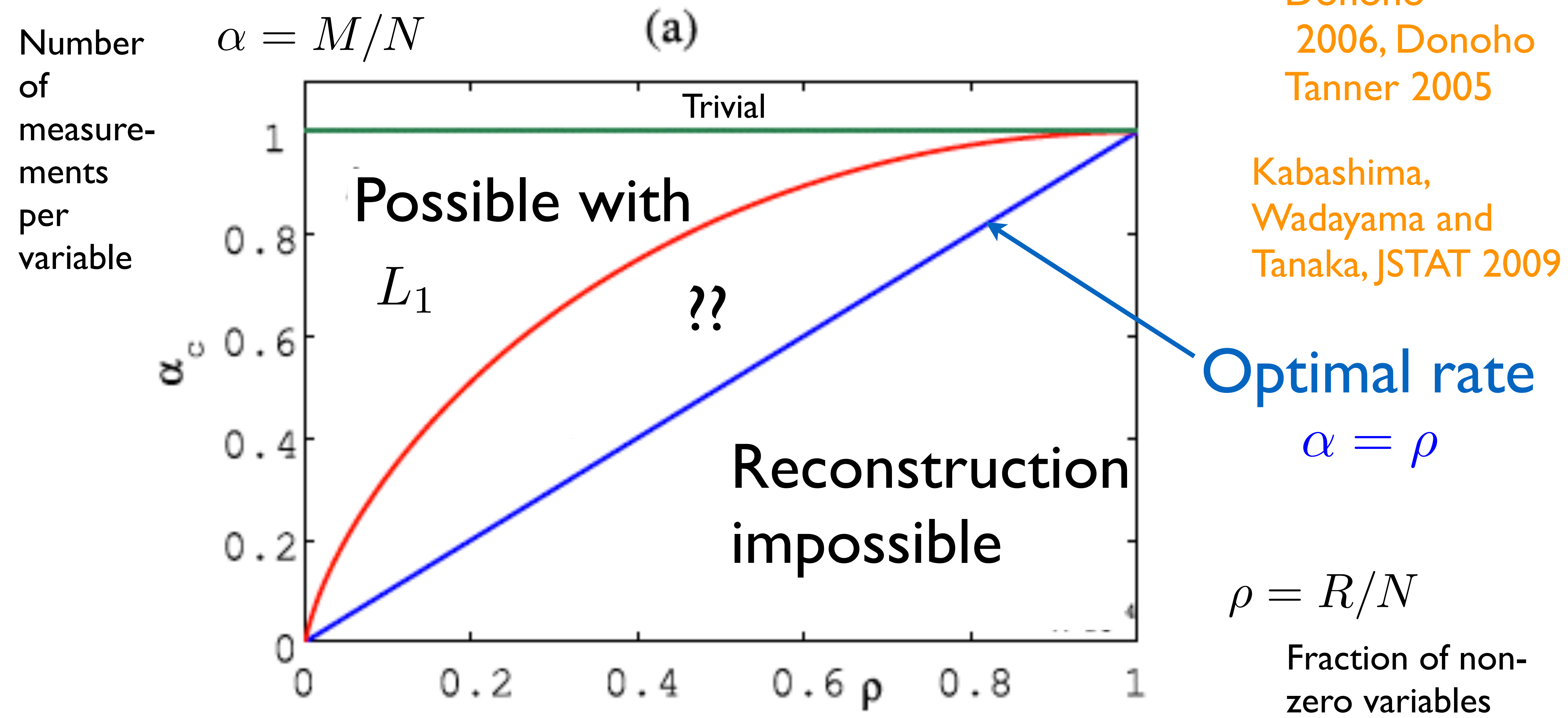
$\rho = R/N$

Fraction of non-zero variables

L_1: Find a N - component vector x such that the M equations $y = Fx$ are satisfied and $||x||$ is minimal

Gaussian random matrix

Phase diagram - random Gaussian F



Donoho 2006, Donoho Tanner 2005

Kabashima, Wadayama and Tanaka, JSTAT 2009

L_1: Find a N - component vector x such that the M equations $y = Fx$ are satisfied and $||x||$ is minimal

Gaussian random matrix

Alternative approach, able to reach the optimal rate $\alpha = \rho$

Krzakala Sausset Mézard Sun Zdeborova 2011

- Probabilistic approach
- Message passing reconstruction of the signal
- Careful design of the measurement matrix

NB: each of these three ingredients is crucial

Alternative approach, able to reach the optimal rate $\alpha = \rho$

Krzakala Sausset Mézard Sun Zdeborova 2011

- Probabilistic approach
- Message passing reconstruction of the signal
- Careful design of the measurement matrix

NB: each of these three ingredients is crucial

Assumption: original signal components are independent and sparse (in an appropriate basis)

$s_i = 0$ with probability $1 - \rho_0$

Step 1: Probabilistic approach to compressed sensing

Probability $P(x)$ that the signal is x :

I) x must be compatible with the measurements:

$$\sum_i F_{\mu i} x_i = y_\mu$$

II) A priori measure on x favours sparsity

« Gauss-Bernoulli » prior:

with probability ρ : $x_i = 0$

with probability $1-\rho$: drawn from Gaussian

Theorem: with this measure, the original signal $x = s$ is the most probable (even for wrong prior: not obvious!)

Step 2: Sampling from the constrained measure

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

Gaussian ϕ

Step 2: Sampling from the constrained measure

Prior

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

Gaussian ϕ

Step 2: Sampling from the constrained measure

Prior

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

Gaussian ϕ

constraints

Step 2: Sampling from the constrained measure

Prior

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

Gaussian ϕ

«Native configuration»= stored signal $x_i = s_i$ is infinitely more probable than other configurations.

Efficient sampling? Not so easy.

constraints

Step 2: Sampling from the constrained measure

Prior

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

Gaussian ϕ

constraints

«Native configuration»= stored signal $x_i = s_i$ is infinitely more probable than other configurations.
Efficient sampling? Not so easy.

«Mean field»:
belief propagation
(spin glass mean-field equations, TAP)

Step 2: Sampling from the constrained measure

Prior

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

Gaussian ϕ

«Native configuration»= stored signal $x_i = s_i$ is infinitely more probable than other configurations.

Efficient sampling? Not so easy.

constraints

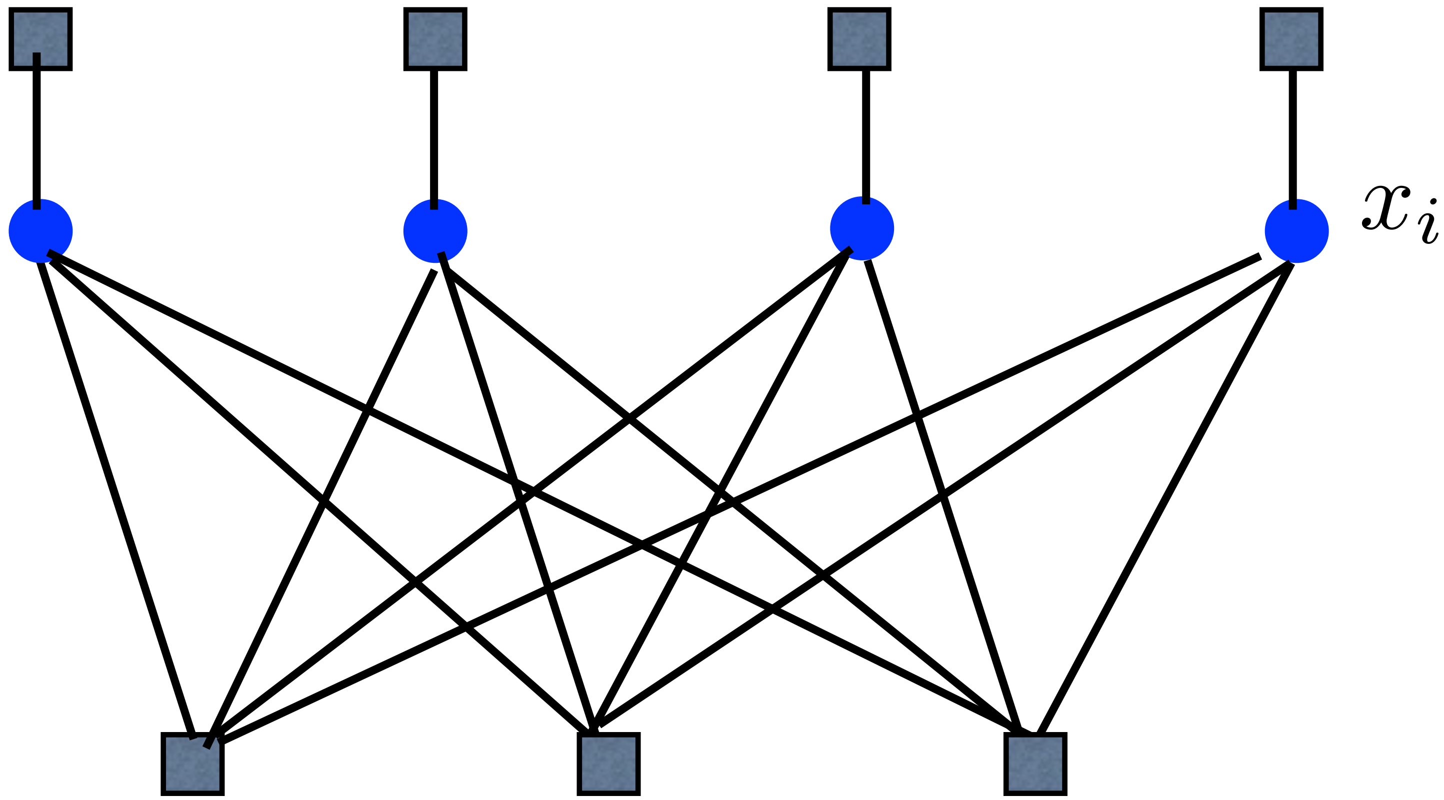
«Mean field»:
belief propagation
(spin glass mean-field equations, TAP)

Each constraint involves all the variables: «long-range» weak interactions (e.g. Curie Weiss model for magnets).

⟶ Mean field is exact*

Belief propagation = mean field equations

«Factor graph»

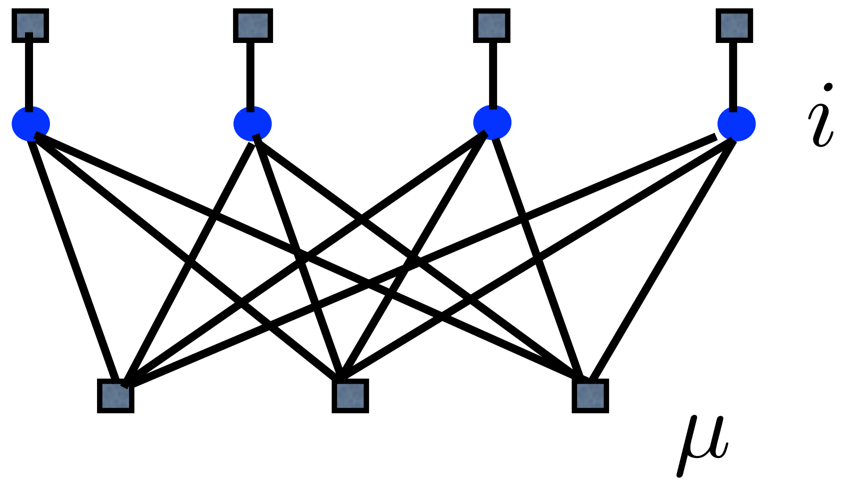


«variables»

constraints

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

Belief propagation = mean-field like equations



Local order parameters:

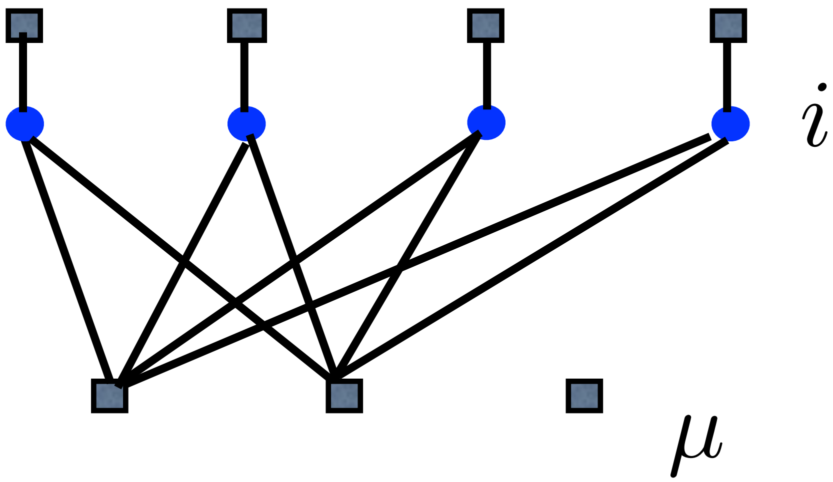
$$a_{i\to\mu} = \langle x_i \rangle_\mu$$

$$v_{i\to\mu} = \langle x_i^2 \rangle_\mu - \left(\langle x_i \rangle_\mu\right)^2$$

where $\langle \cdot \rangle_\mu$ denotes the mean, in absence of constraint μ

(«cavity»-type measure)

Belief propagation = mean-field like equations



Local order parameters:

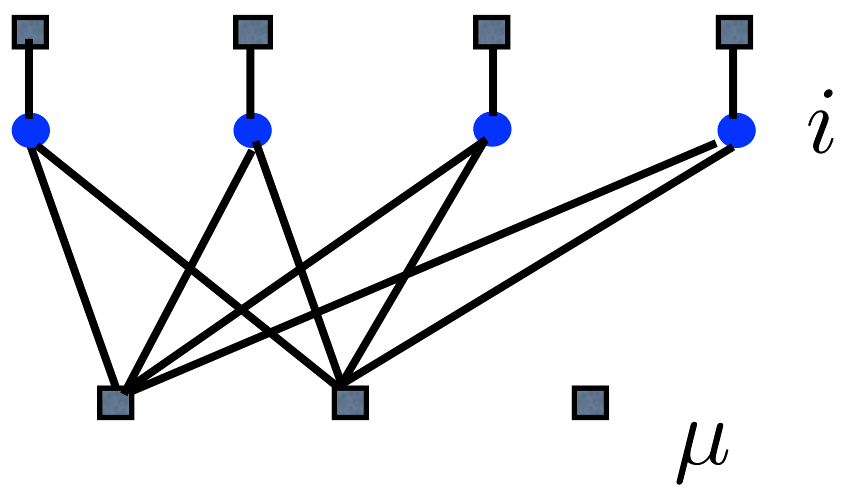
$$a_{i\to\mu} = \langle x_i \rangle_\mu$$

$$v_{i\to\mu} = \langle x_i^2 \rangle_\mu - \left(\langle x_i \rangle_\mu\right)^2$$

where $\langle \cdot \rangle_\mu$ denotes the mean, in absence of constraint μ

(«cavity»-type measure)

Belief propagation = mean-field like equations



Local order parameters:

$$a_{i\to\mu} = \langle x_i \rangle_\mu$$

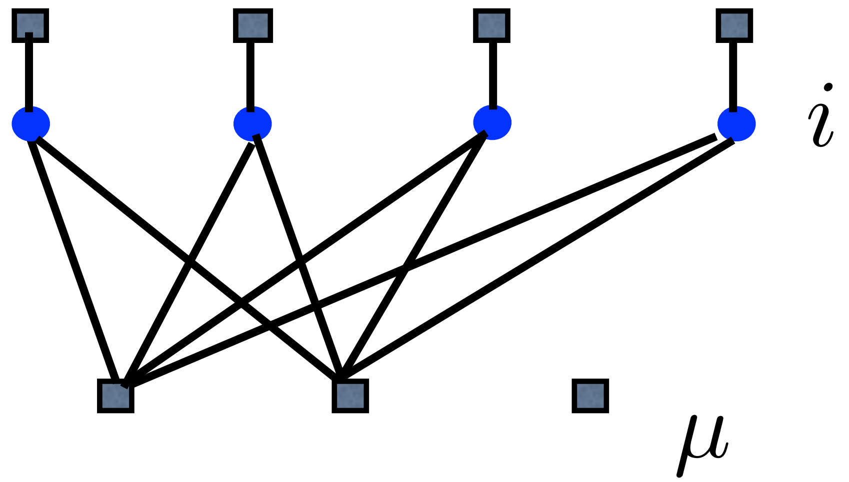
$$v_{i\to\mu} = \langle x_i^2 \rangle_\mu - \left(\langle x_i \rangle_\mu\right)^2$$

where $\langle \cdot \rangle_\mu$ denotes the mean, in absence of constraint μ

(«cavity»-type measure)

Closed self-consistent equations relating these order parameters («BP», «TAP», «G-AMP»,...)

