Statistical physics approach to compressed sensing

Marc Mézard
LPTMS, Université Paris Sud, CNRS

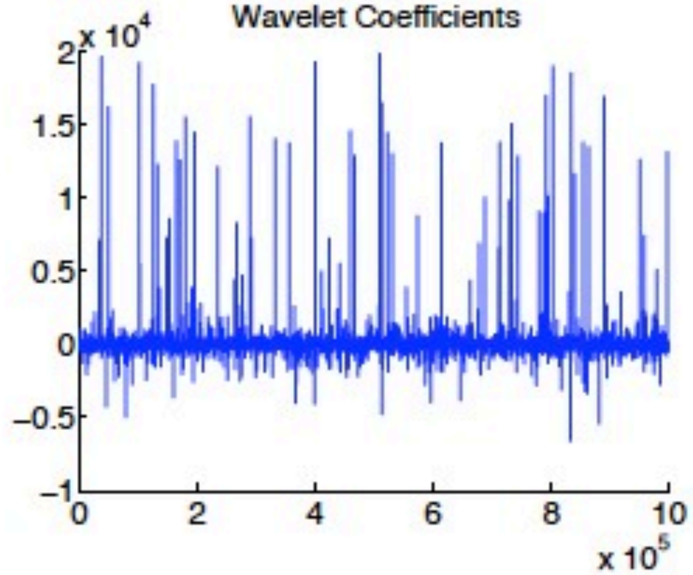
collaboration with

Florent Krzakala(ESPCI), **François Sausset** (LPTMS),**Yifan Sun** (ESPCI), **Lenka Zdeborova**(IPhT)

arXiv:1109.4424

Phys. Rev. X 2, 021005, (2012) (open access))

Sparse signals

From 65.536 wavelet coefficients, keep 25.000

(From Candes-Wakin)

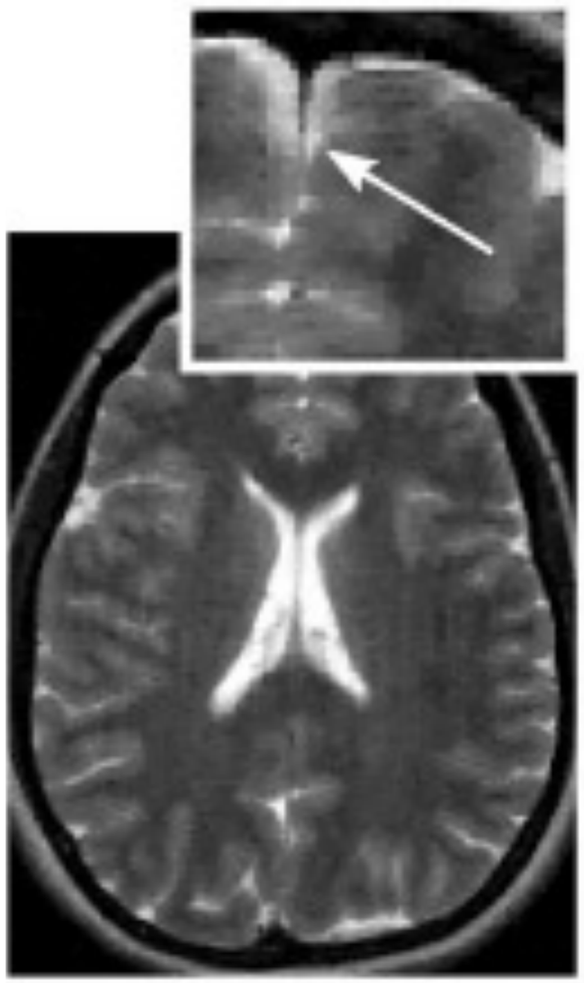
Exploited for data compression (JPEG). More recently: data acquisition (...,Donoho, Candes-Romberg-Tao, 2006,+...)

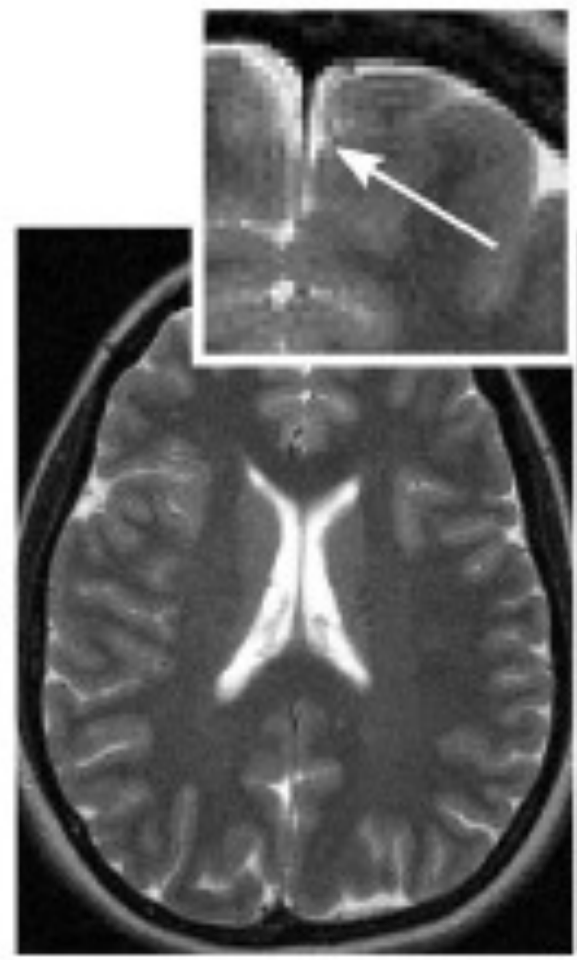
Compressed sensing

Acquire N bit data by doing measurements on much less than N bits (possible if signal is compressible, i.e. it has much less than N bits of information).

Possible applications:
- Rapid Magnetic Resonance Imaging
- Tomography, microscopy
- Image acquisition (single-pixel camera)
- Infer regulatory interactions among many genes using only a limited number of experimental conditions
- Possible relevance in information processing in the brain (e.g. uncover original signal from compressed signal sent by retina
- ...

An example from magnetic resonance imaging





Left: image acquired with compressed sensing: acceleration 2.5

Lustig et al.,

The simplest problem: getting a signal from some measurement= linear transforms

Consider a system of linear measurements

$$y = Fx$$

Measurements

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

(e.g. wavelet components)

Signal $x = \begin{pmatrix} x^1 \\ . \\ . \\ . \\ . \\ x^N \end{pmatrix}$

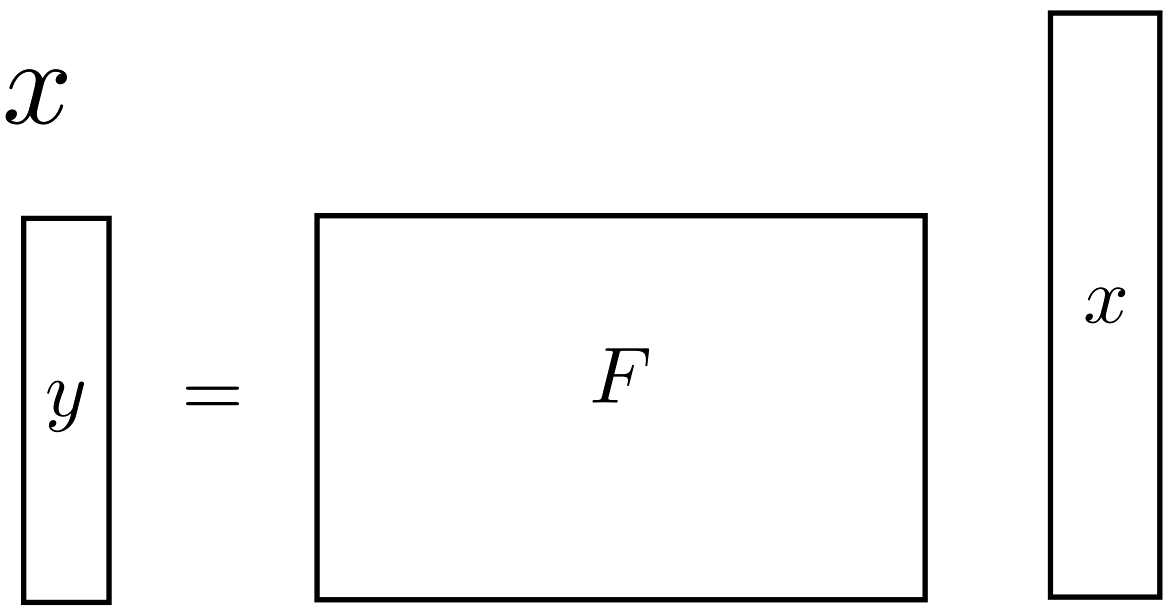
$F = M \times N$ matrix

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

The problem: $y = Fs$ and x is sparse, i.e. it has R components $\neq 0$

$R < M < N$ y is observed, F is known. Find s

Study the linear system $y = Fx$



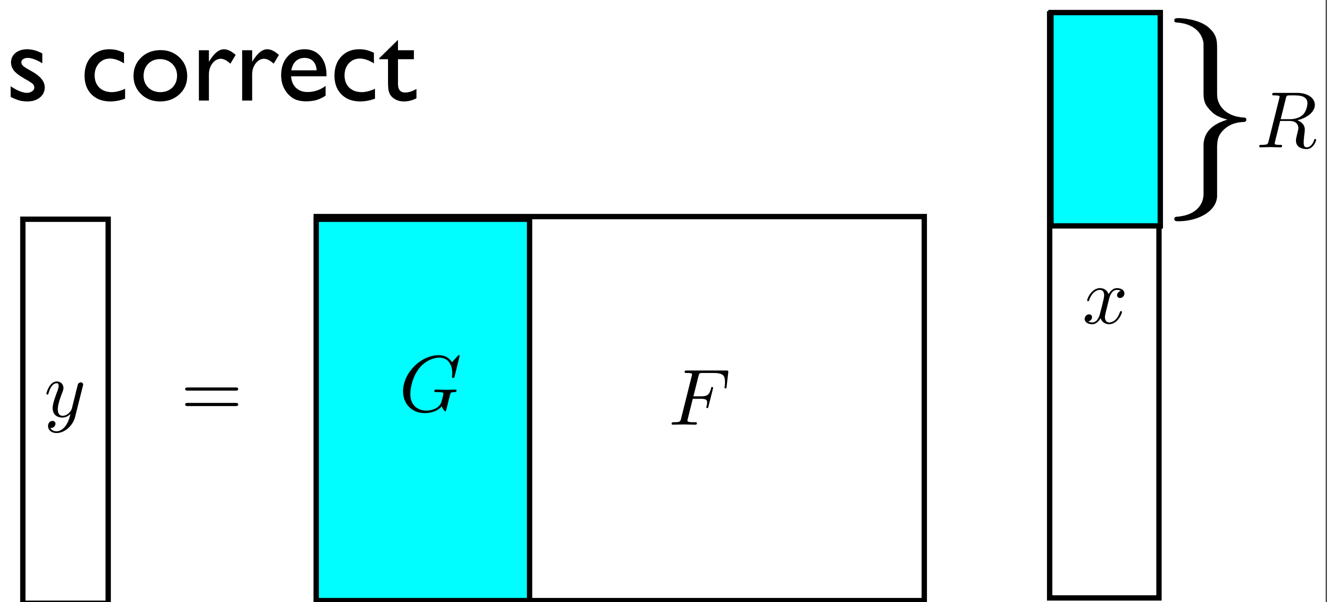
Exploit the sparsity of the original s

The problem: $y = Fs$ and s is sparse

R components $\neq 0$

⇒ Study the linear system $y = Fx$

A 'simple' solution: guess the positions where $x_i \neq 0$ and check if it is correct



e.g. $x_1, \ldots, x_R \neq 0$

$G = \{\ R$ first columns of $F\ \}$

Solve : $y^\mu = \sum_{i=1}^{R} G^{\mu i} x_i \qquad \mu = 1, \ldots, M$

$R < M$ ⇒ too many equations

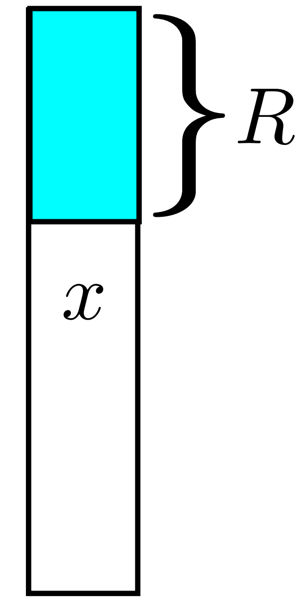
⇒ generically inconsistent (no solution), except if the guess of locations of $x_i \neq 0$ was correct

The problem: $y = Fs$ and s is sparse

R components $\neq 0$

➡ Study the linear system $y = Fx$

A 'simple' solution: guess the positions where $x_i \neq 0$ and check if it is correct



e.g.

$G = \{$

Solve

$\binom{N}{R}$ possible guesses

$R < M$

➡ generically inconsistent (no solution), except if the guess of locations of $x_i \neq 0$ was correct

Compressed sensing as an optimization problem: the L_1 norm approach

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

Hopefully: $x = s$

$||x||_0$: number of non-zero components

$$||x||_p = \sum_i |x_i|^p$$

Ideally, use $||x||_0$. In practice, use $||x||_1$

Compressed sensing as an optimization problem: the L_1 norm approach

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

Worst-case analysis: How many equations are needed in order to get the correct result for any initial sparse signal? Candès-Tao, Donoho

Typical-case analysis: How many equations are needed in order to get the correct result for almost all initial sparse signals and measurement matrices, drawn from some measure (e.g. $F_{\mu i}$ = iid Gaussian variables)

Phase diagram of the L_1 norm approach

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

Hardest and most interesting regime:

$N \gg 1$ variables

$R = \rho N$ non-zero variables

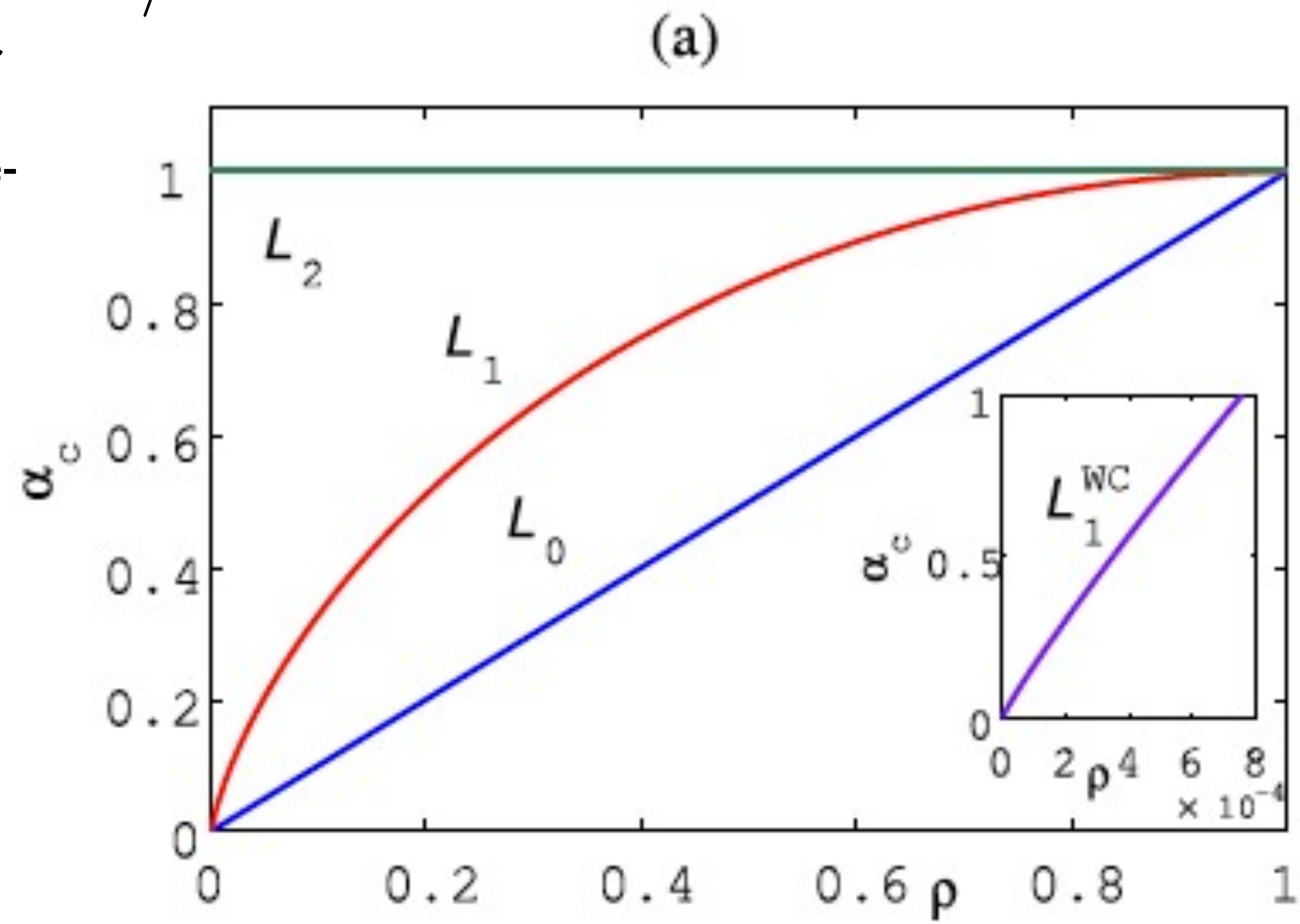
$M = \alpha N$ equations

Typical-case analysis: phase diagram in the plane ρ, α

Phase diagram

$\alpha = M/N$

Number of measure-ments per variable



Donoho 2006, Donoho Tanner 2005

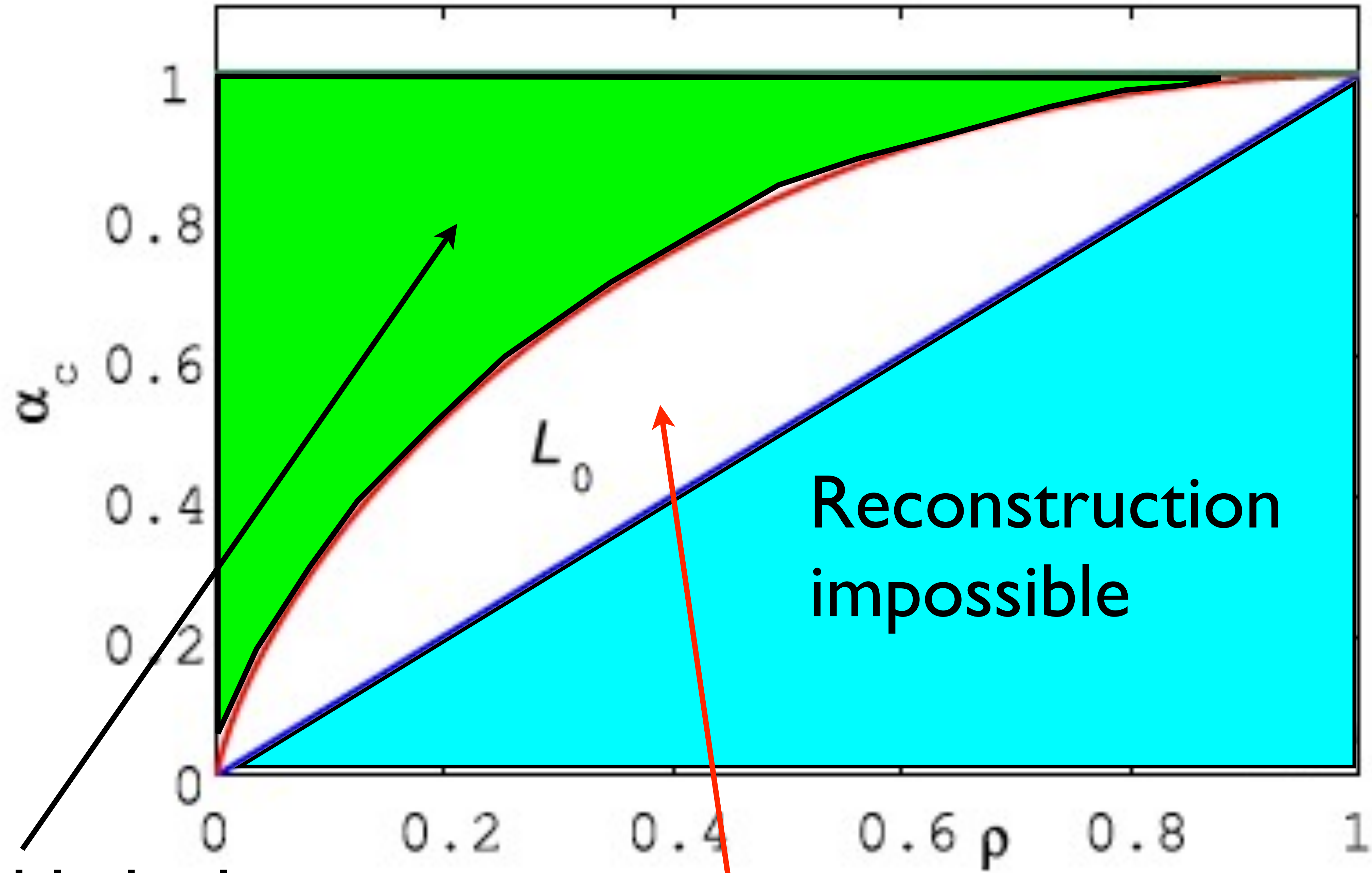
Kabashima, Wadayama and Tanaka, JSTAT 2009