Belief propagation = mean-field like equations



Local order parameters:

$$a_{i\to\mu} = \langle x_i \rangle_\mu$$

$$v_{i\to\mu} = \langle x_i^2 \rangle_\mu - \left(\langle x_i \rangle_\mu\right)^2$$

where $\langle \cdot \rangle_\mu$ denotes the mean, in absence of constraint μ

(«cavity»-type measure)

Closed self-consistent equations relating these order parameters («BP», «TAP», «G-AMP»,...)

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

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

Technical parenthesis

Gaussian-projected BP («relaxed-BP»)

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

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

... (TAP +cavity method for SK model)...,
Kabashima Saad,
Guo Wang,
Rangan ⟶ CS

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

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

Large connectivity: simplification by projection of the messages on their first two moments

Technical parenthesis

Gaussian-projected BP («relaxed-BP»)

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

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

... (TAP +cavity method for SK model)...,
Kabashima Saad,
Guo Wang,
Rangan ⟶ CS

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

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

Large connectivity: simplification by projection of the messages on their first two moments

Performance of the probabilistic approach + message passing + parameter learning

- Simulations
- Analytic study of the large N limit
(replica method, cavity method)

Analytic study: cavity equations, density evolution, replicas, state evolution

Replica method allows to compute the «free entropy»

$$\Phi(D) = \lim_{N \to \infty} \frac{1}{N} \log P(D)$$

where $P(D)$ is the probability that reconstructed x is at distance D from original signal s.

$$D = \frac{1}{N} \sum_i (x_i - s_i)^2$$

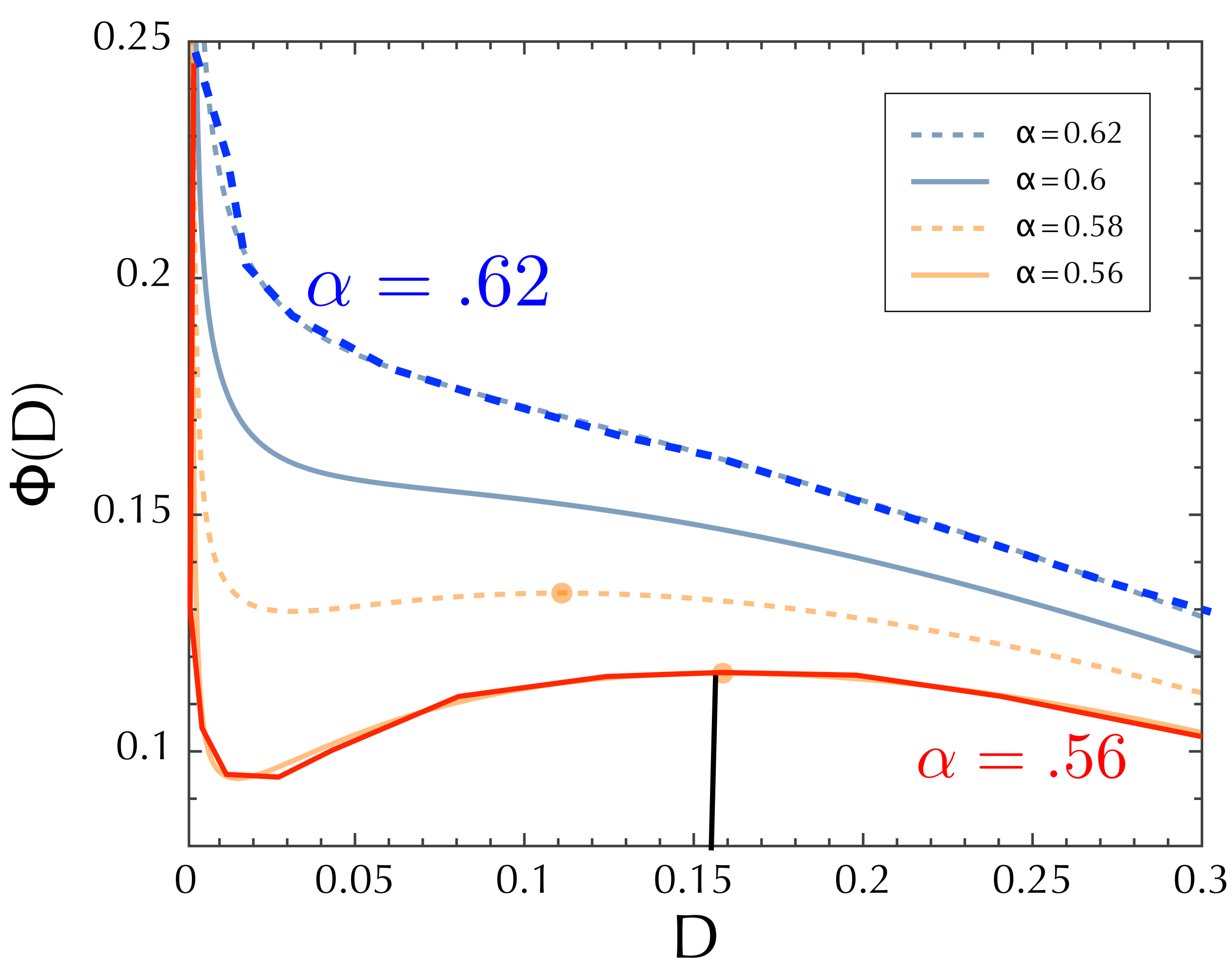
Cavity method shows that the order parameters of the BP iteration flow according to the gradient of the replica free entropy Φ («density evolution» eqns)

→ analytic control of the BP equations

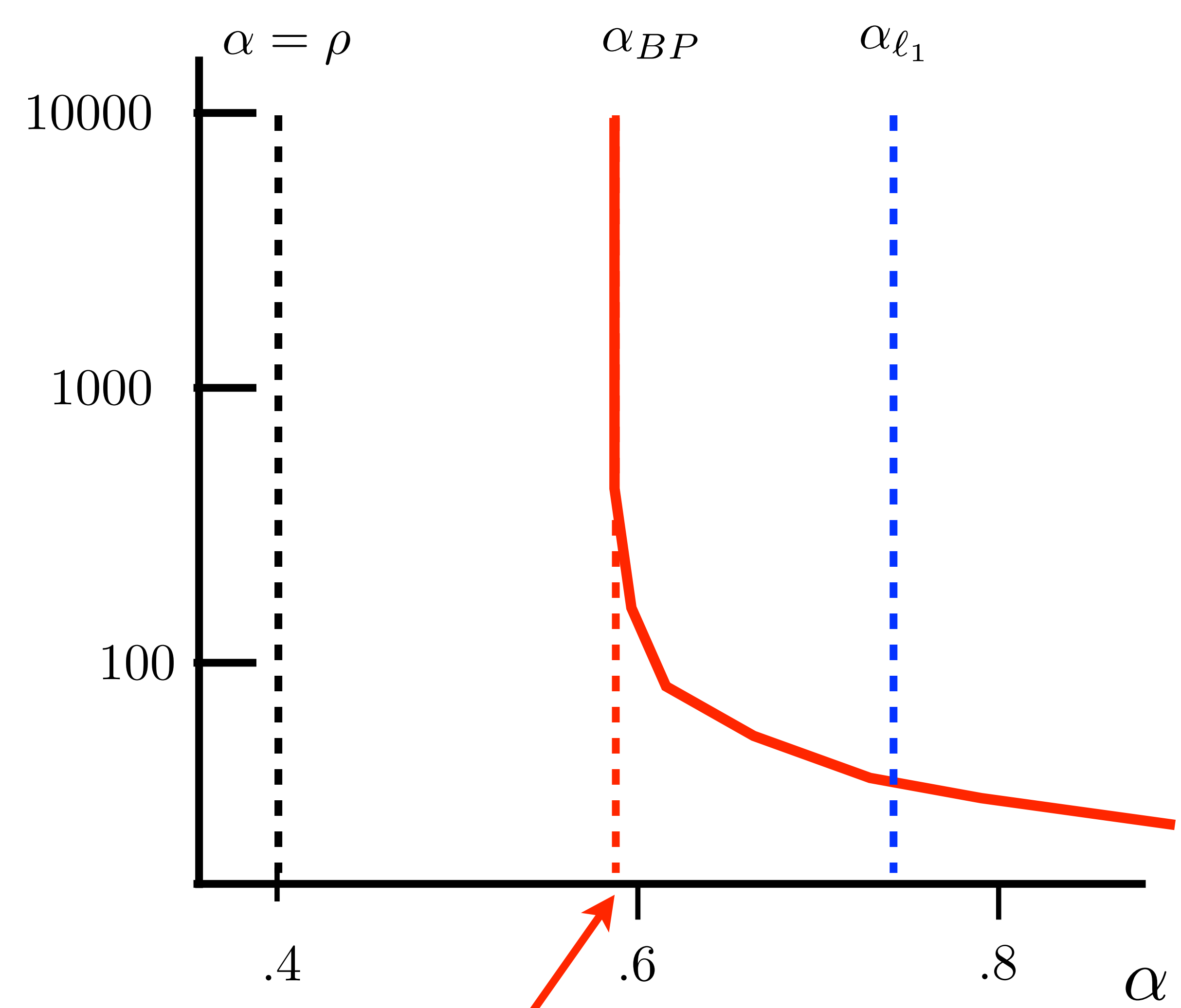
NB rigorous: Bayati Montanari, Lelarge Montanari