$\rho = R/N$

Fraction of non-zero variables

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

Gaussian random matrix



Possible by linear programming
Efficient message passing solution

Donoho Maleki Montanari; (Kabashima MM)

Possible by enumeration, using a time $O(e^N)$

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

Step 1:Probabilistic approach to compressed sensing

Signal generated from: $P_0(\mathbf{s}) = \prod_{i=1}^{N} [(1-\rho_0)\delta(s_i) + \rho_0 \phi_0(s_i)]$

Probabilistic decoding using:

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

NB: $(\rho, \phi(x))$ may be *distinct* from true signal distribution $(\rho_0, \phi_0(x))$: no need of prior knowledge of signal

Theorem: if $\rho_0 < 1$, $\rho < 1$, $\alpha > \rho_0$, F random Gaussian, in the large N limit the maximum of $P(\mathbf{x})$ is at $\mathbf{x} = \mathbf{s}$

⇒ Sampling from $P(\mathbf{x})$ is optimal, even if we do not know the correct ρ_0 , ϕ_0

Not intuitive!

Hint of why the theorem is correct

Theorem: if $\rho_0 < 1$, $\rho < 1$, $\alpha > \rho_0$, F random Gaussian, in the large N limit the maximum of $P(\mathbf{x})$ is at $\mathbf{x} = \mathbf{s}$

$$Z(D) \equiv \lim_{\Delta \to 0} \int \prod_{i=1}^{N} \mathrm{d}x_i \, \mathbb{I}\left(\sum_{i=1}^{N} (x_i - s_i)^2 / N = D \right)$$
$$\prod_{i=1}^{N} [(1-\rho)\delta(x_i) + \rho\phi(x_i)] \prod_{\mu=1}^{M} \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{1}{2\Delta}[\sum_{i=1}^{N} F_{\mu i}(x_i - s_i)]^2}$$

In the limit $\Delta \to 0$ when $D \to 0$ (i.e. $x = s$) there is a factor $\delta(0)^{(1-\rho_0+\alpha)N}$ and N integrals over x_i

Therefore $\lim_{D \to 0} Z(D) = \infty$ when $\alpha > \rho_0$. (Proof...)

while $Z(D>0)$ is finite. (Proof by first moment (annealed) bound)

Step 2: belief propagation-based reconstruction with parameter learning

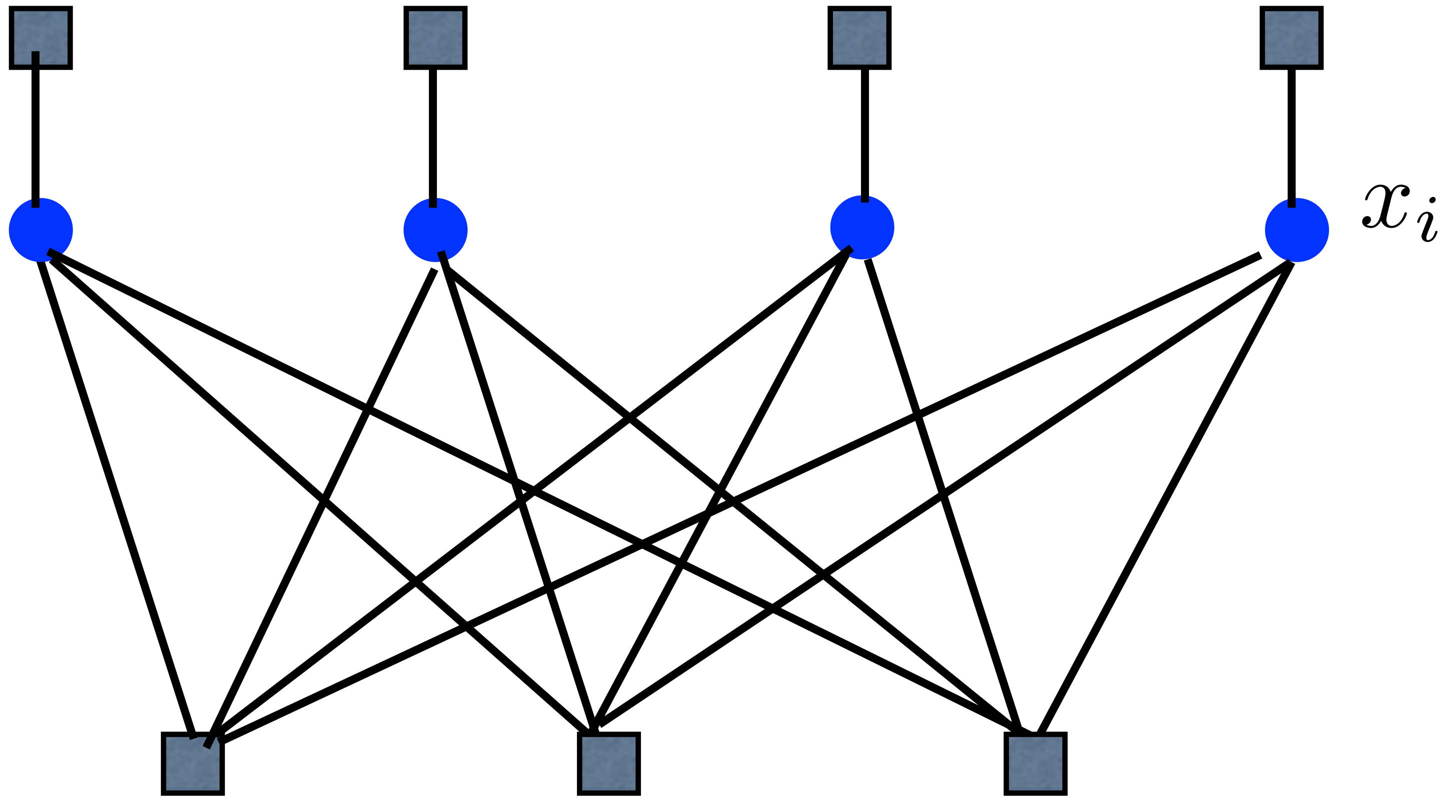
$$P(\mathbf{x}) = \prod_{i=1}^{N}[(1-\rho)\delta(x_i) + \rho\phi(x_i)] \quad \prod_{\mu=1}^{P} \delta\left(y_\mu - \sum_i F_{\mu i} x_i\right)$$ Gaussian ϕ

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

Use **belief propagation**, with **gaussian-approximated** messages, and **parameter learning** of (ρ, ϕ) .

Belief propagation = replica symmetric cavity method on a given instance

«spins»

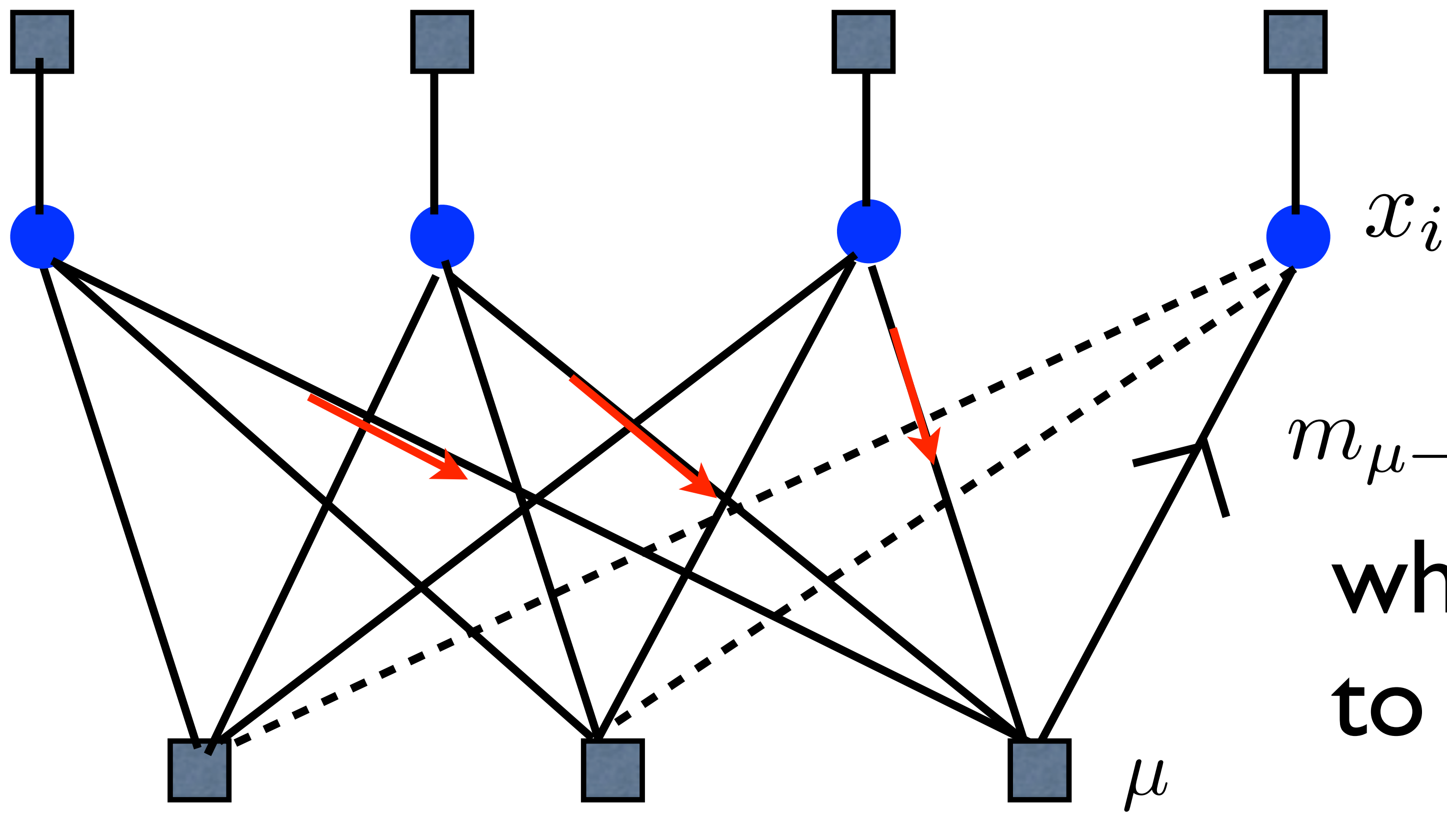


«Factor graph»

constraints

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

«spins»



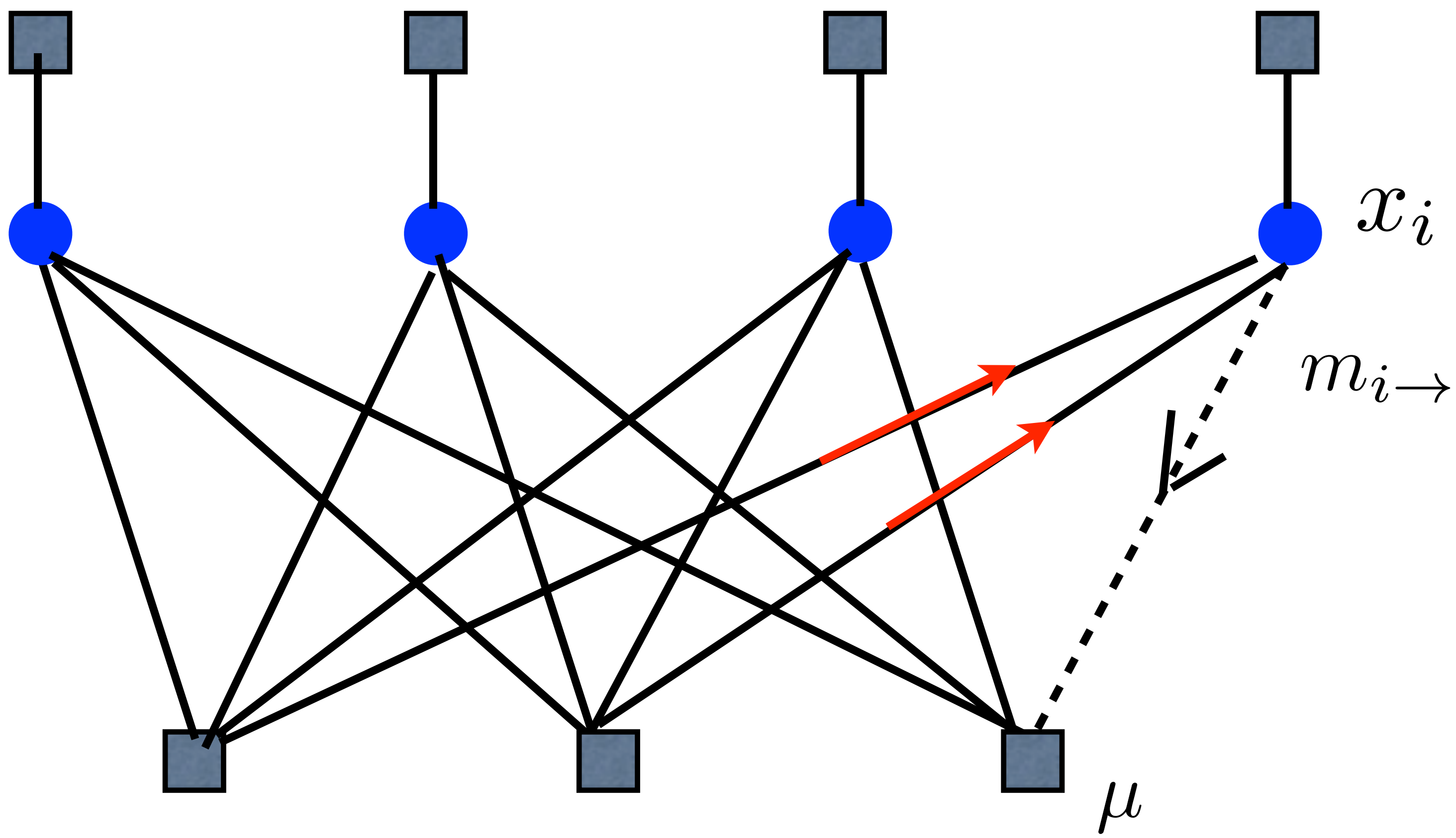
$m_{\mu\to i}(x_i)$: distribution of x_i when it is connected only to μ

constraints = interactions

$$m_{\mu\to i}(x_i) = \frac{1}{Z^{\mu\to i}} \int \prod_{j\neq i} \mathrm{d}x_j \prod_{j\neq i} m_{j\to\mu}(x_j)\, \delta\left(\sum_j F_{\mu j} x_j - y_\mu\right)$$

$m_{j\to\mu}(x_j)$: distribution of x_j when μ is absent

Message passing for compressed sensing



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

when it is disconnected from μ

Belief Propagation equations

$$m_{i\to\mu}(x_i) = \frac{1}{Z^{i\to\mu}} \left[(1-\rho)\delta(x_i) + \rho\phi(x_i)\right] \prod_{\gamma\neq\mu} m_{\gamma\to i}(x_i)$$

$$m_{\mu\to i}(x_i) = \frac{1}{Z^{\mu\to i}} \int \prod_{j\neq i} \mathrm{d}x_j \prod_{j\neq i} m_{j\to\mu}(x_j)\, \delta\left(\sum_j F_{\mu j} x_j - y_\mu\right)$$

NB: Validity of BP equations?

- Fully connected system, «couplings» $F_{\mu i} = O(1/\sqrt{N})$

  SK like. small correlations within one single pure state

- If the prior matches the original signal distribution ($\rho = \rho_0,\ \phi(x) = \phi_0(x)$), then a Nishimori condition guarantees that there is a single pure state $\longrightarrow$ BP OK

$$m_{i\to\mu}(x_i) = \frac{1}{Z^{i\to\mu}} \left[(1-\rho)\delta(x_i) + \rho\phi(x_i)\right] \prod_{\gamma\neq\mu} m_{\gamma\to i}(x_i)$$

$$m_{\mu\to i}(x_i) = \frac{1}{Z^{\mu\to i}} \int \prod_{j\neq i} \mathrm{d}x_j \prod_{j\neq i} m_{j\to\mu}(x_j)\ \delta\left(\sum_j F_{\mu j} x_j - y_\mu\right)$$

Message passing for compressed sensing

Large connectivity: simplification by projection on first two moments (cf SK model).

One message = two numbers (mean, variance)

Belief Propagation equations

$$m_{i\to\mu}(x_i) = \frac{1}{Z^{i\to\mu}} \left[(1-\rho)\delta(x_i) + \rho\phi(x_i)\right] \prod_{\gamma\neq\mu} m_{\gamma\to i}(x_i)$$

$$m_{\mu\to i}(x_i) = \frac{1}{Z^{\mu\to i}} \int \prod_{j\neq i} \mathrm{d}x_j \prod_{j\neq i} m_{j\to\mu}(x_j)\, \delta\left(\sum_j F_{\mu j} x_j - y_\mu\right)$$

Gaussian-projected BP («relaxed-BP»)

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

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

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

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

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

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

Parameter learning

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

Parameters: $\rho, \ \overline{x}, \ \sigma$

(taking Gaussian $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\overline{x})^2)/(2\sigma^2)}$)

Express the Bethe free-entropy $\log Z$ in terms of the BP messages.

Update the parameters $\rho, \ \overline{x}, \ \sigma$ at each iteration by moving in the direction of the gradient of $\log Z$

→ Find the parameters which maximize Z

Performance of the probabilistic approach + message passing + parameter learning

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

$F_{\mu i}$ iid Gaussian, variance $1/N$

- Simulations
- Analytic study of the large N limit

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

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

Quenched disorder:

$F_{\mu i}$ iid Gaussian, variance $1/N$

$y_\mu = \sum_{i=1}^{N} F_{\mu i} x_i^0$ where x_i^0 are iid distributed from

$$(1-\rho_0)\delta(x_i^0) + \rho_0 \phi_0(x_i)$$

Infinite range weak interactions...

Replica computation:

$$E(\log Z) = \lim_{n\to 0} \frac{E(Z^n) - 1}{n}$$

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

$$E(Z^n) = \max_{D,V} \; e^{Nn\phi(D,V)}$$

Φ is known

Order parameters:

$$D = \frac{1}{N}\sum_i (\langle x_i \rangle - s_i)^2 \qquad V = \frac{1}{N}\sum_i \left(\langle x_i^2 \rangle - \langle x_i^2 \rangle\right)$$

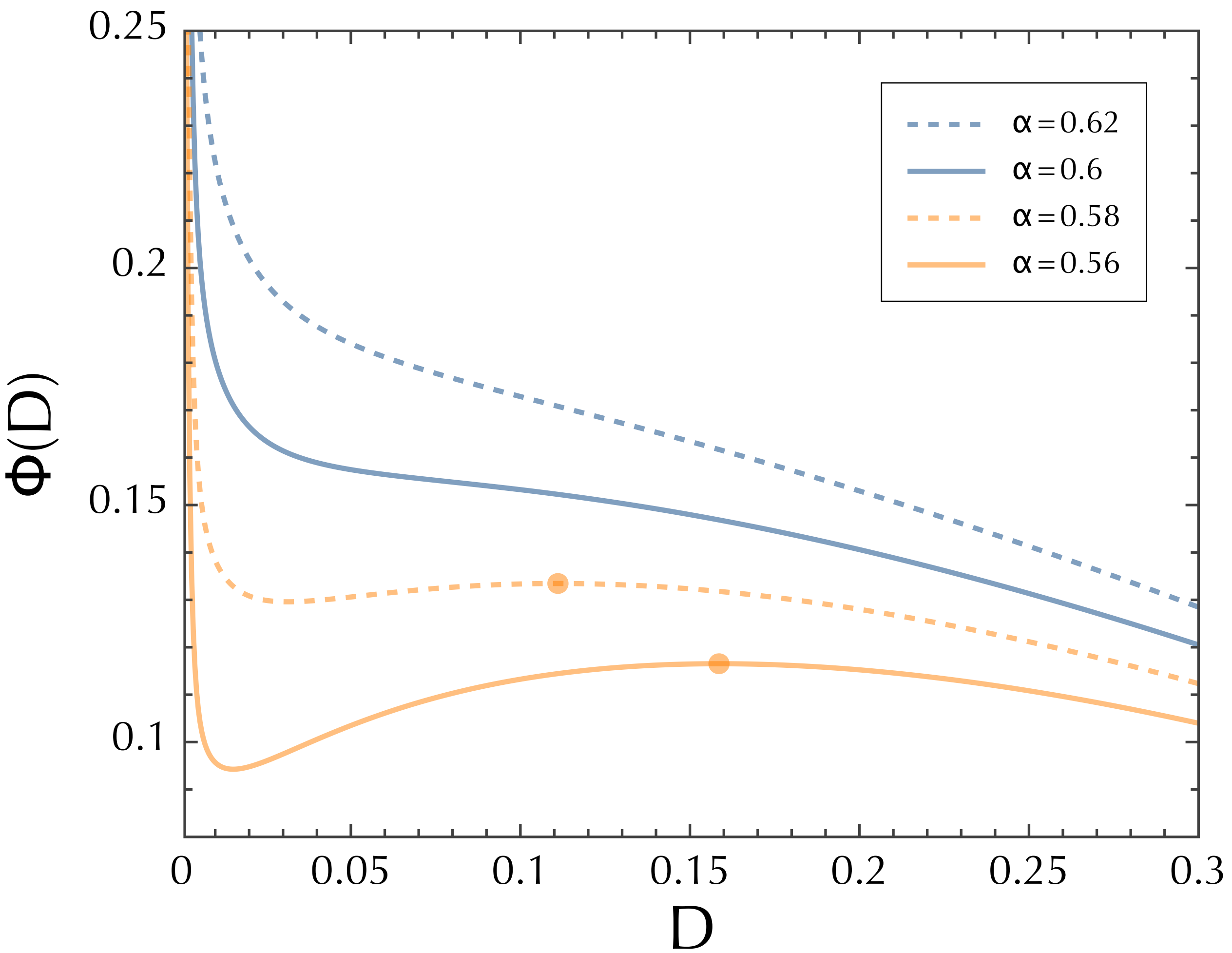
Cavity approach shows that the order parameters of the BP iteration flow according to the gradient of the replica free entropy Φ