Free entropy $\sim \log P(D)$

BP convergence time $\rho_0 = .4$



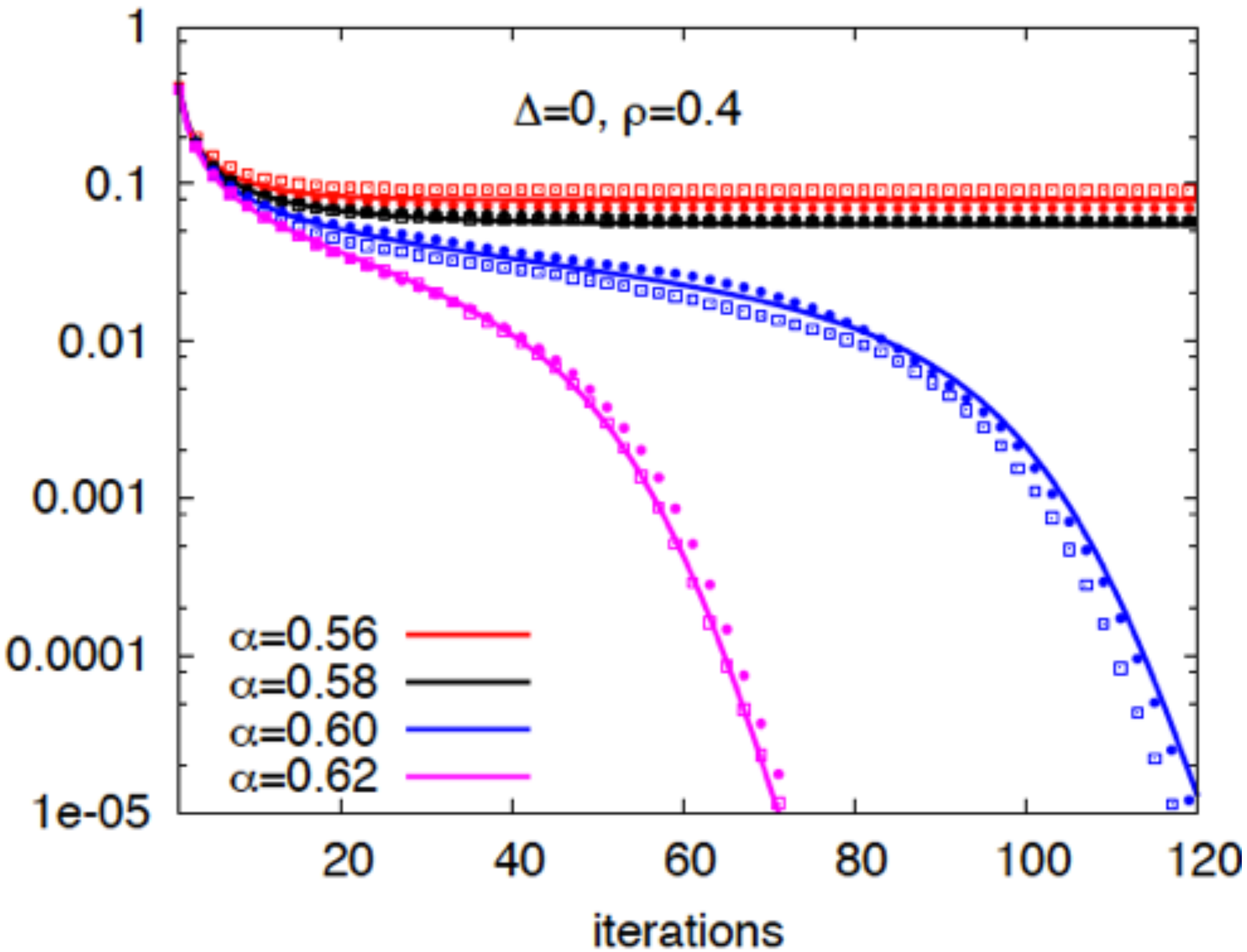
distance to native state



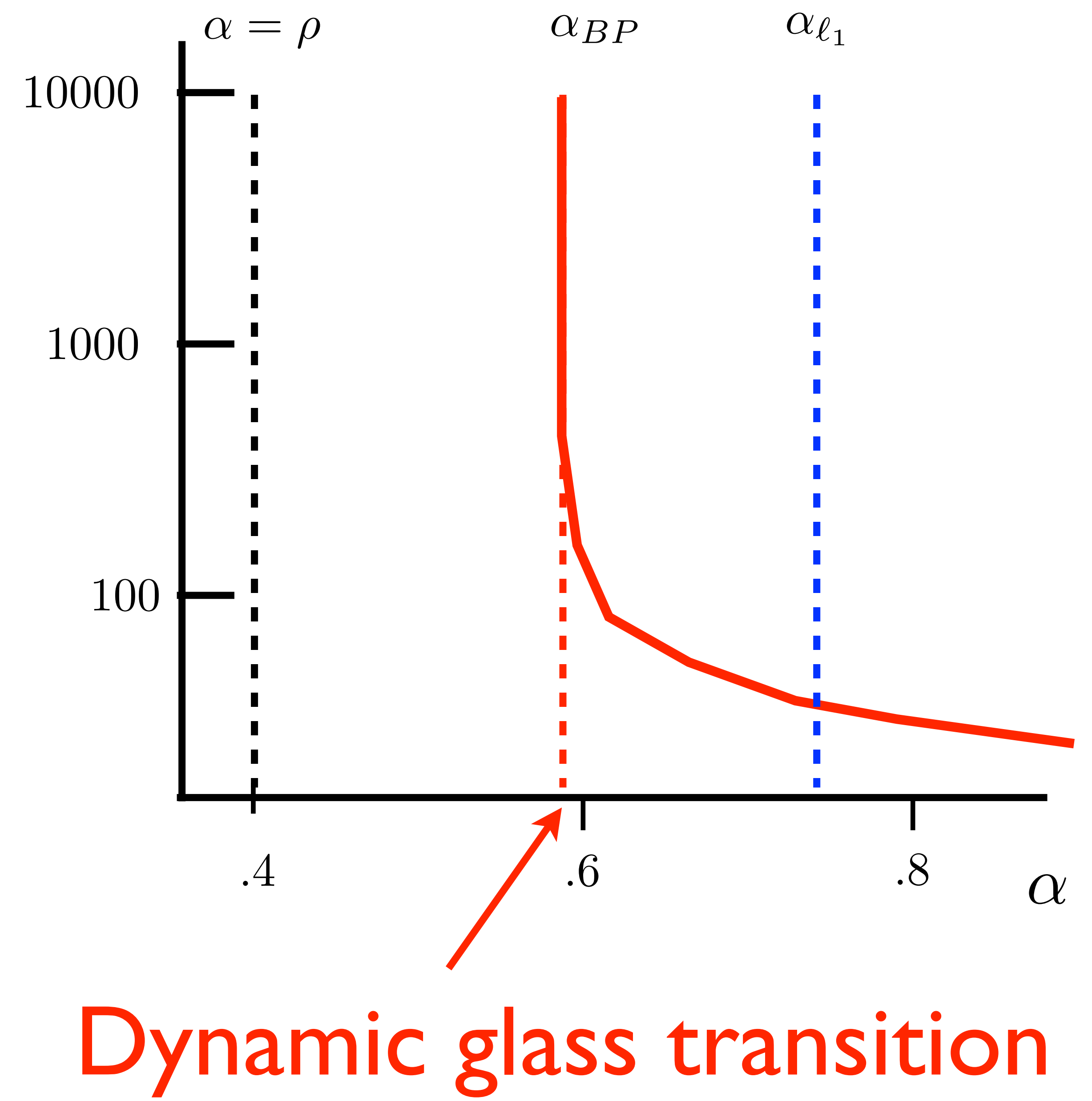
Dynamic glass transition

When α is too small, BP is trapped in a **glass phase**

Error

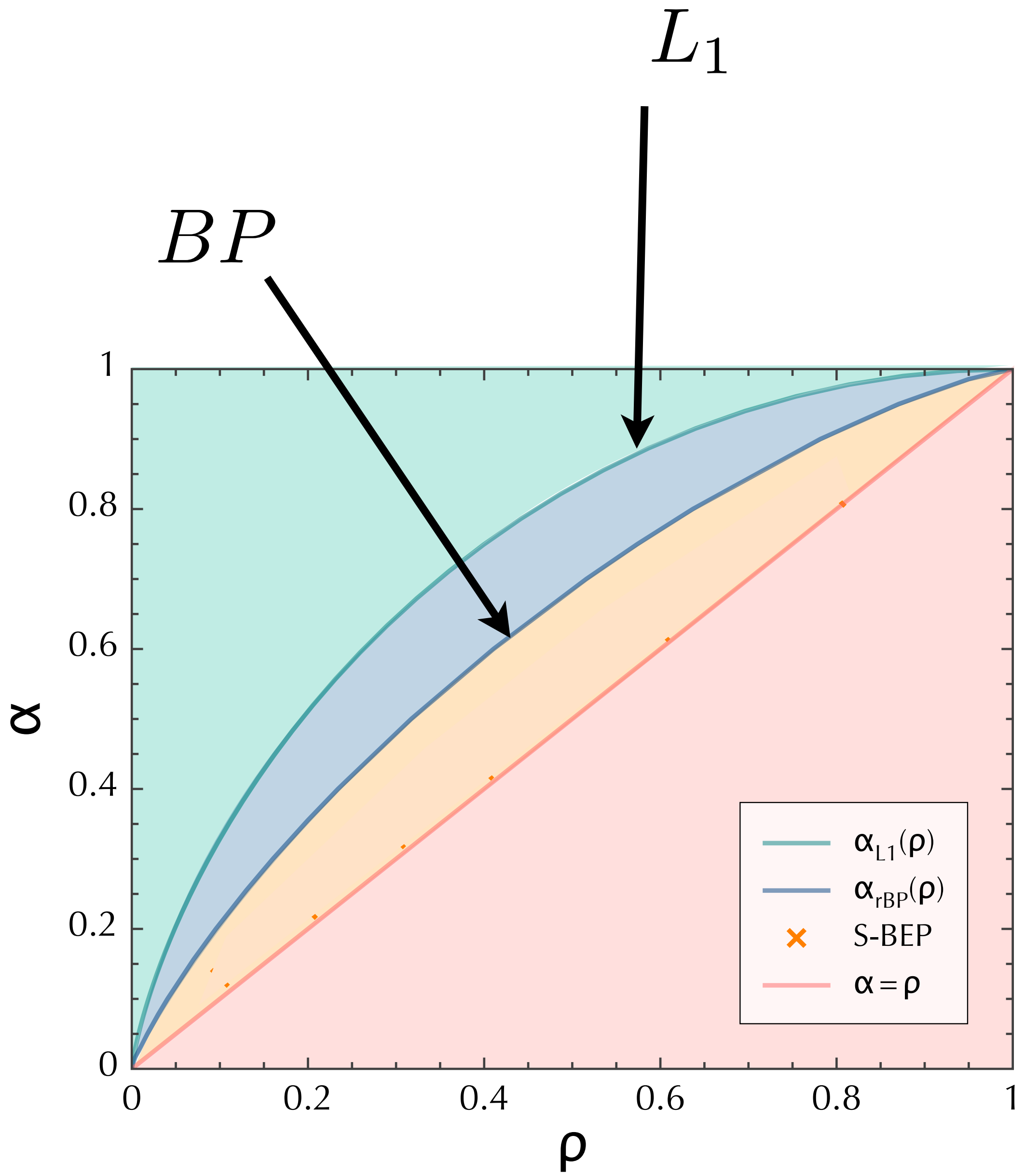


BP convergence time $\rho_0 = .4$

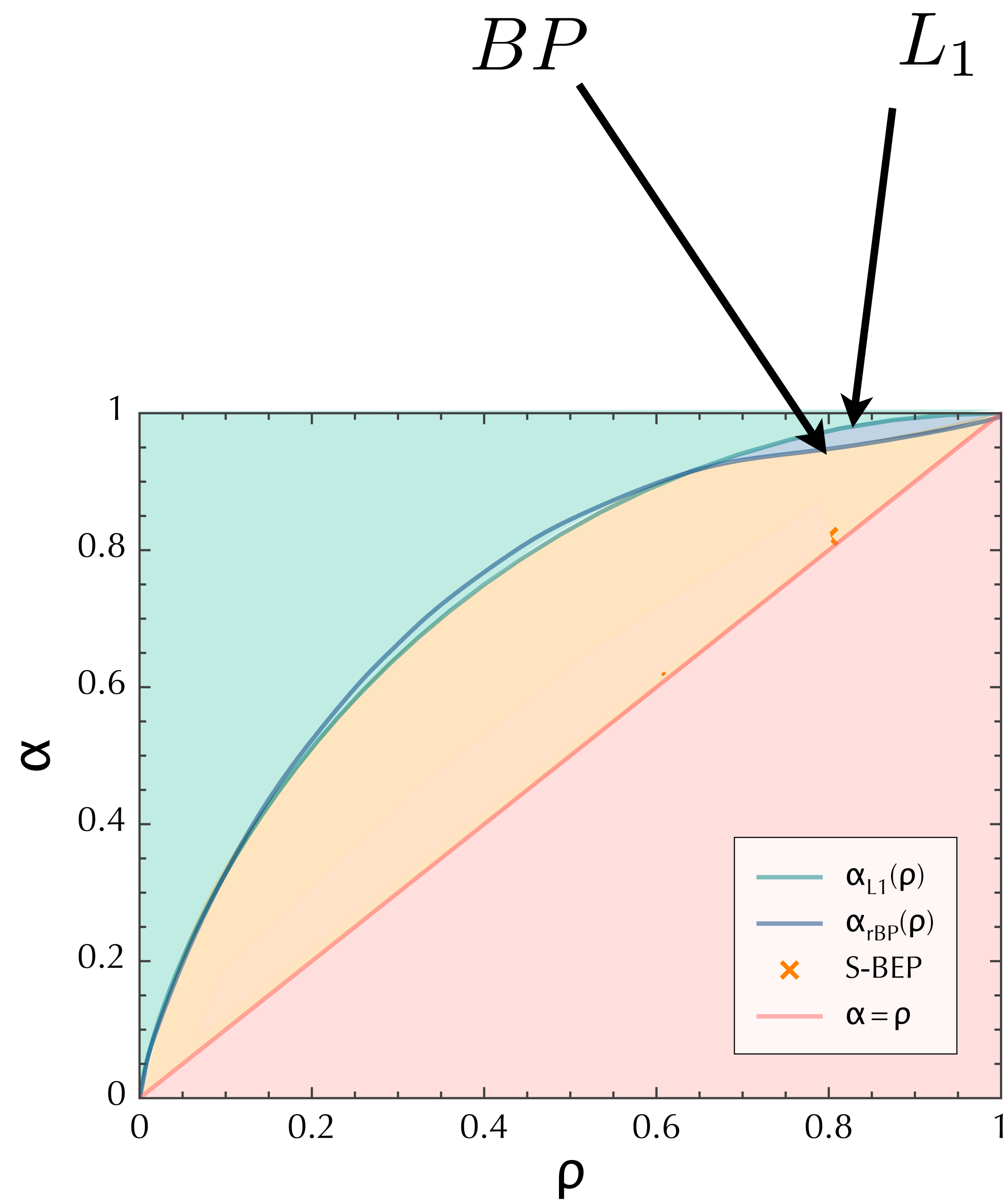


NB comparison of theory (replica, cavity, density evolution) and numerical experiment

Performance of BP with parameter learning: phase diagram



Gaussian signal



Binary signal

Step 3: design the measurement matrix in order to get around the glass transition

Getting around the glass trap: design the matrix F so that one nucleates the naive state (crystal nucleation idea,
...borrowed from error correcting codes!)

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

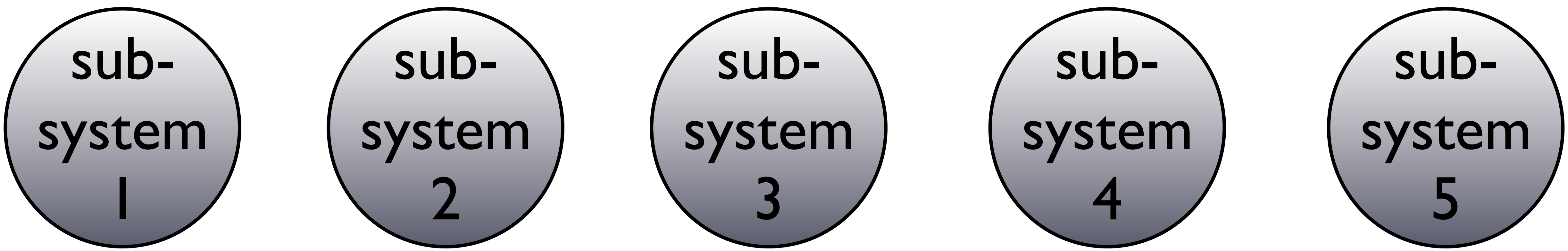
«Seeded BP» ; «Spatial coupling »

Nucleation and seeding

How to help the system find the « crystal » , getting around the glass trap?