NB: Replica symmetric expression of Φ is OK only on the Nishimori line: $\rho = \rho_0 \quad \phi = \phi_0$

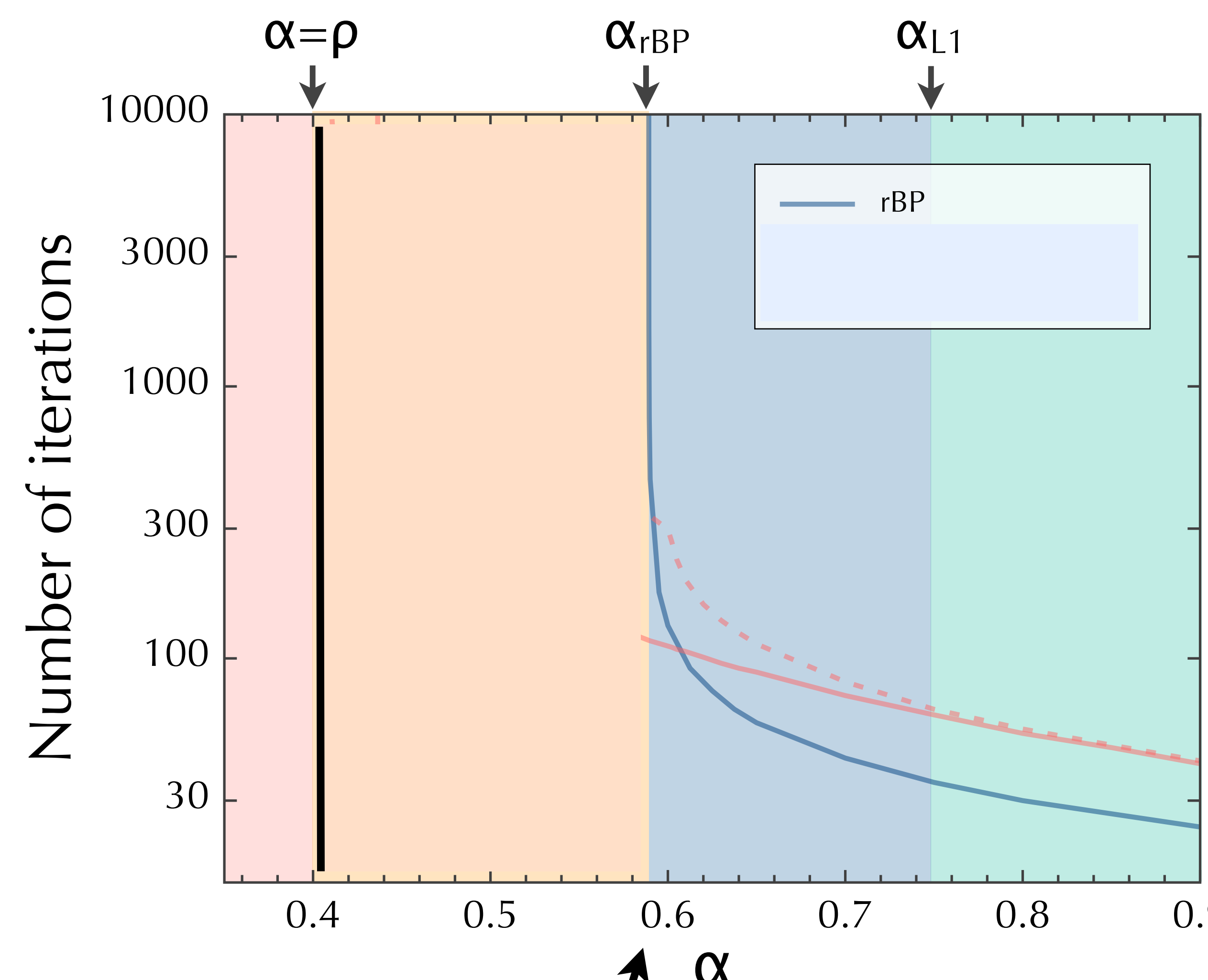
Free entropy

$\rho_0 = .4$

BP convergence time



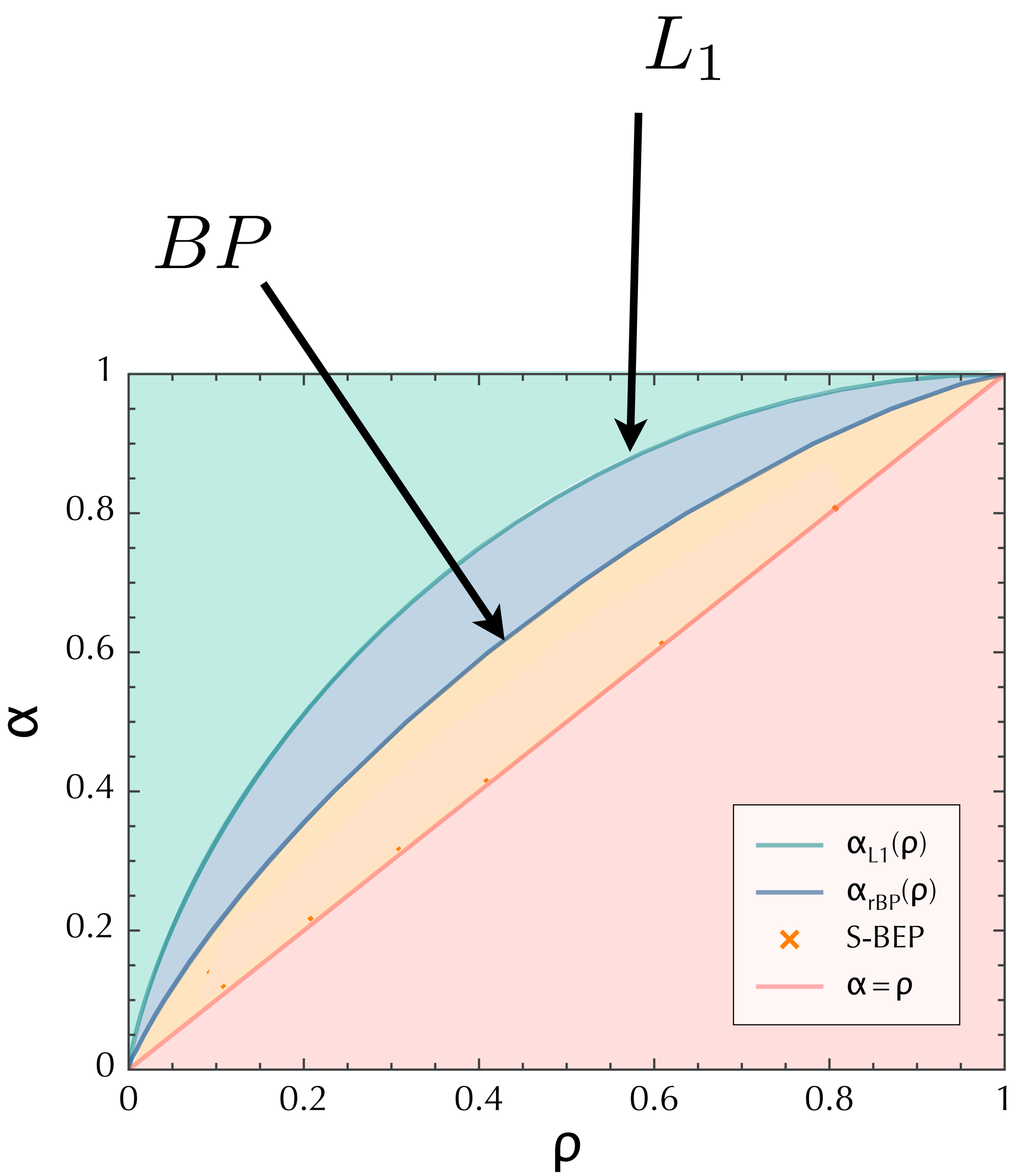
distance to native state



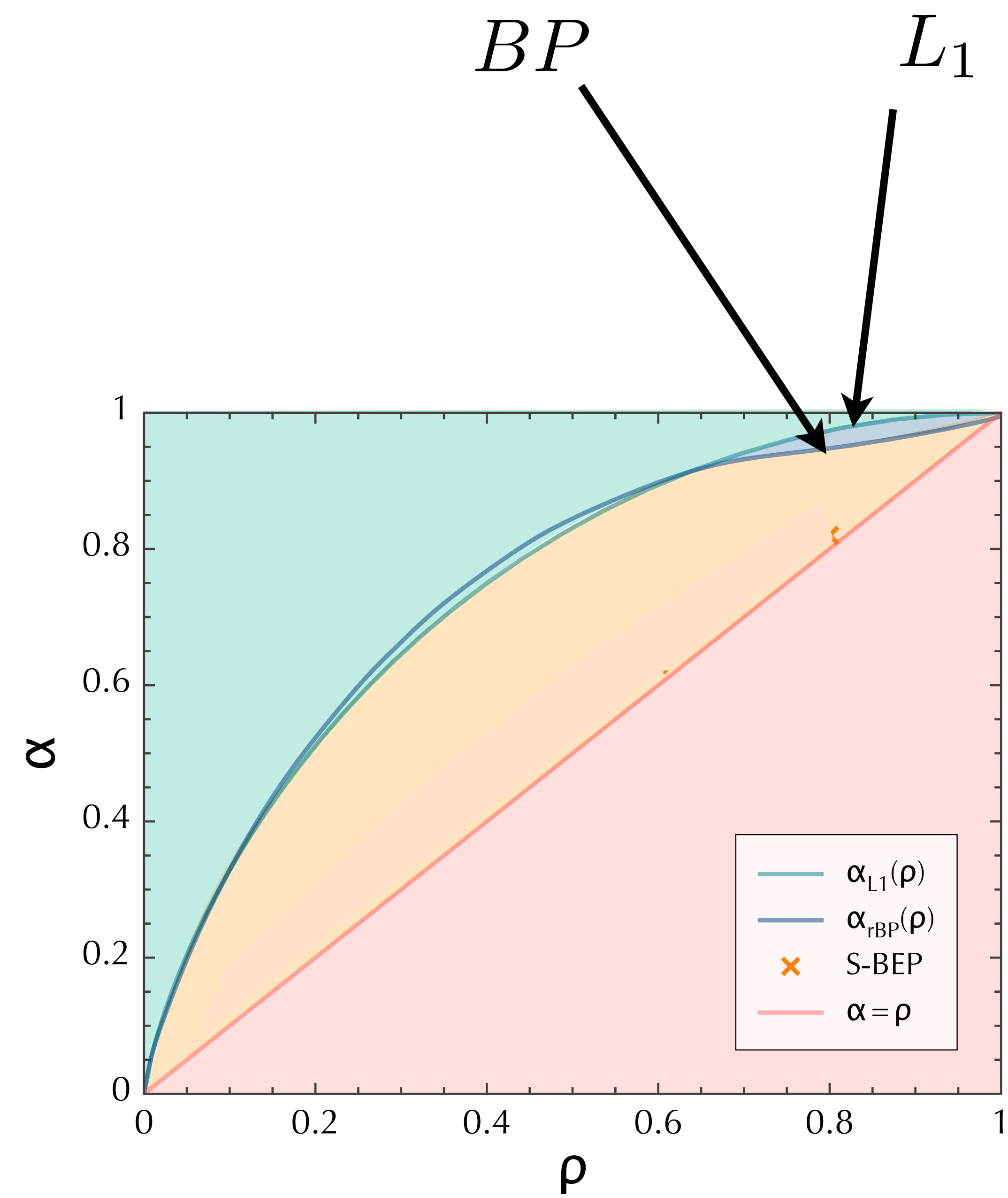
Dynamic glass transition

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

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!)

Hassani Macris Urbanke

Nucleation and seeding

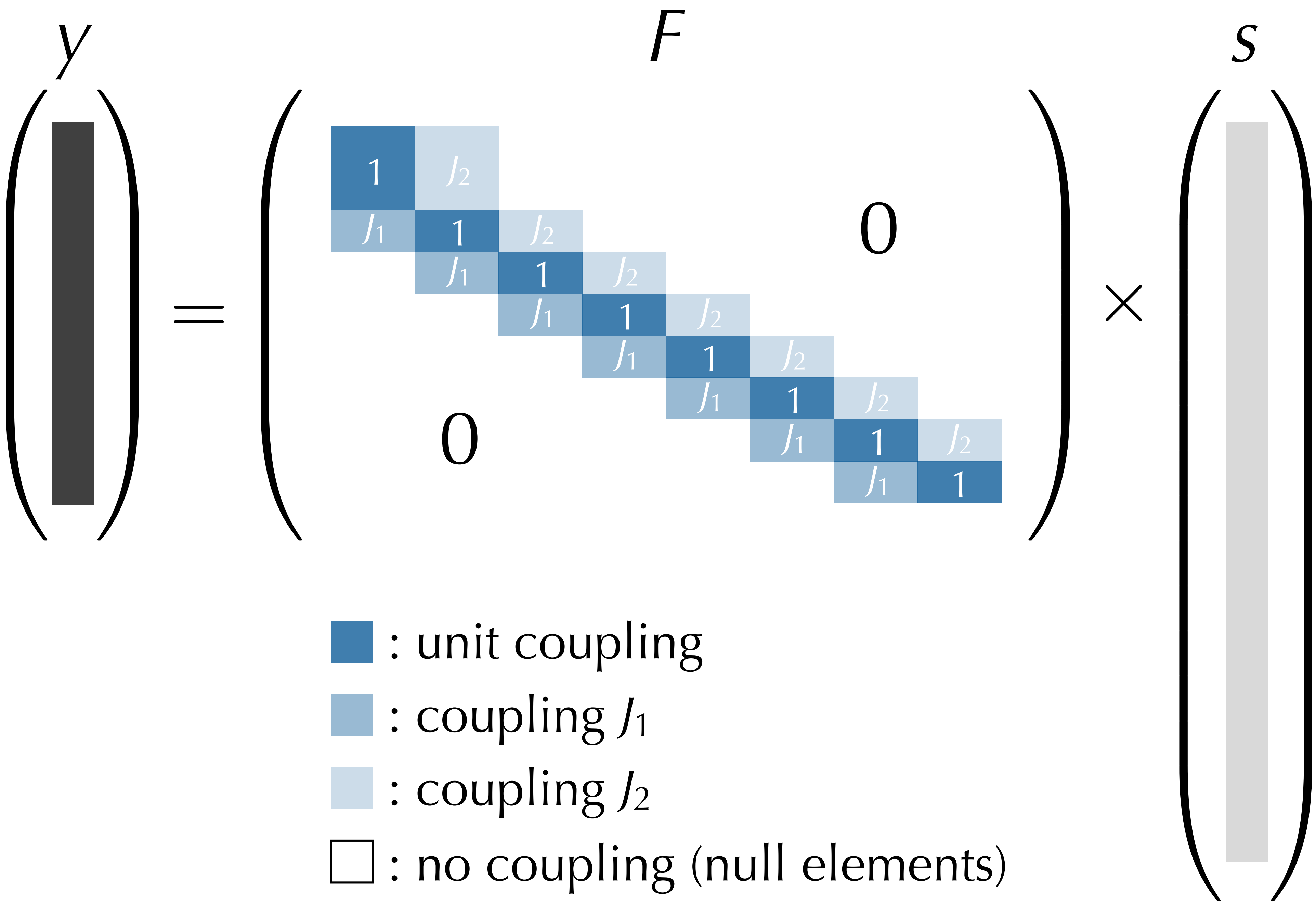
Step 3: design the measurement matrix in order to get around the glass transition