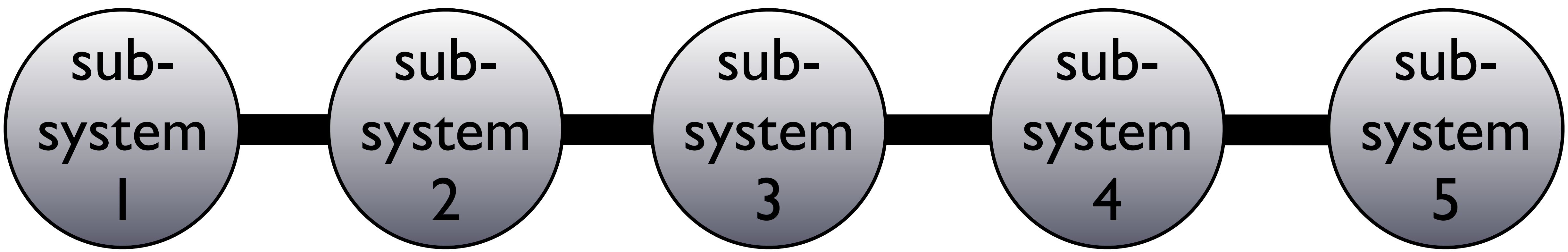
Mixed "mean-field" and one-dimensional system:

1) Create many "mean-field" sub-systems



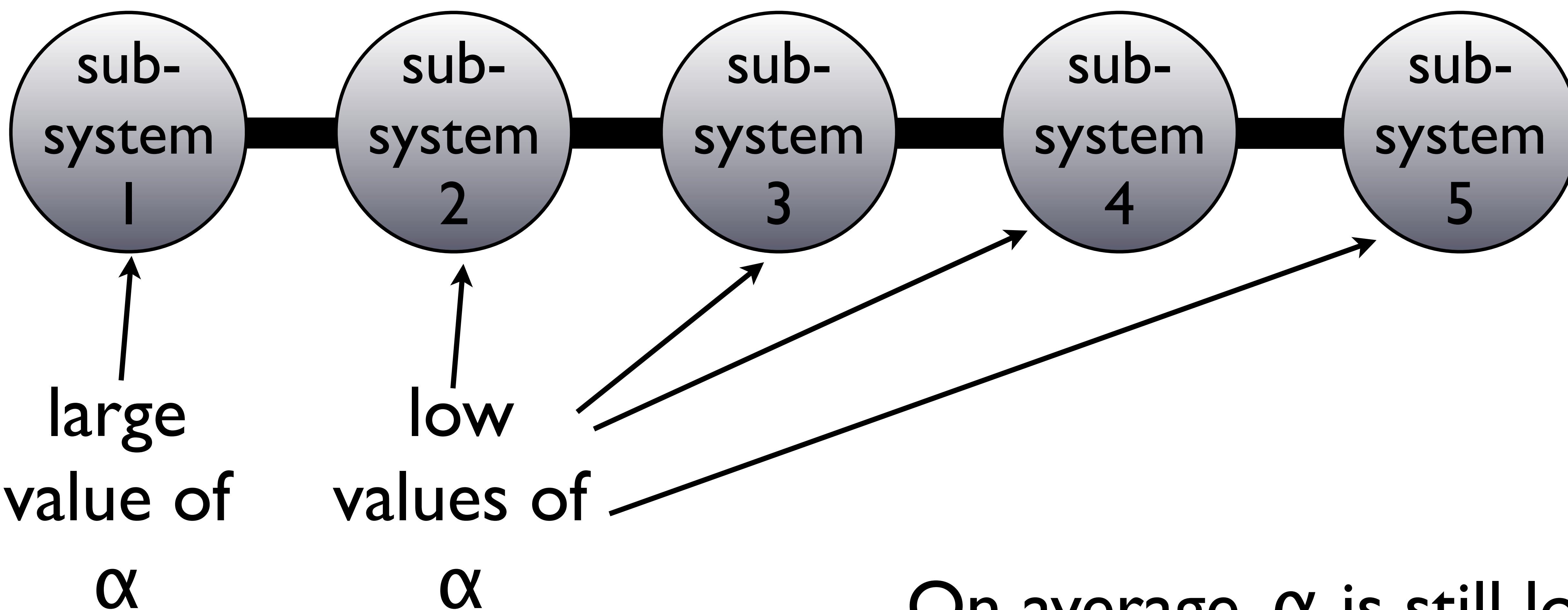
Mixed "mean-field" and one-dimensional system:

2) Add a first neighbor coupling



Mixed "mean-field" and one-dimensional system:

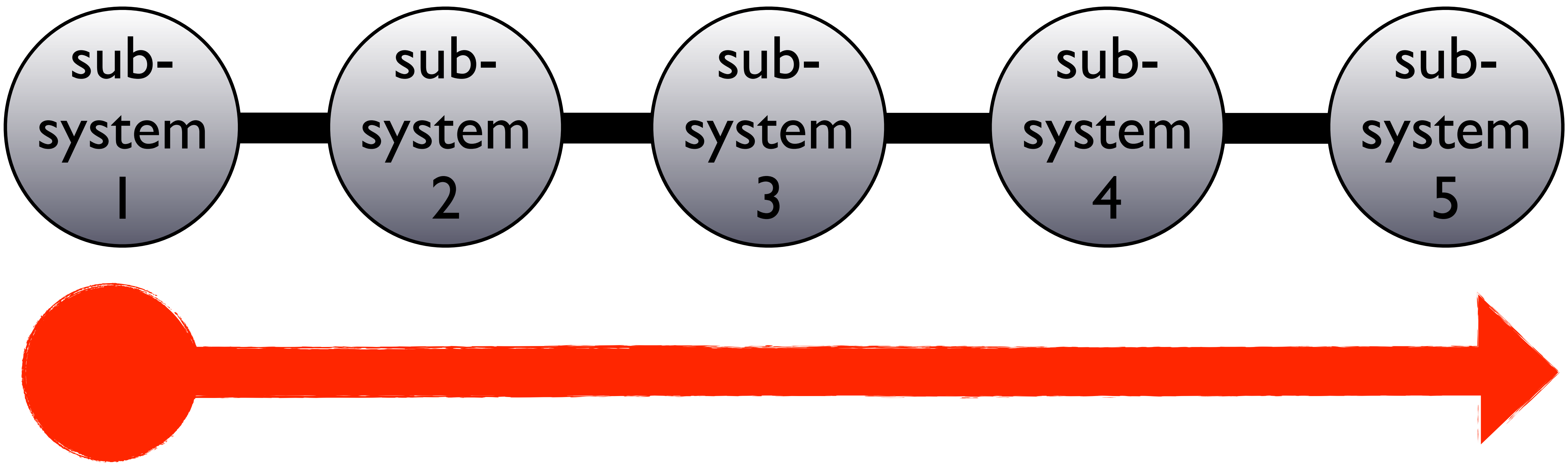
3) Choose parameters such that the first system is in the region of the phase diagram where there is no metastability

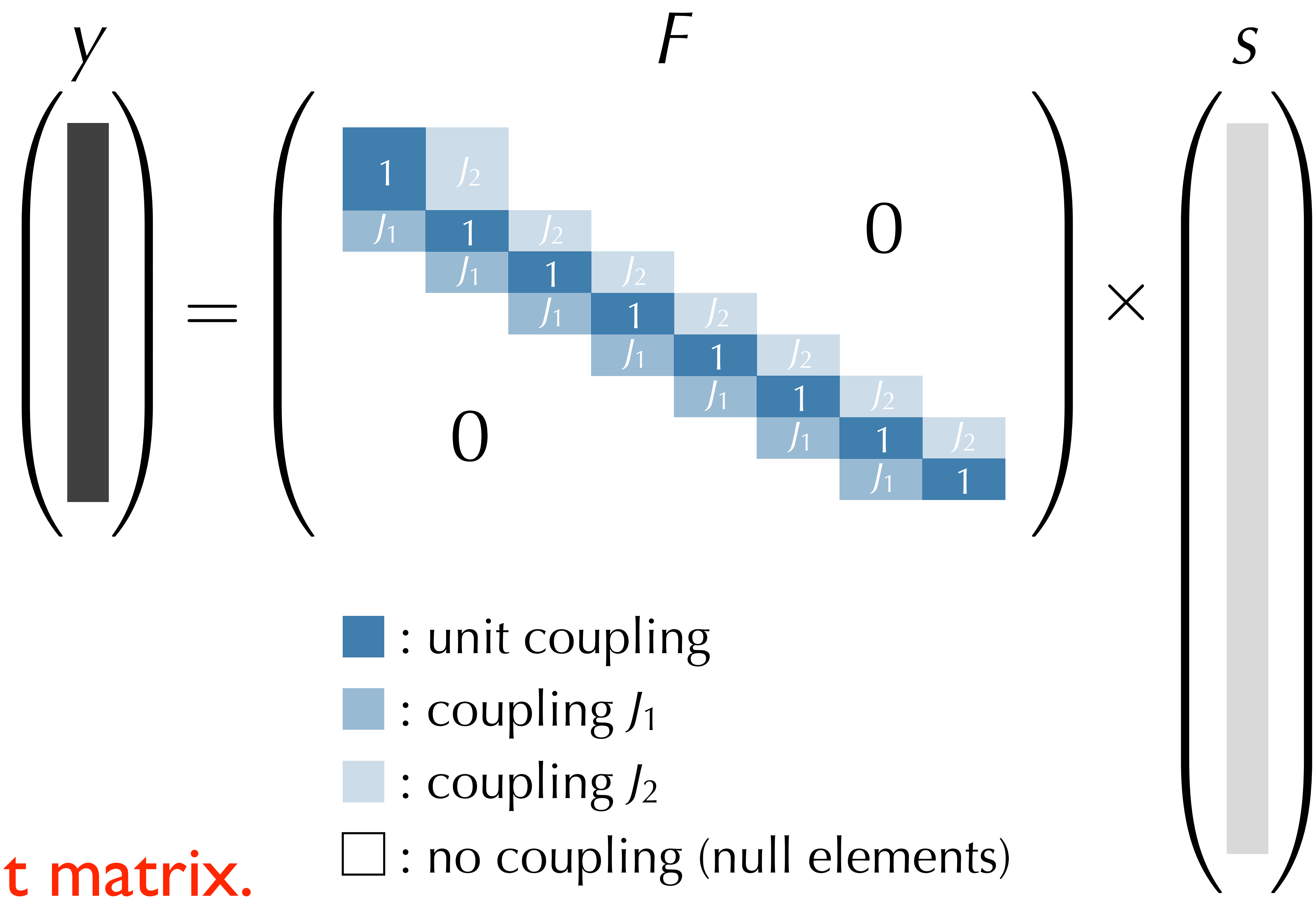


On average, α is still low !

Mixed "mean-field" and one-dimensional system:

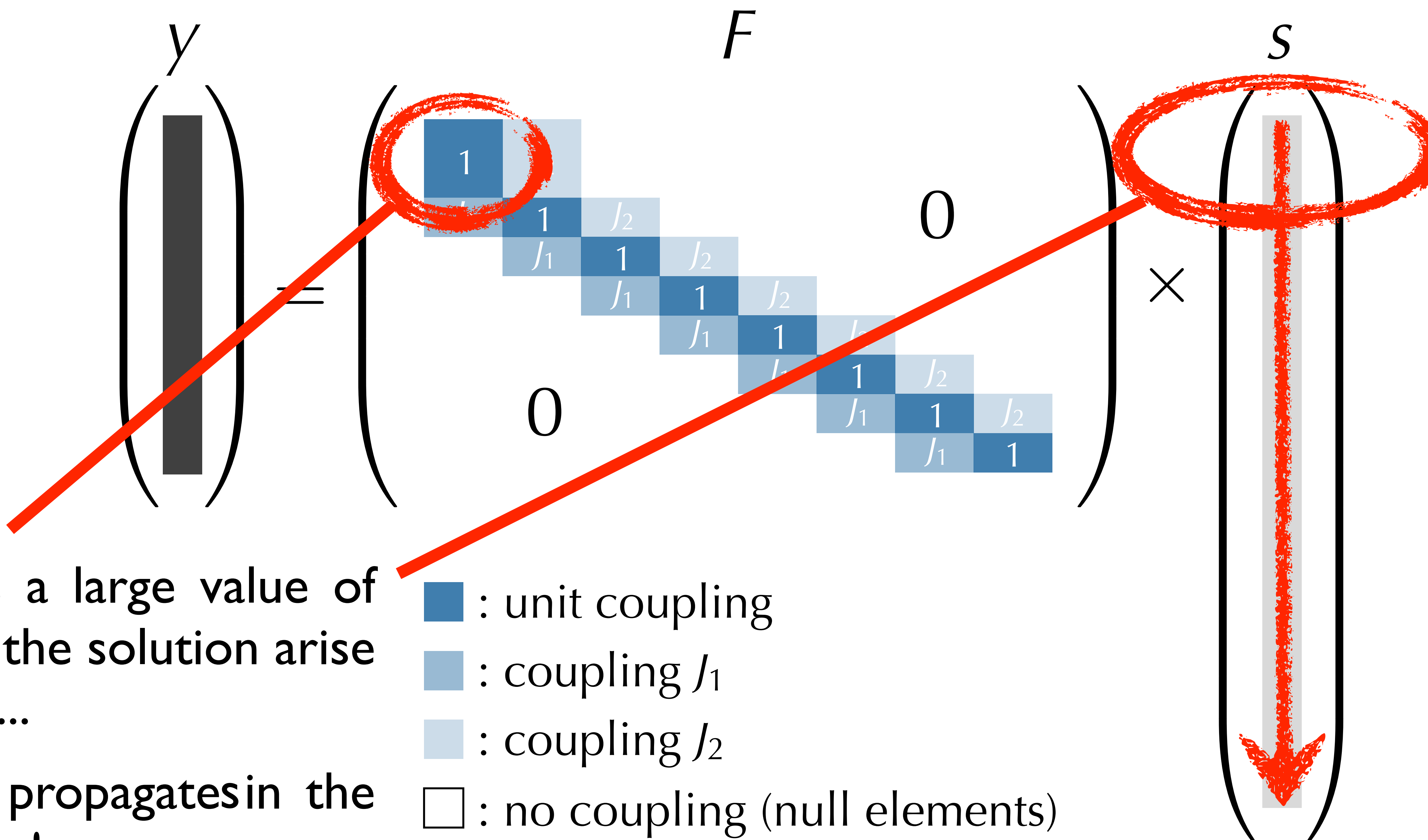
4) The solution will appear in the first sub-system (with large α), and then propagate in the system





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 propagates in the whole system!

$$L = 8$$

$$N_i = N/L$$

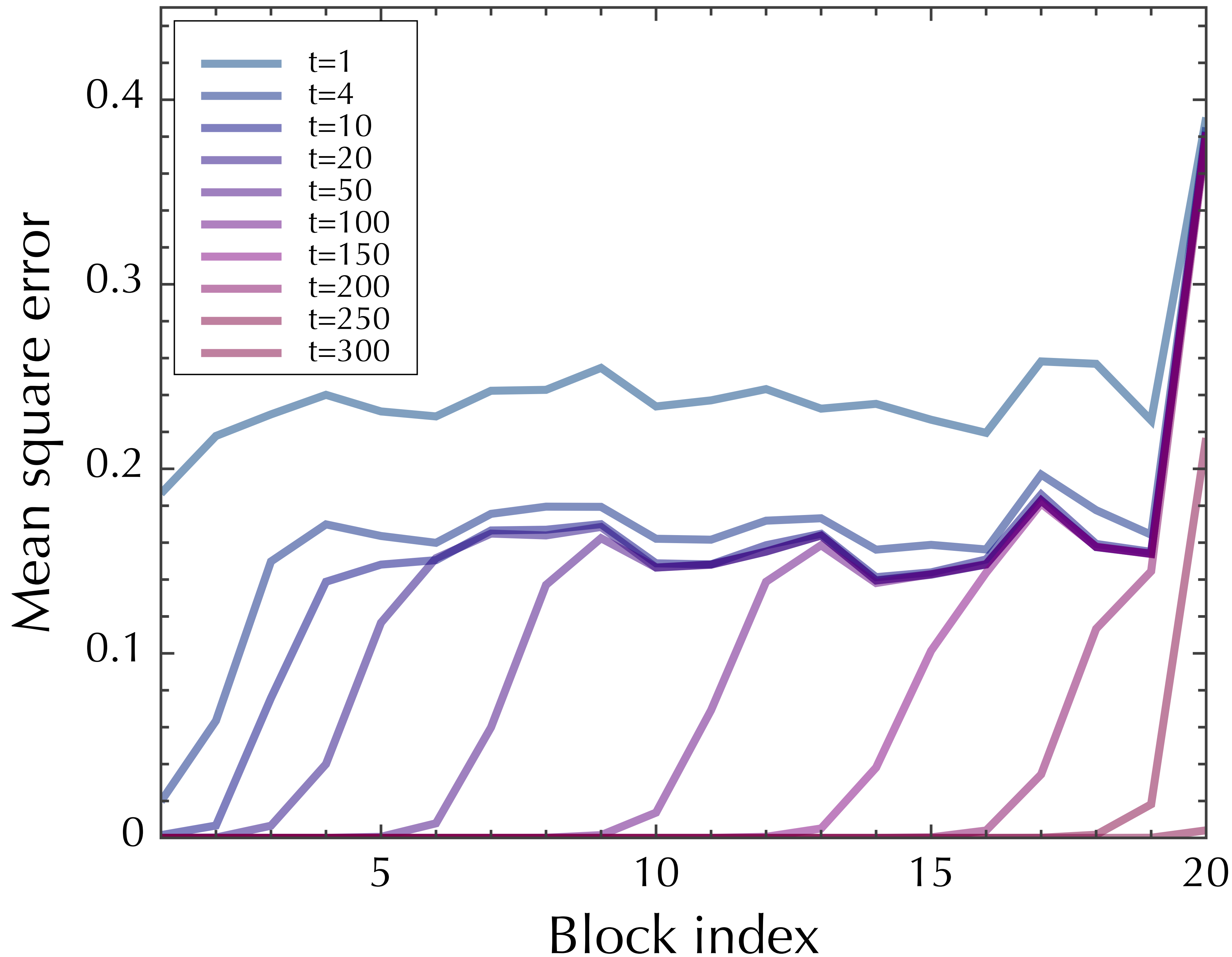
$$M_i = \alpha_i N/L$$

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

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

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

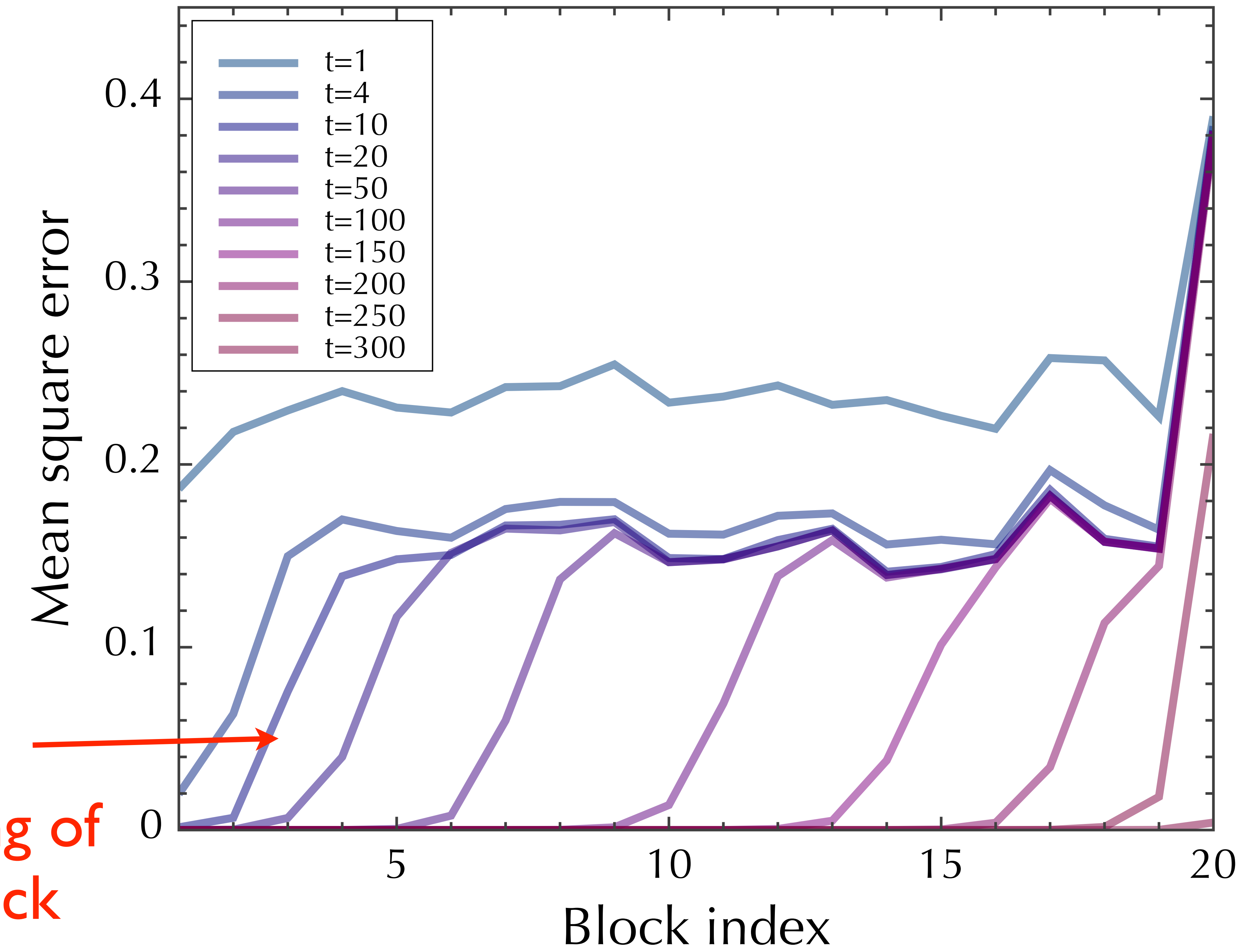
Numerical study



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

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

Numerical study



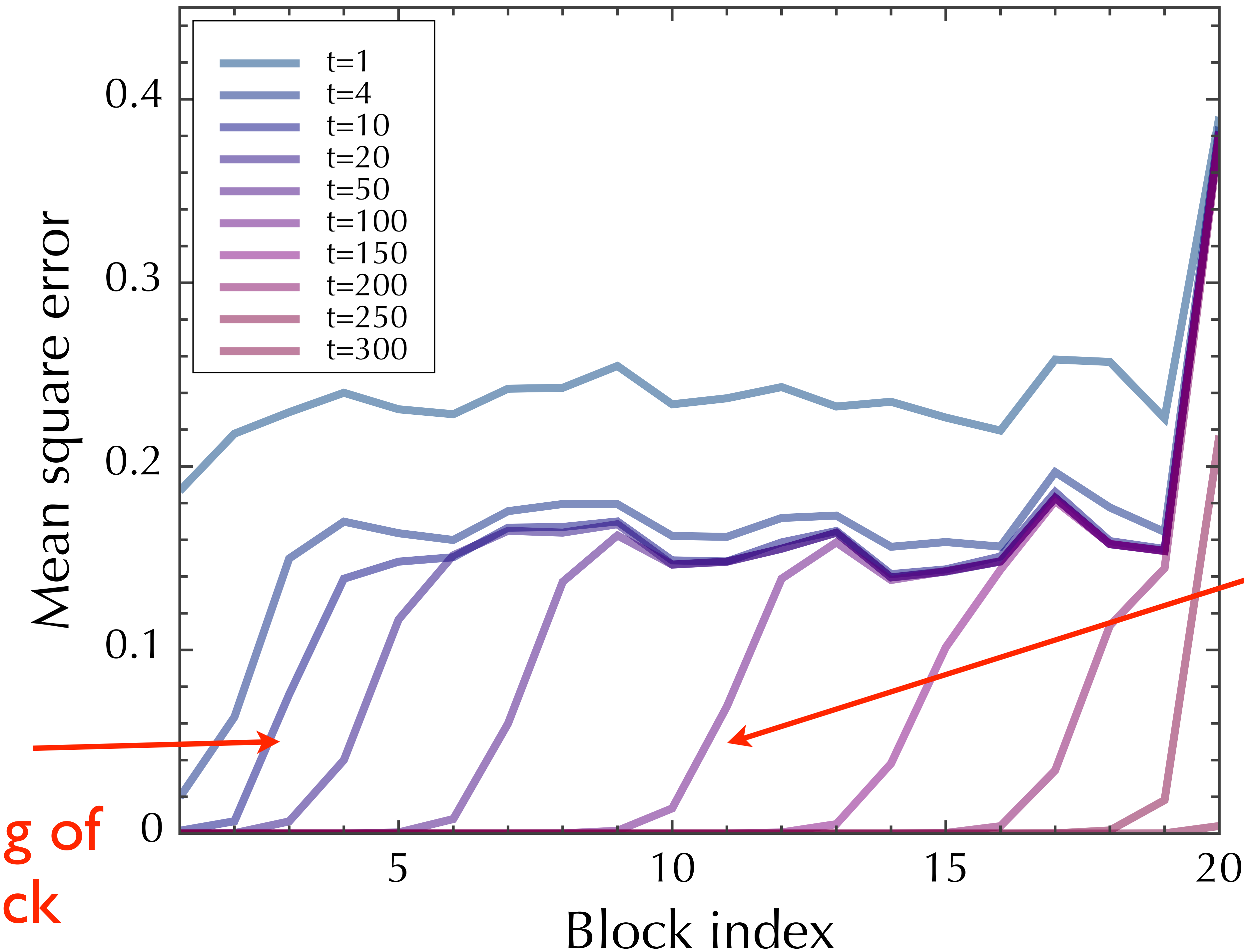
$t = 10$

decoding of first block

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

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

Numerical study



$t = 10$ decoding of first block

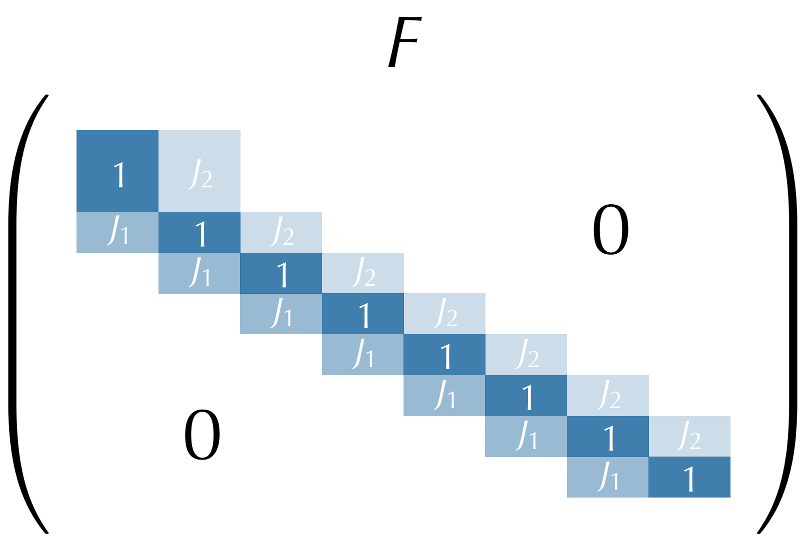
$t = 100$ decoding of blocks 1 to 9

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

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

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

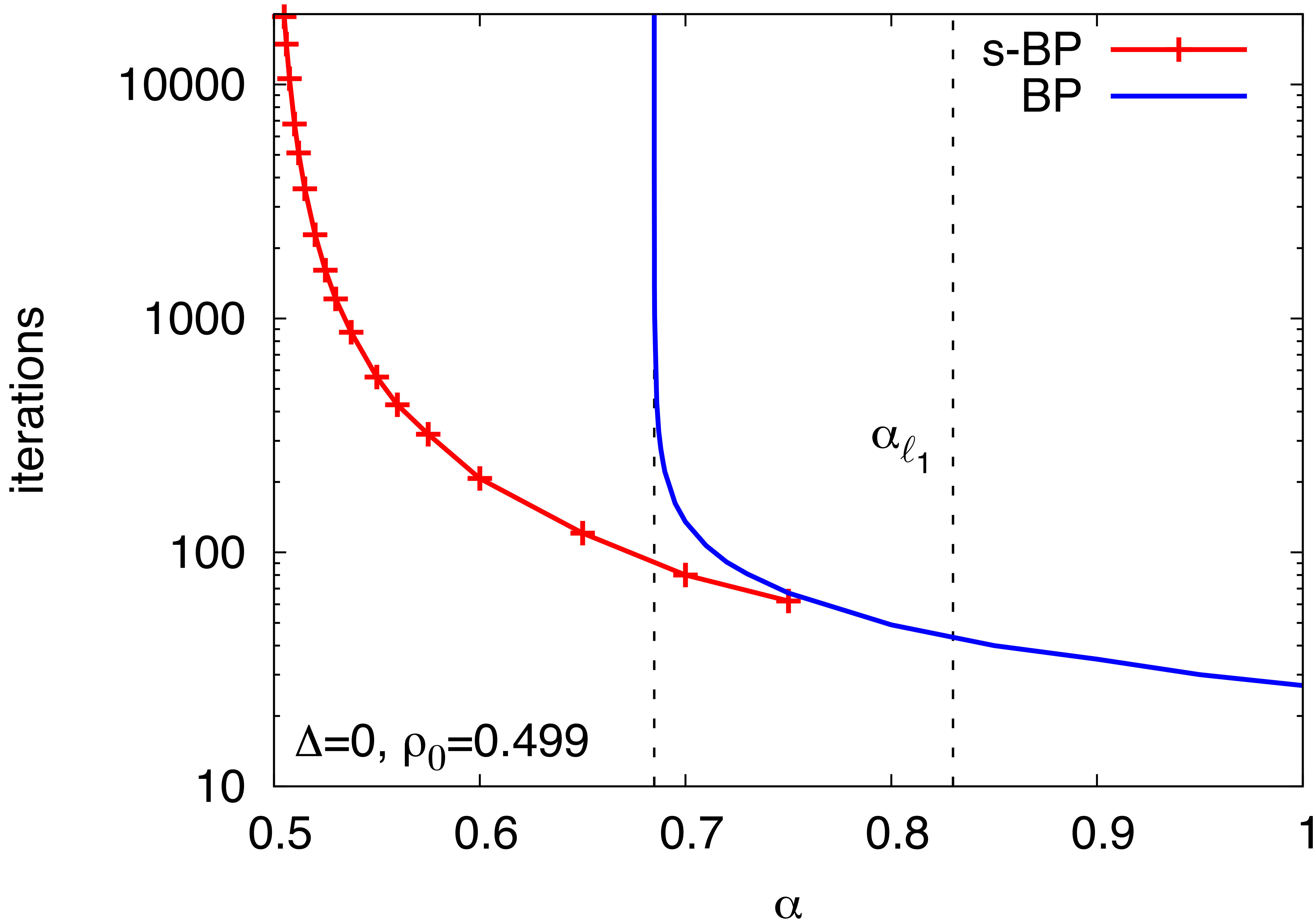
$$Z = \int \prod_{j=1}^{N} \mathrm{d}x_j \prod_{i=1}^{N} [(1-\rho)\delta(x_i) + \rho\phi(x_i)] \prod_{\mu=1}^{M} \delta\left(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i\right)$$