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

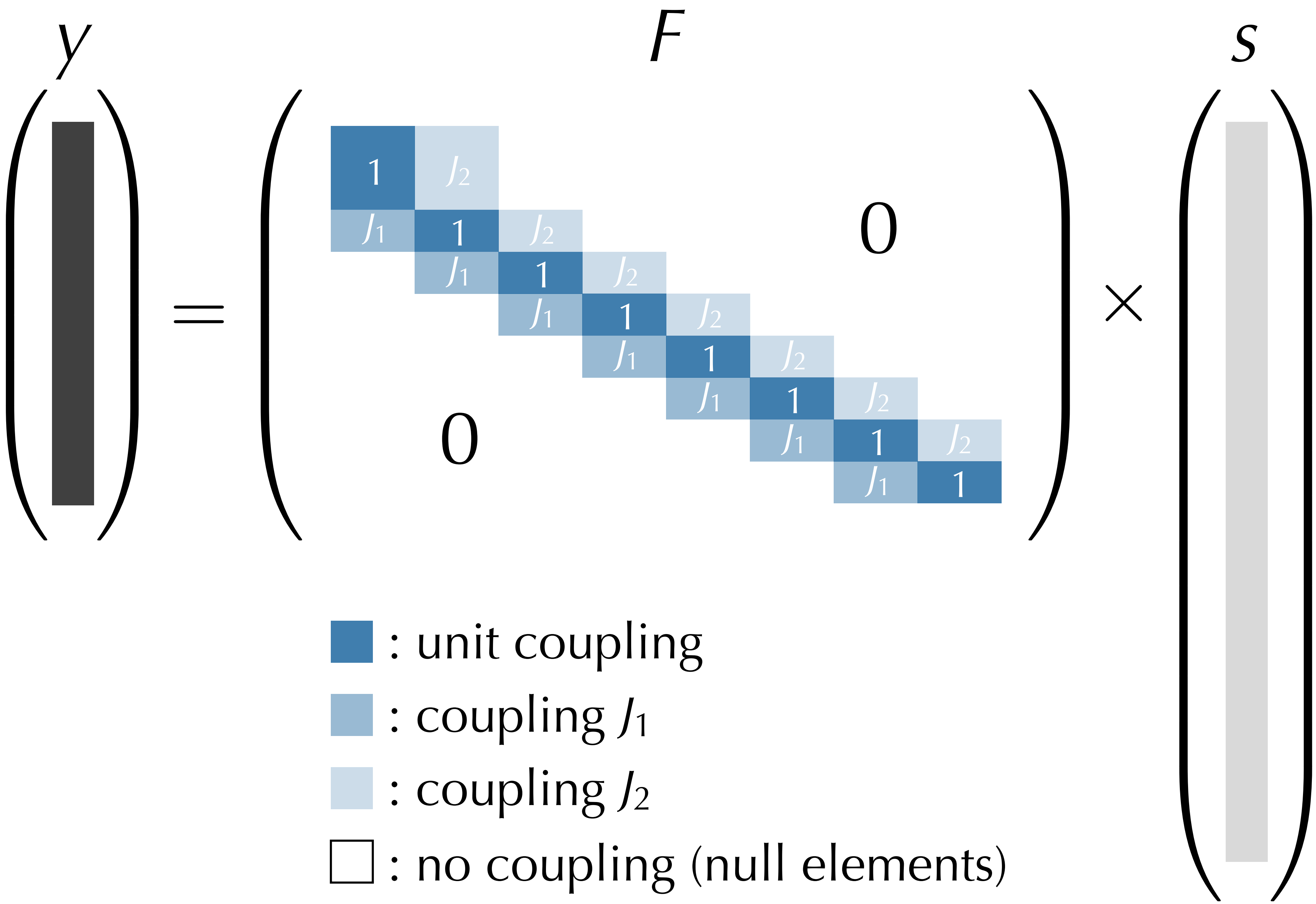
⟶ Seeded BP

Group the variables and the measurements into L blocks

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



Structured measurement matrix.
Variances of the matrix elements



$$L = 8$$

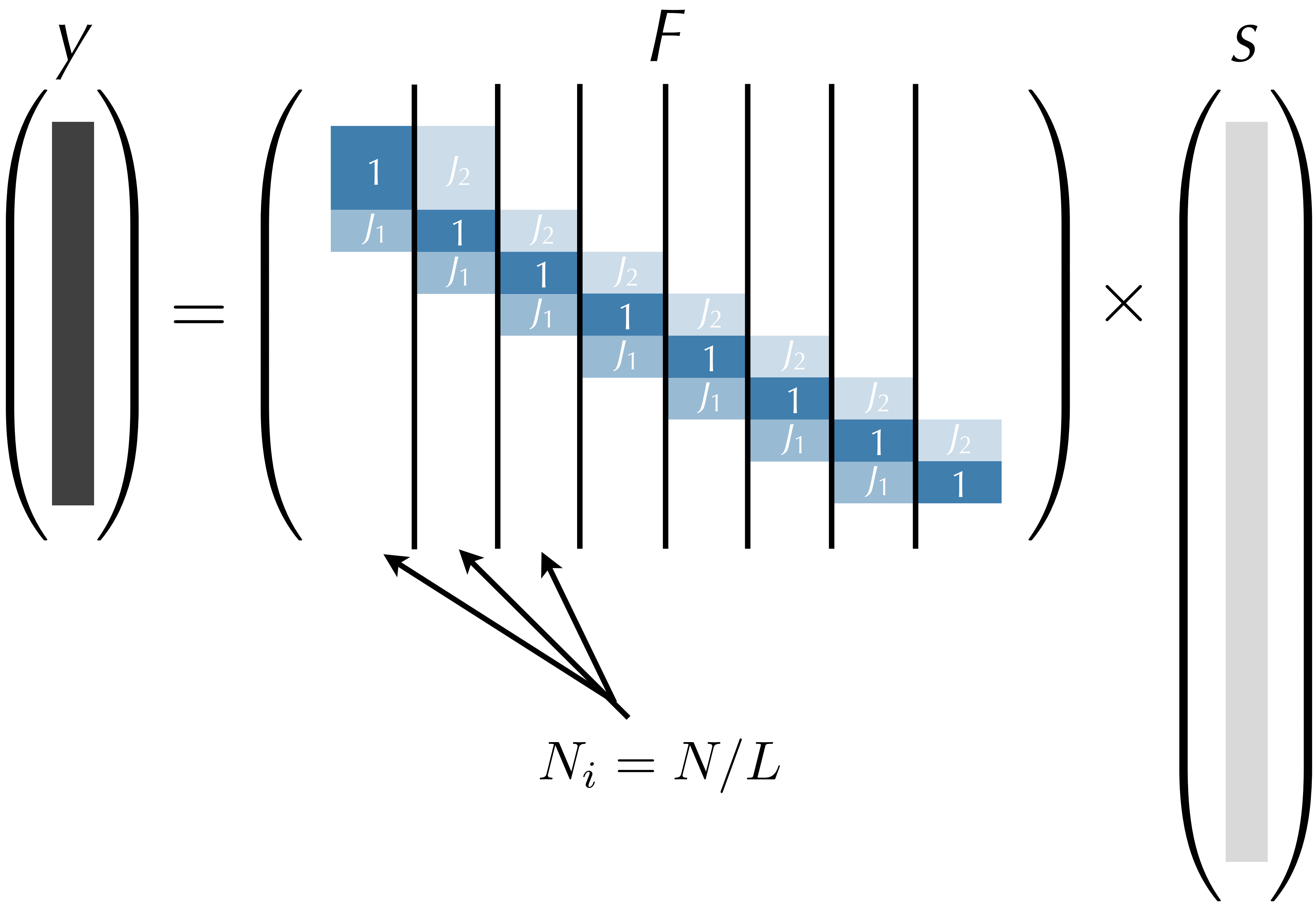
$$N_i = N/L$$

$$M_i = \alpha_i N/L$$

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

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

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



$L = 8$

$N_i = N/L$

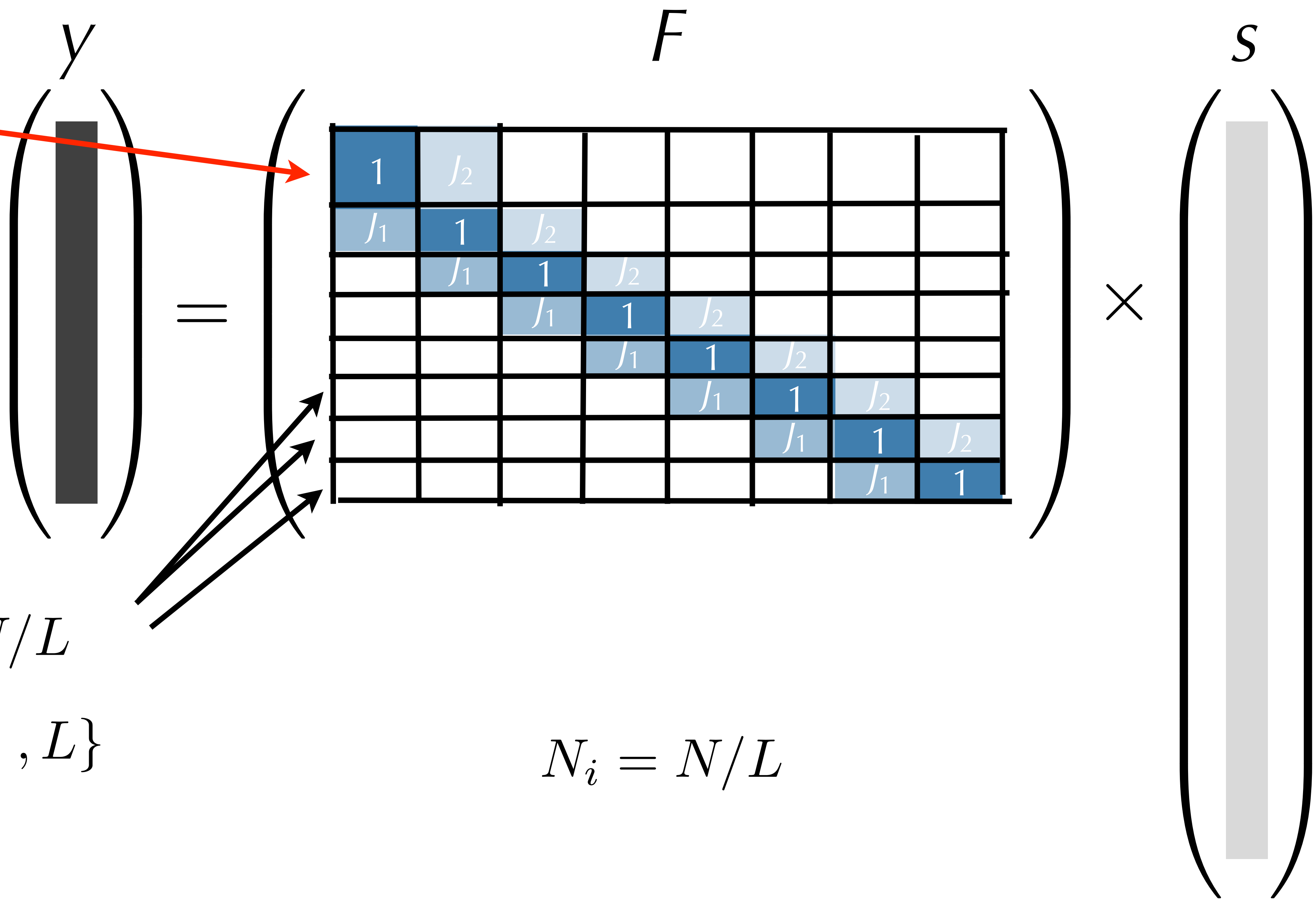
$M_i = \alpha_i N/L$

$\alpha_1 > \alpha_{BP}$

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

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

$M_1 = \alpha_1 N/L$



$M_i = \alpha' N/L$

$i \in \{2, \cdots, L\}$

$L = 8$

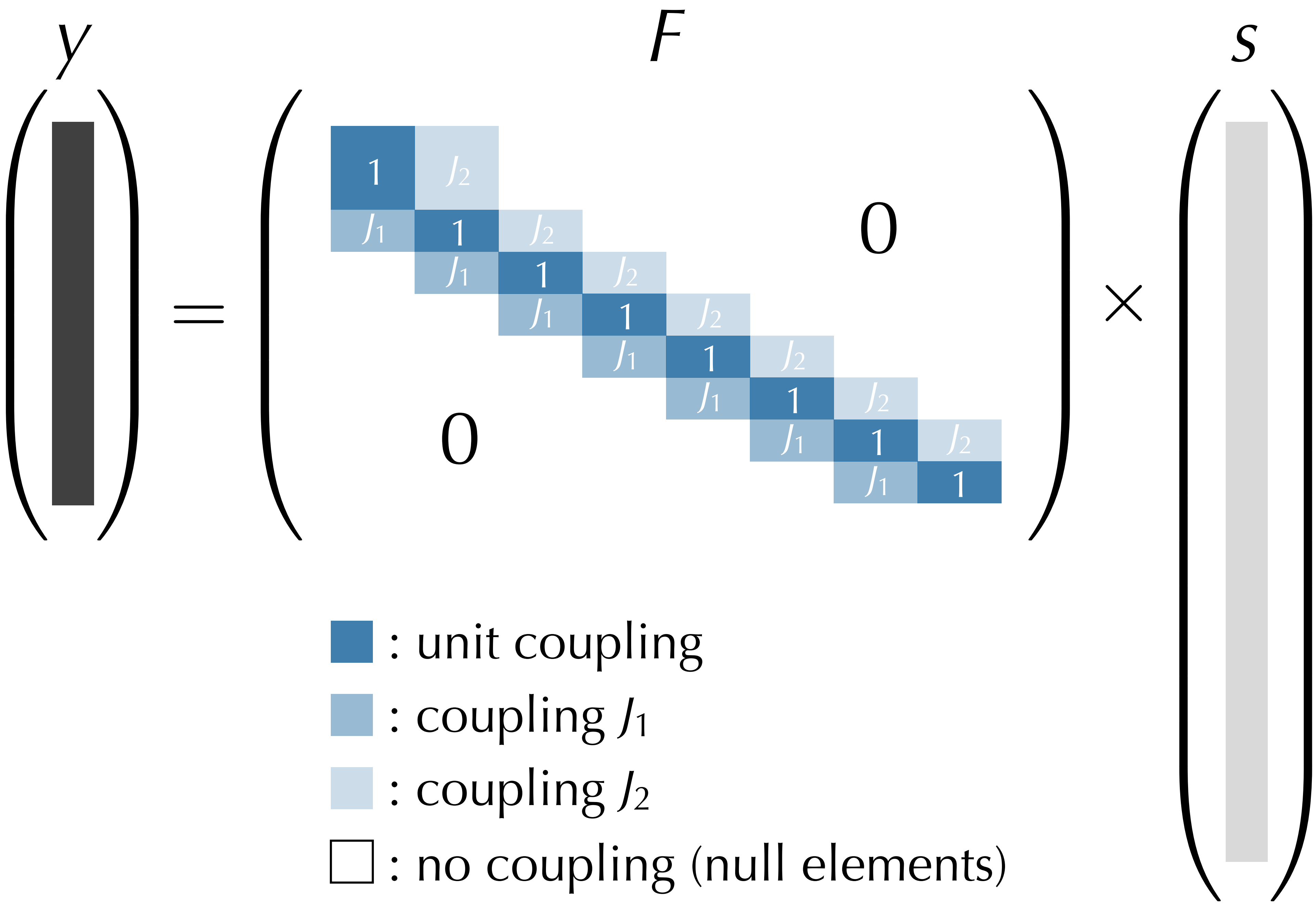
$N_i = N/L$

$M_i = \alpha_i N/L$

$\alpha_1 > \alpha_{BP}$

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

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



$$L = 8$$

$$N_i = N/L$$

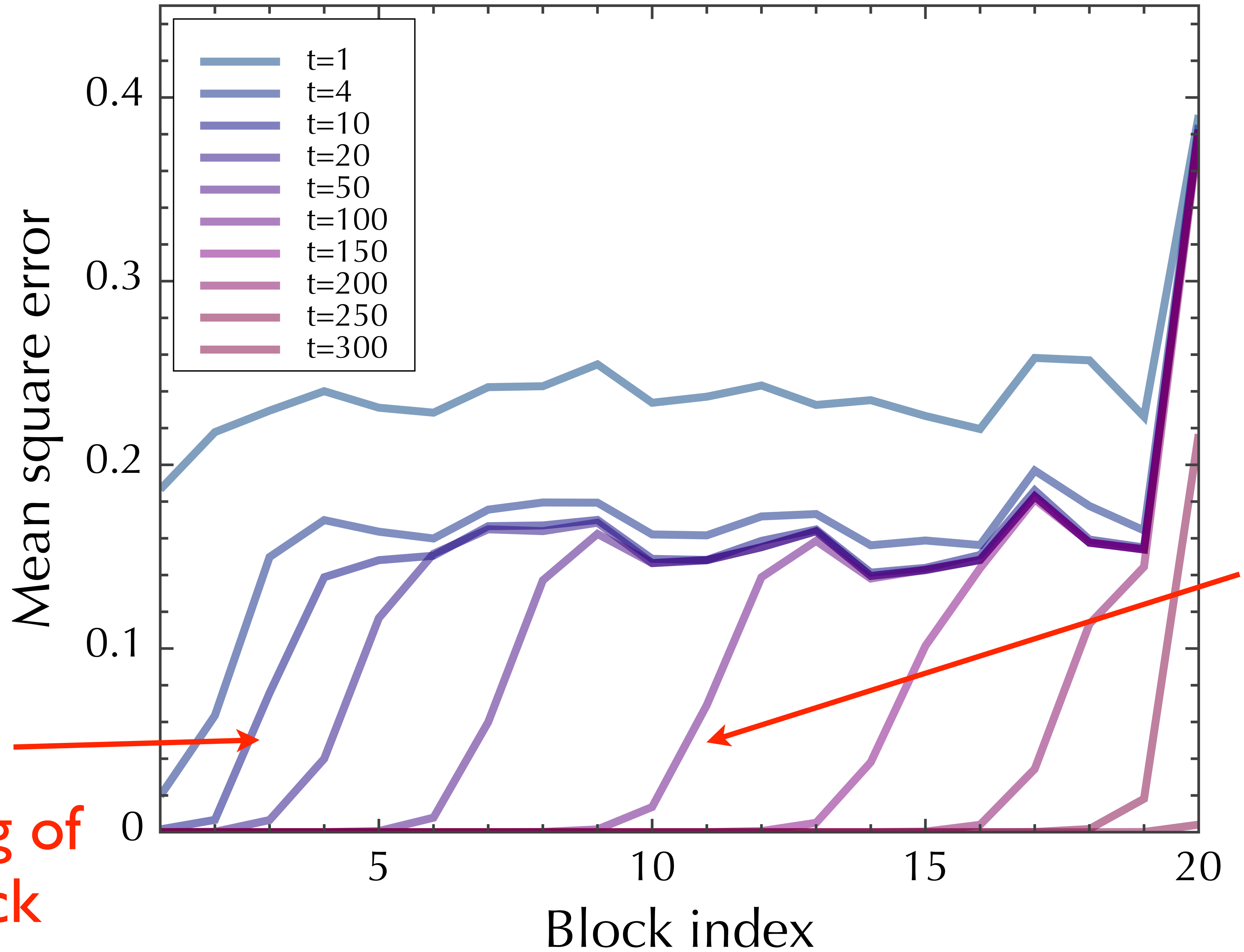
$$M_i = \alpha_i N/L$$

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

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

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