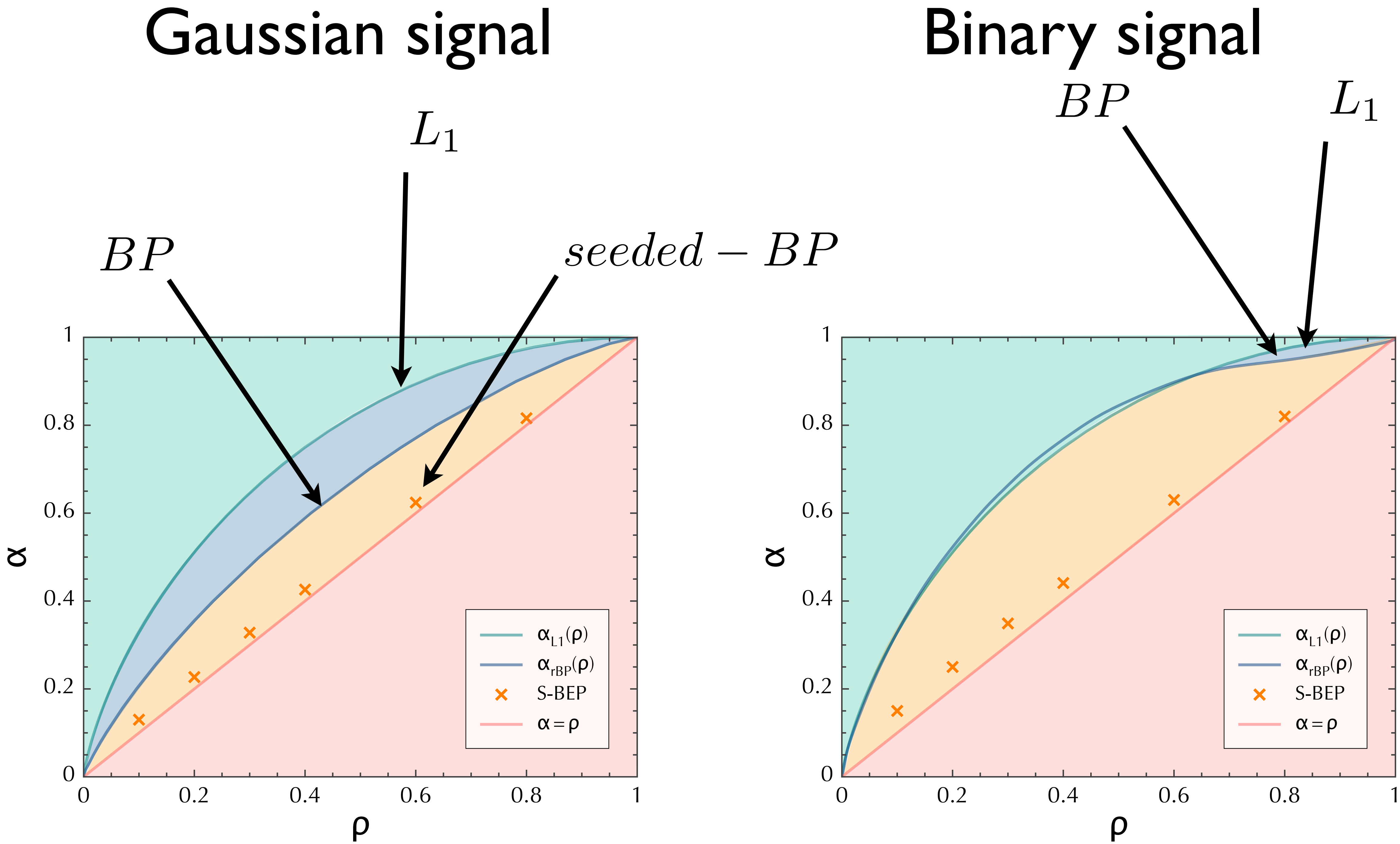
- Simulations
- Analytic approaches (replicas and cavity)

$$\rightarrow \alpha_c = \rho_0$$

Recent proof: Donoho Javanmard Montanari



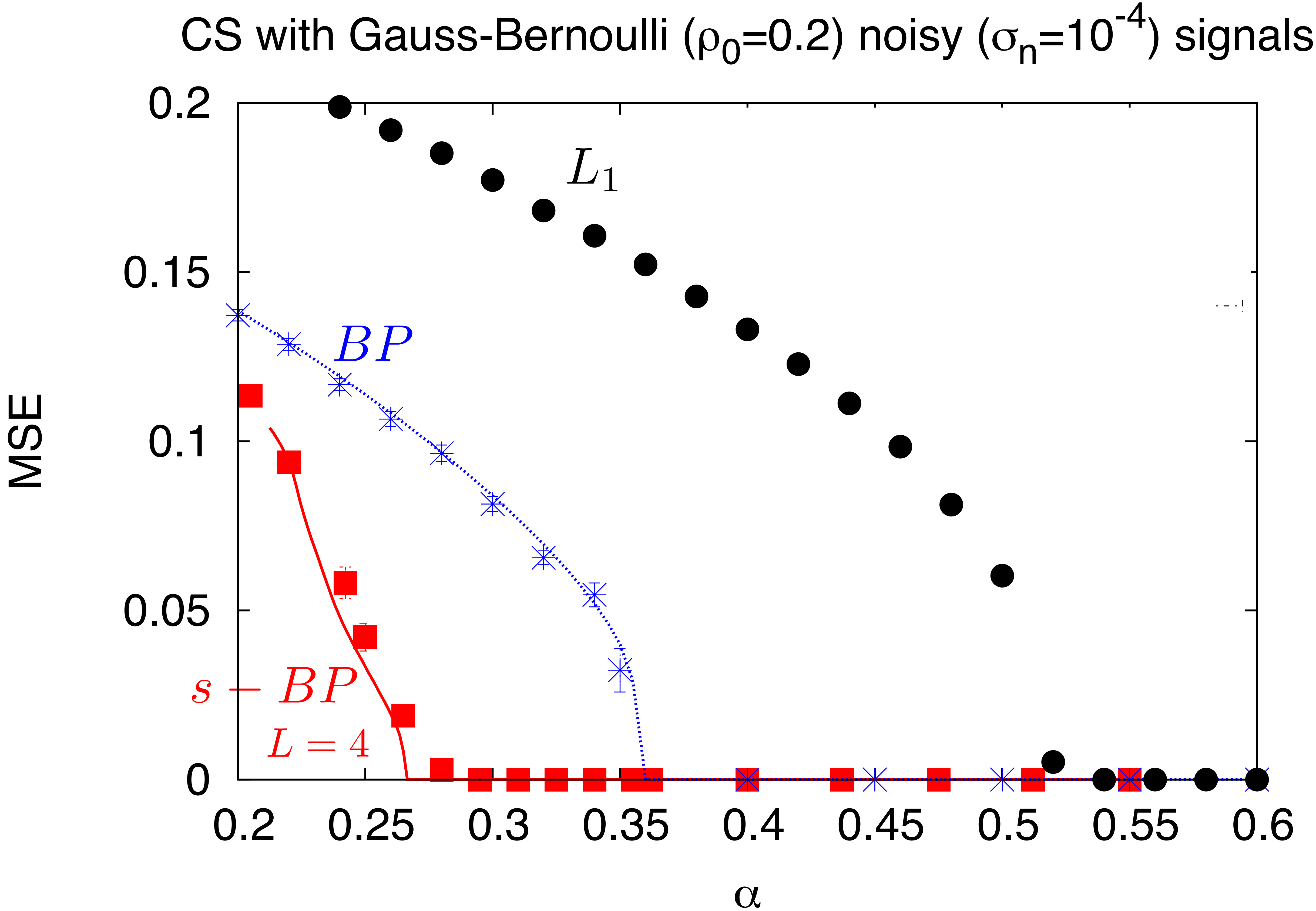
s-BP
BP
iterations
10000
1000
100
10
0.5
0.6
0.7
0.8
0.9
1
α
α_{ℓ_1}
Δ=0, ρ_0=0.499



Theory: seeded-BP threshold at $\alpha = \rho$ when $L \to \infty$

L_1 phase transition line moves up when using seeding F

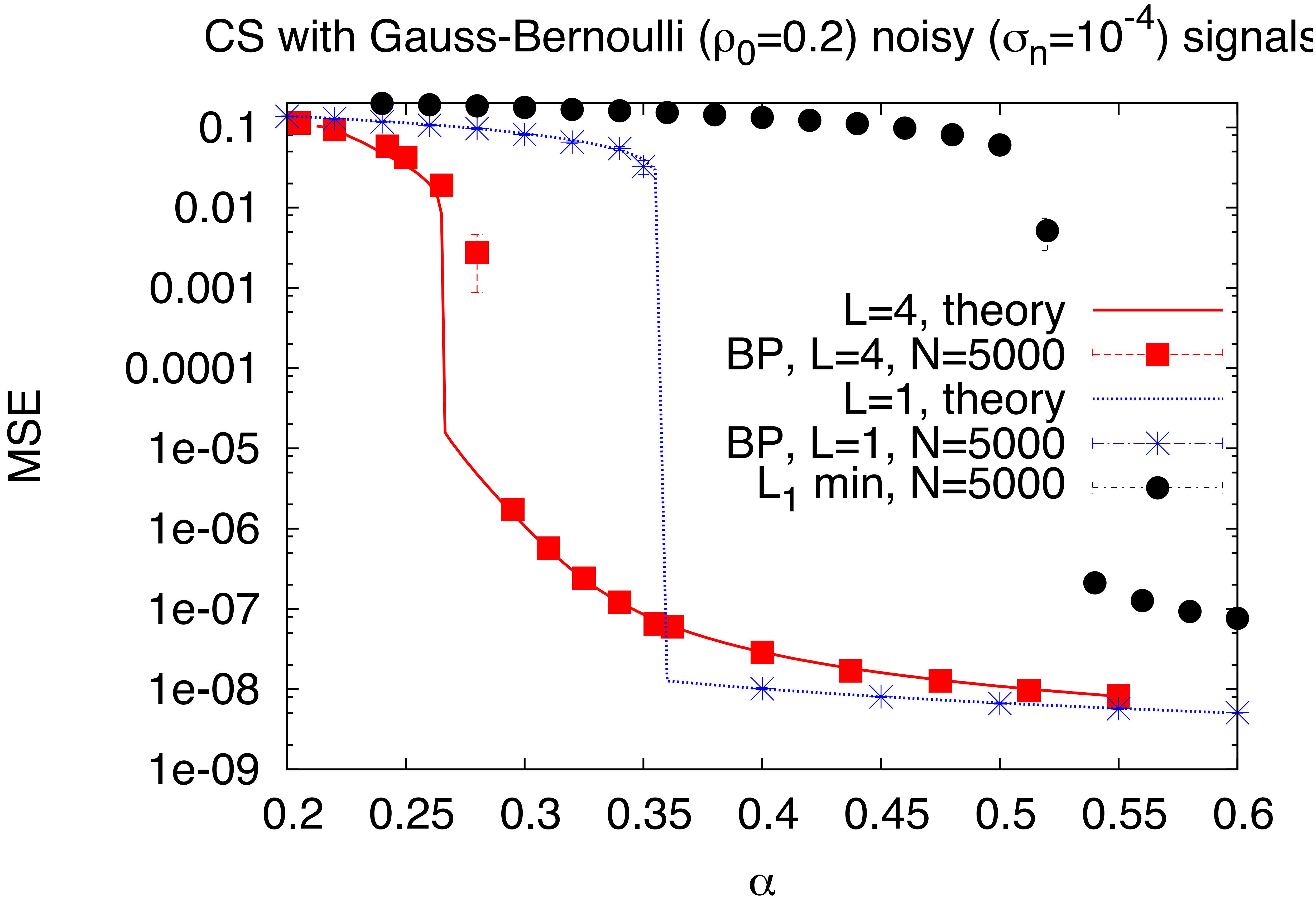
Noise



CS with Gauss-Bernoulli (ρ_0=0.2) noisy (σ_n=10^{-4}) signals
MSE
0.2
0.15
0.1
0.05
0
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
α
L_1
BP
$s-BP$
$L=4$

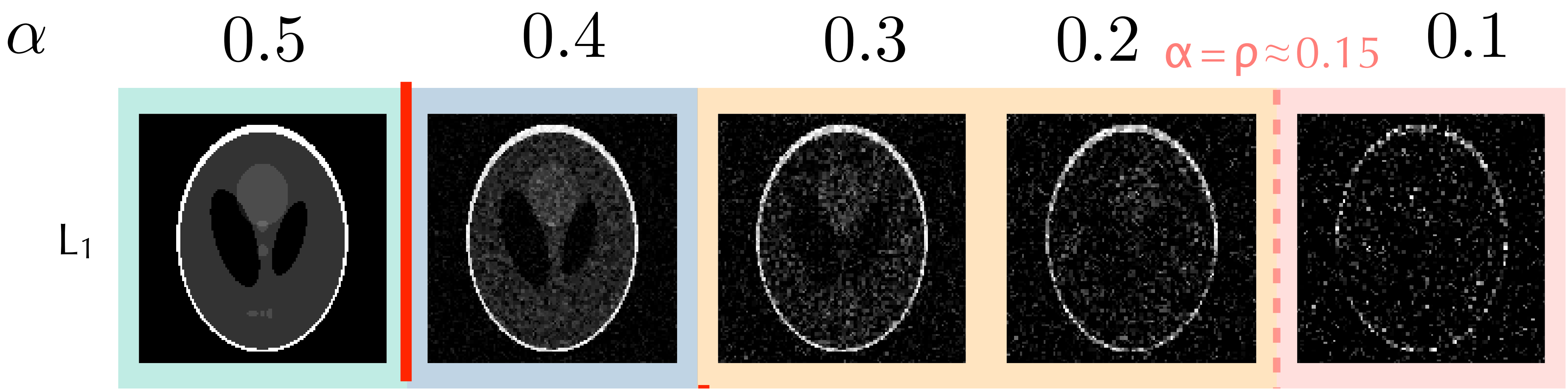
$N = 5000$

Noise

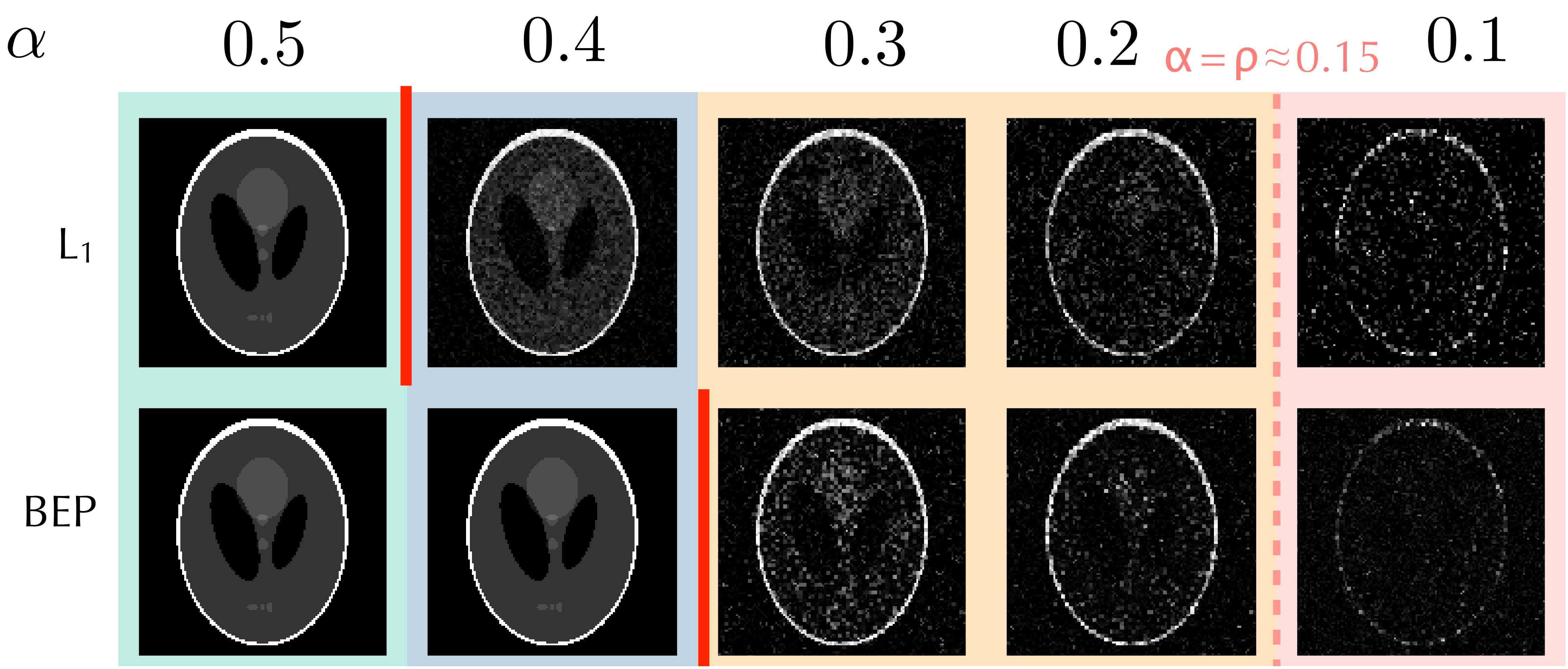

CS with Gauss-Bernoulli (ρ_0=0.2) noisy (σ_n=10^{-4}) signals
MSE
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
1e-08
1e-09
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
α
L=4, theory
BP, L=4, N=5000
L=1, theory
BP, L=1, N=5000
L_1 min, N=5000

Optimal performance on artificial signals (sparse but with independent components).

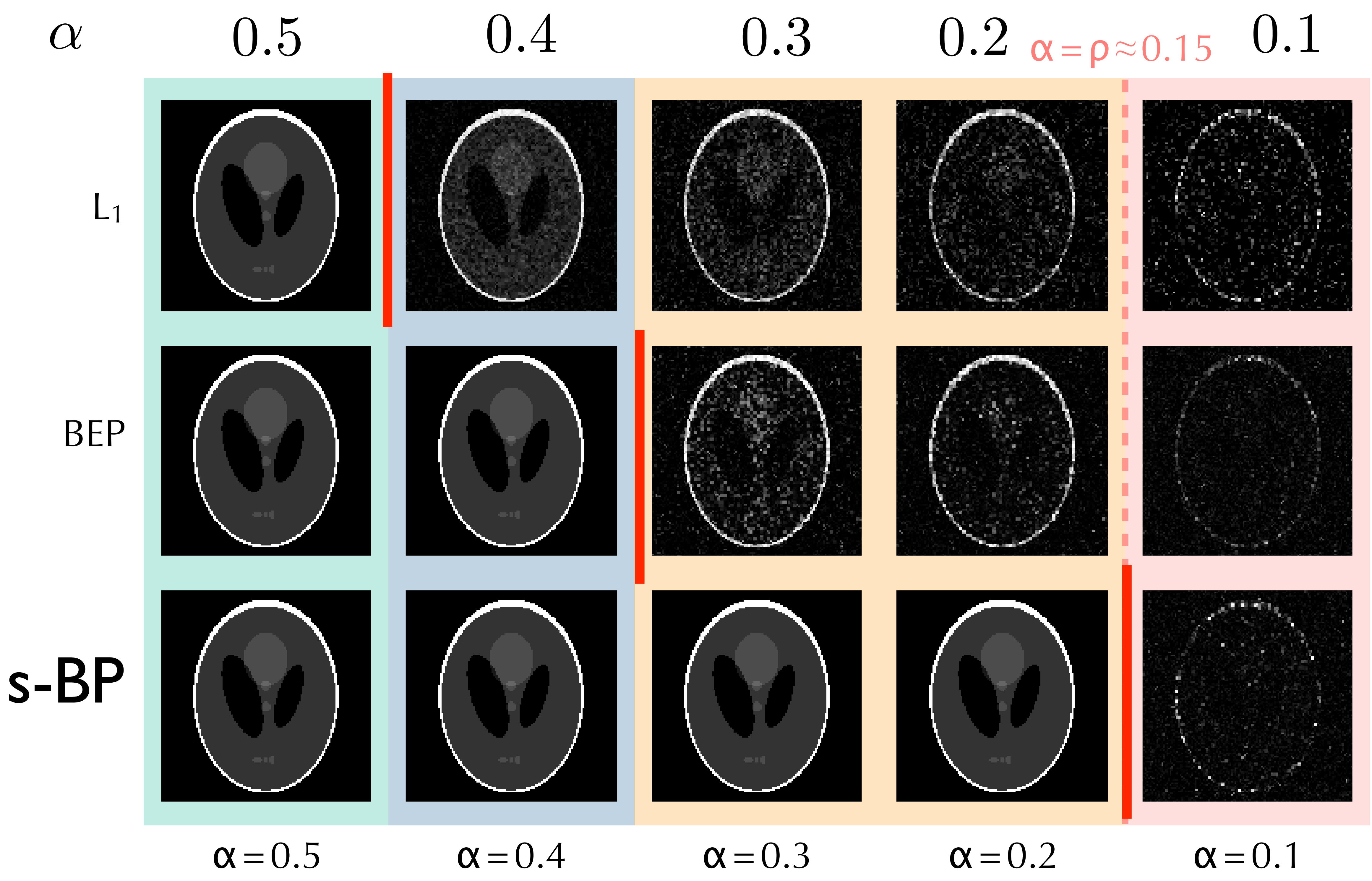
More realistic signals?



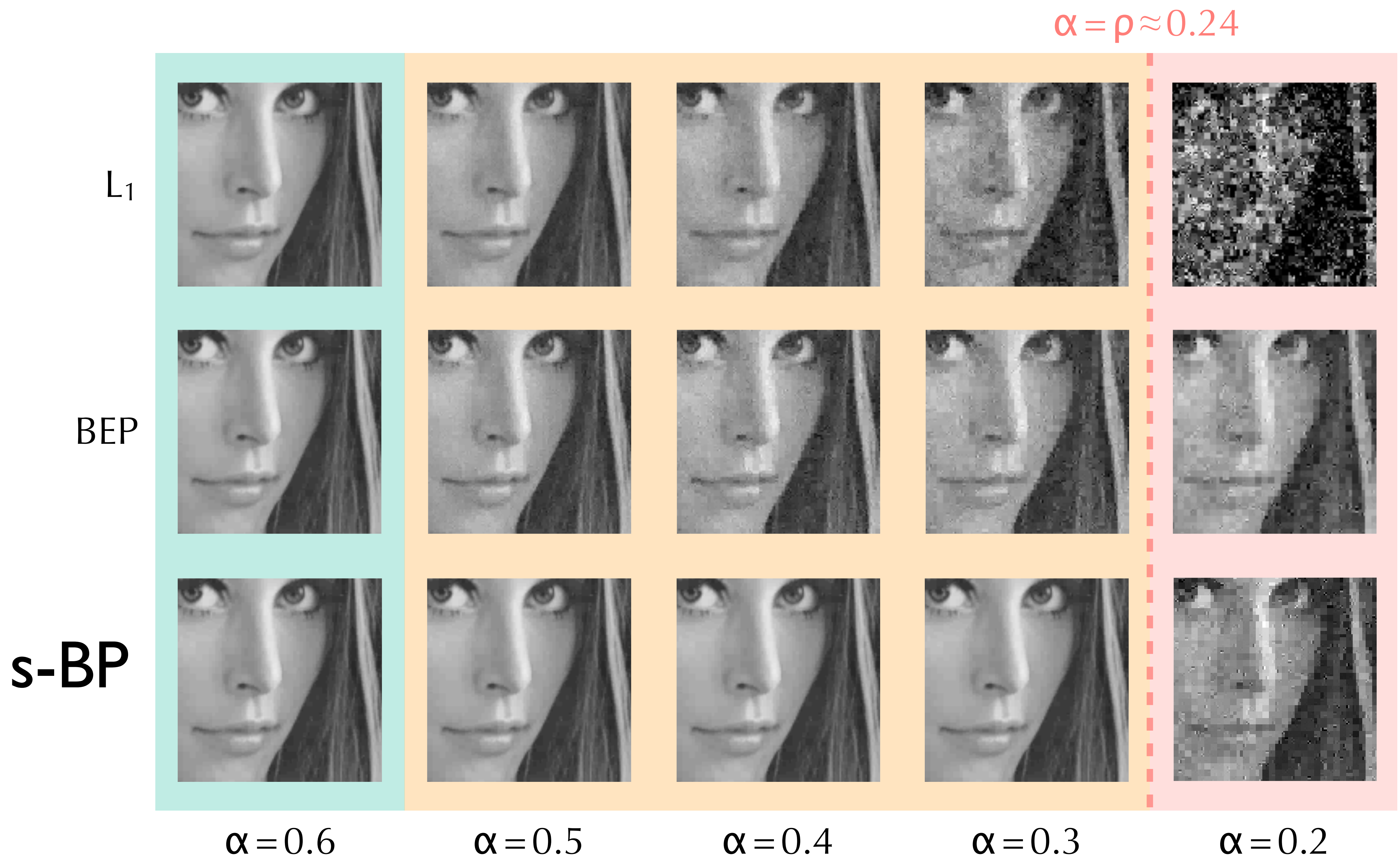
Shepp-Logan phantom, in the Haar-wavelet representation



Shepp-Logan phantom, in the Haar-wavelet representation

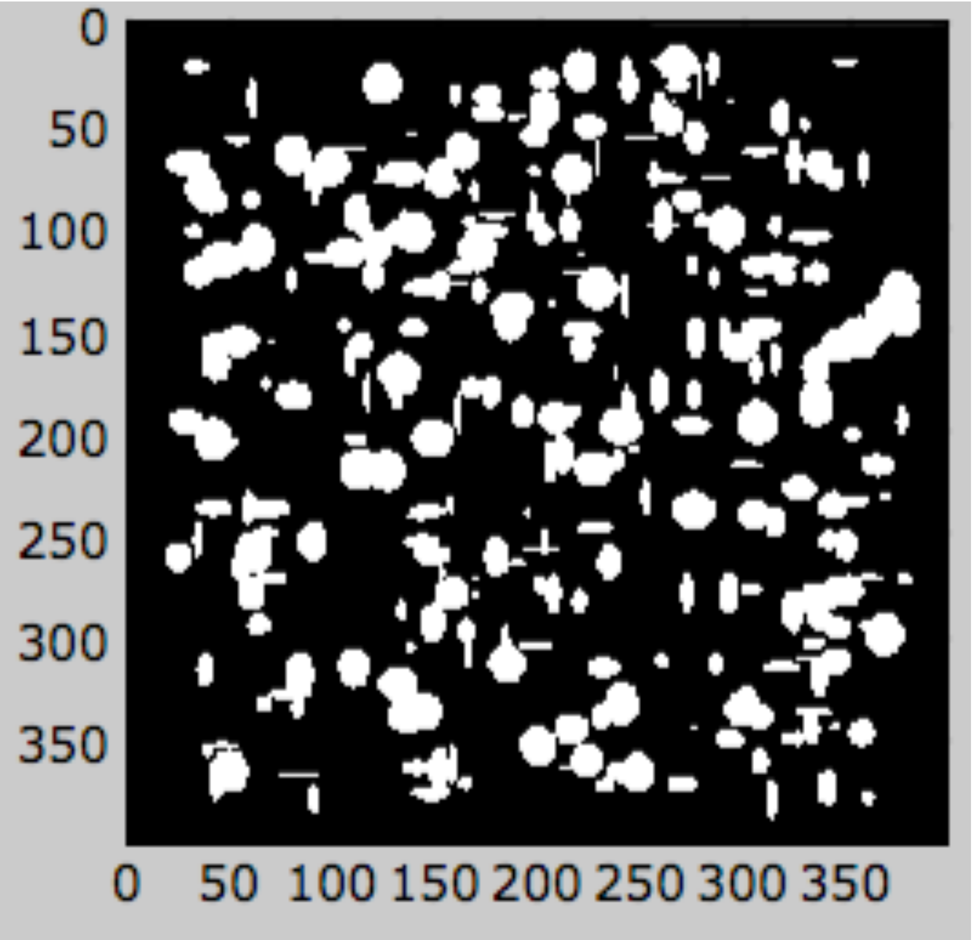


Shepp-Logan phantom, in the Haar-wavelet representation



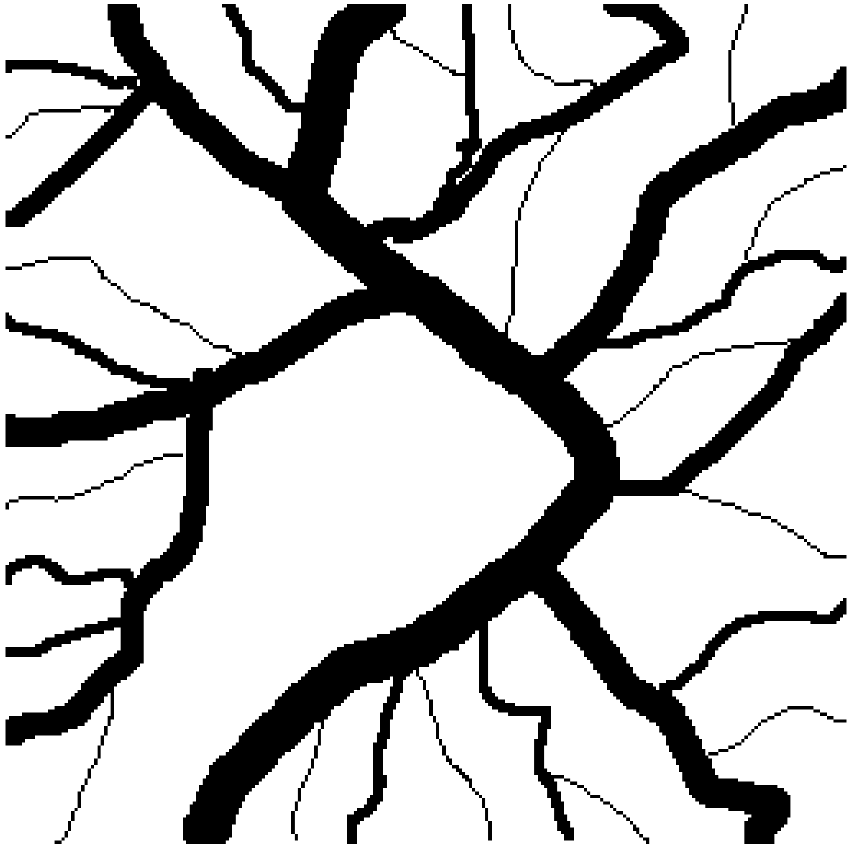
α = ρ ≈ 0.24
L1
BEP
s-BP
α = 0.6
α = 0.5
α = 0.4
α = 0.3
α = 0.2

Discrete tomography



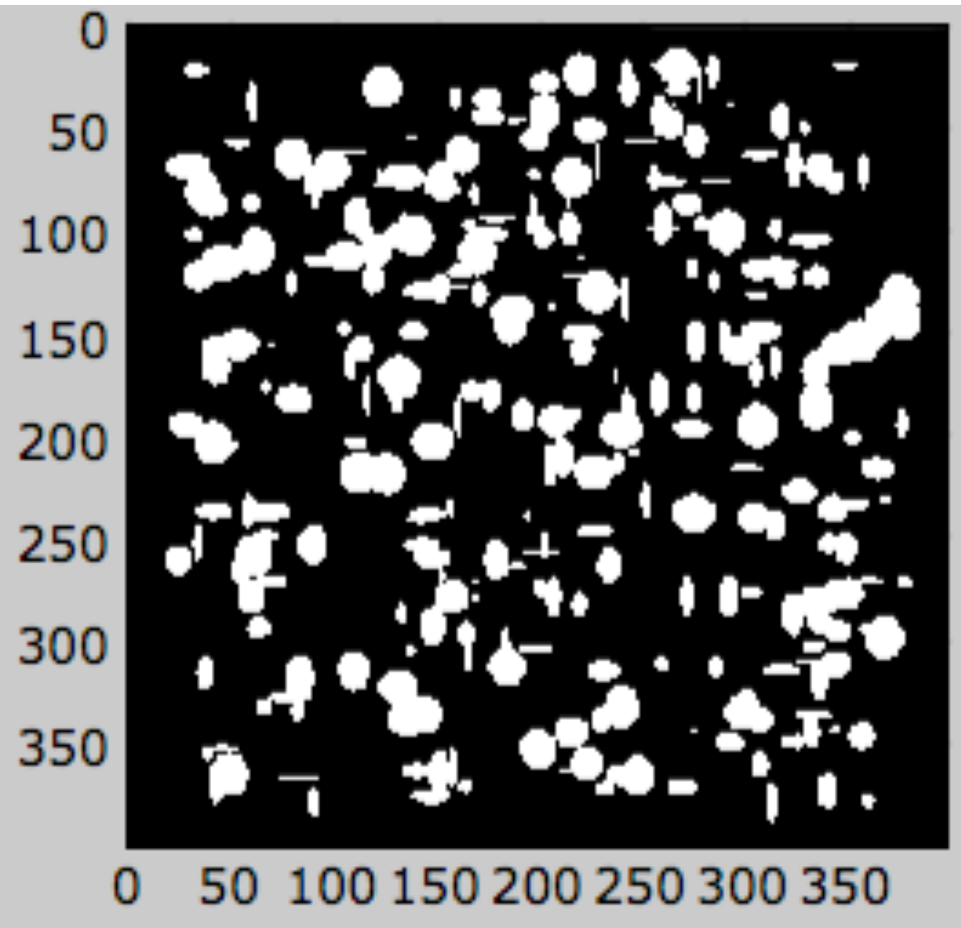
30 angles

14 angles



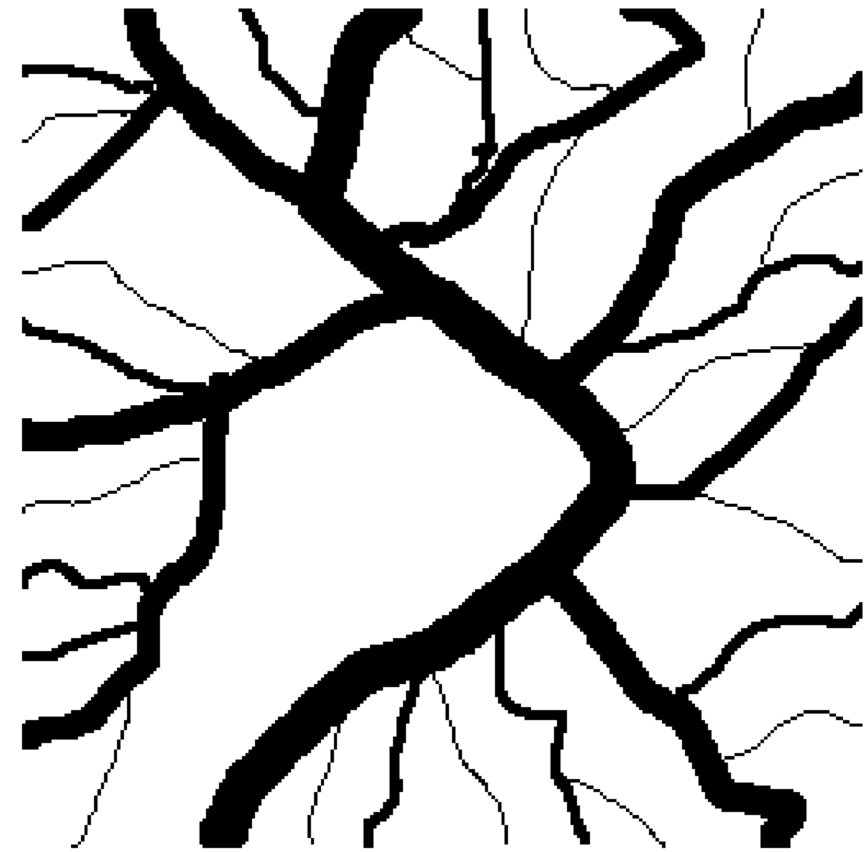
17 angles

Discrete tomography



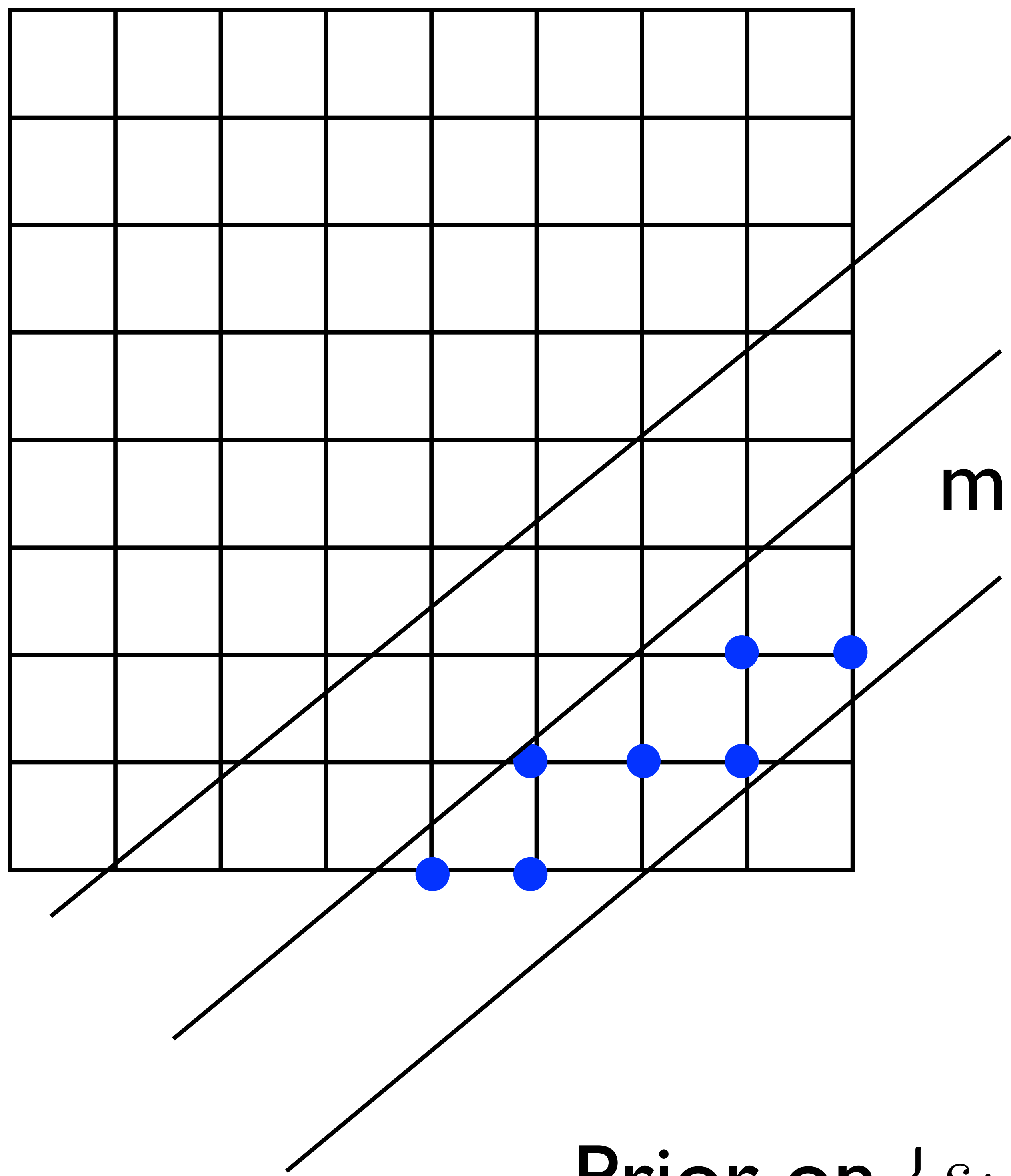
30 angles

14 angles



17 angles

Images = large size, but structured data.
Use hint that it is an image!



measurement μ

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

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

Probabilistic approach

$$P(\{\vec{S}\}) \propto \prod_{\mu=1}^{M} \delta(y_\mu - \sum_{i\in\mu} S_i) \prod_{\mu=1}^{M} e^{J\sum_{i\in\mu} S_i S_{i+1}}$$

Solution of the linear system

Prior on the images

Belief propagation applied to this problem : allows to handle large-size problems

Robust to noise!

Adding a noise to the projections

From 6 angles...

Original

BP

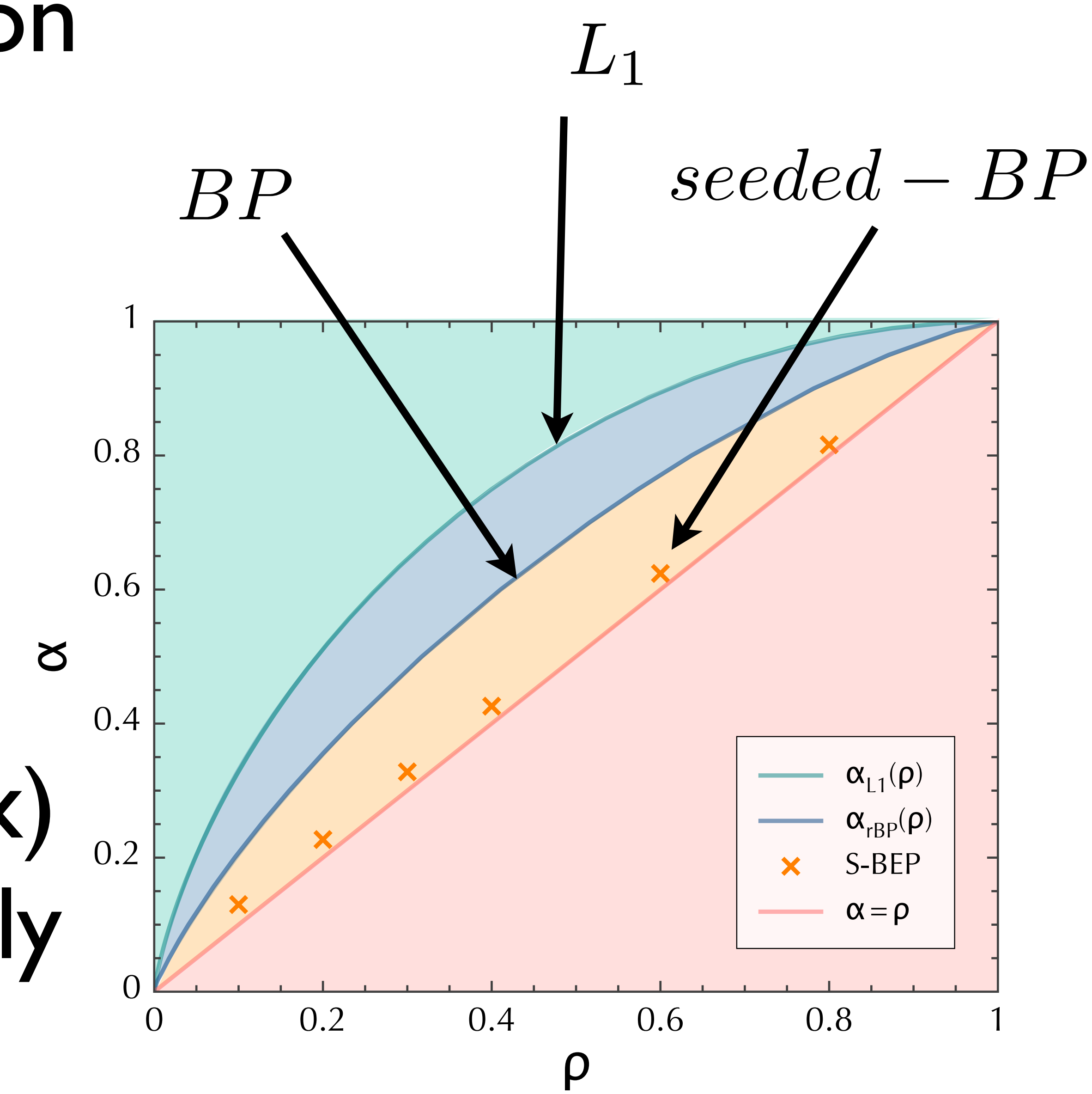
Continuous
+Total Variation

(i.e. LASSO-type problem)

Summary

Progress based on the union of three ingredients:

- Probabilistic approach
- Message passing reconstruction of the signal
- Careful design of the measurement matrix to avoid glass transition



- Robust to noise (signal, matrix)
- Generalizable to approximately sparse signals
- Applications...

Based on joint work with

Jean Barbier (ESPCI), Emmanuelle Gouillart (Saint-Gobain-CNRS), **Florent Krzakala**(ENS), François Sausset (LPTMS), Yifan Sun (ESPCI), **Lenka Zdeborova**(IPhT)

- Phys. Rev. X 2, 021005, (2012) (open access)
- J. Stat. Mech. (2013) P01008
- Inverse Problems 29, 3 (2013) 035003
- arXiv: 1301.5898
- arXiv: 1301.0901
- arXiv: 1207.2079