Numerical study



$t = 10$ decoding of first block

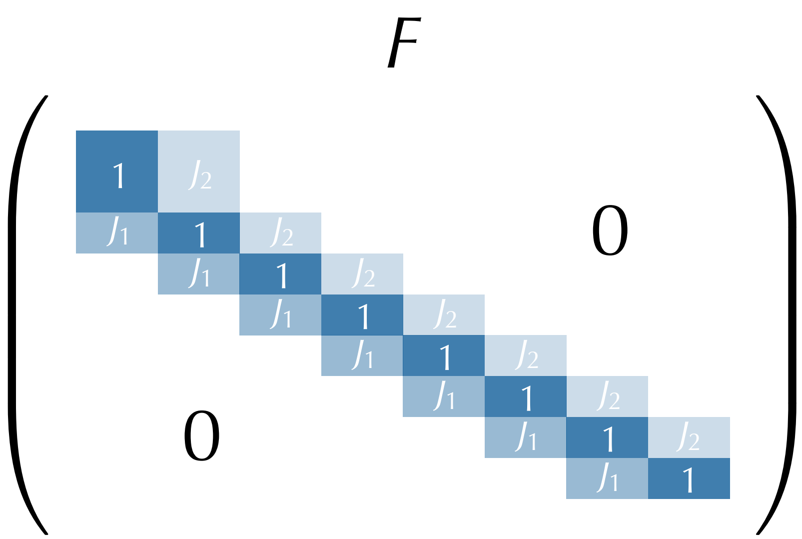
$t = 100$ decoding of blocks 1 to 9

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

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

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

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

$$E(Z^n) = \max_{\{D_r, V_r,\}} e^{Nn\Phi(D_1,V_1,\dots,D_L,V_L)}$$

Φ is known

$2L$ order parameters:

$$D_r = \frac{1}{N/L} \sum_{i \in B_r} \left(\langle x_i \rangle - s_i\right)^2 \qquad V_r = \frac{1}{N/L} \sum_{i \in B_r} \left(\langle x_i^2 \rangle - \langle x_i^2 \rangle\right)$$

Cavity approach shows that the order parameters of the BP iteration +parameter learning flow according to the gradient of the replica free entropy Φ :

- Replica/cavity analysis

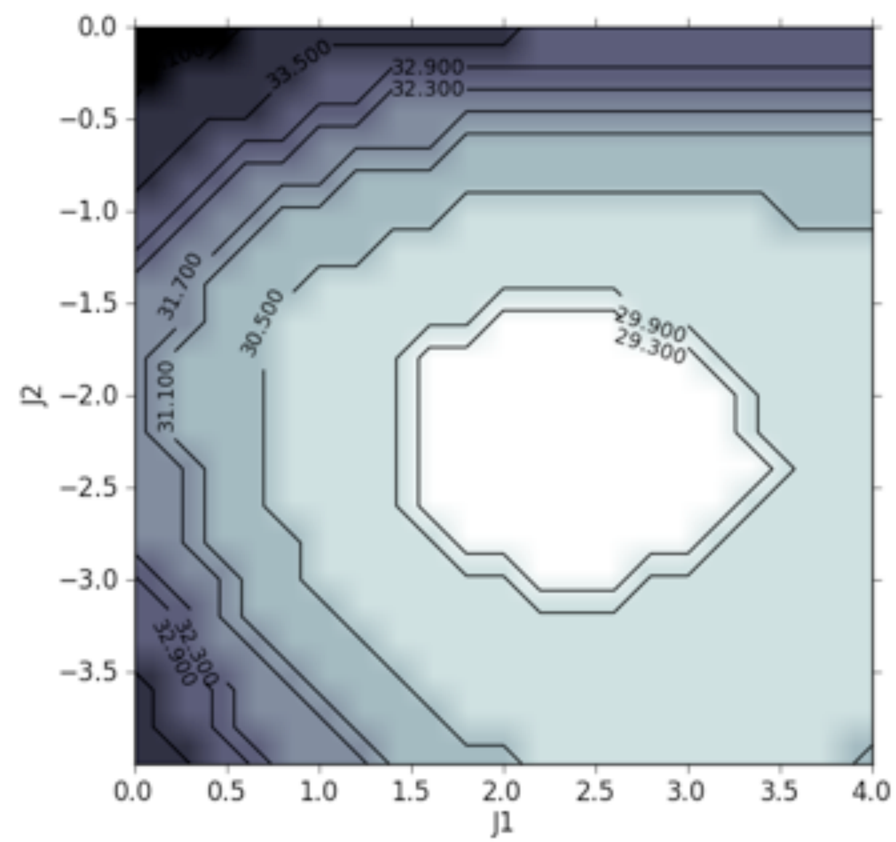
$$E(Z^n) = \max_{D_1,V_1,\ldots,D_L,V_L} e^{Nn\phi(D_1,V_1,\ldots,D_L,V_L)}$$

Time evolution of the s-BP algorithm: follows the gradient of $\phi(D_1, V_1, \ldots, D_L, V_L)$

+ evolution of the parameters

Dynamical system, dimension 2L+3

→ Convergence time for the infinite N system



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

- Replica/cavity analysis

$$E(Z^n) = \max_{D_1,V_1,\ldots,D_L,V_L} e^{Nn\phi(D_1,V_1,\ldots,D_L,V_L)}$$

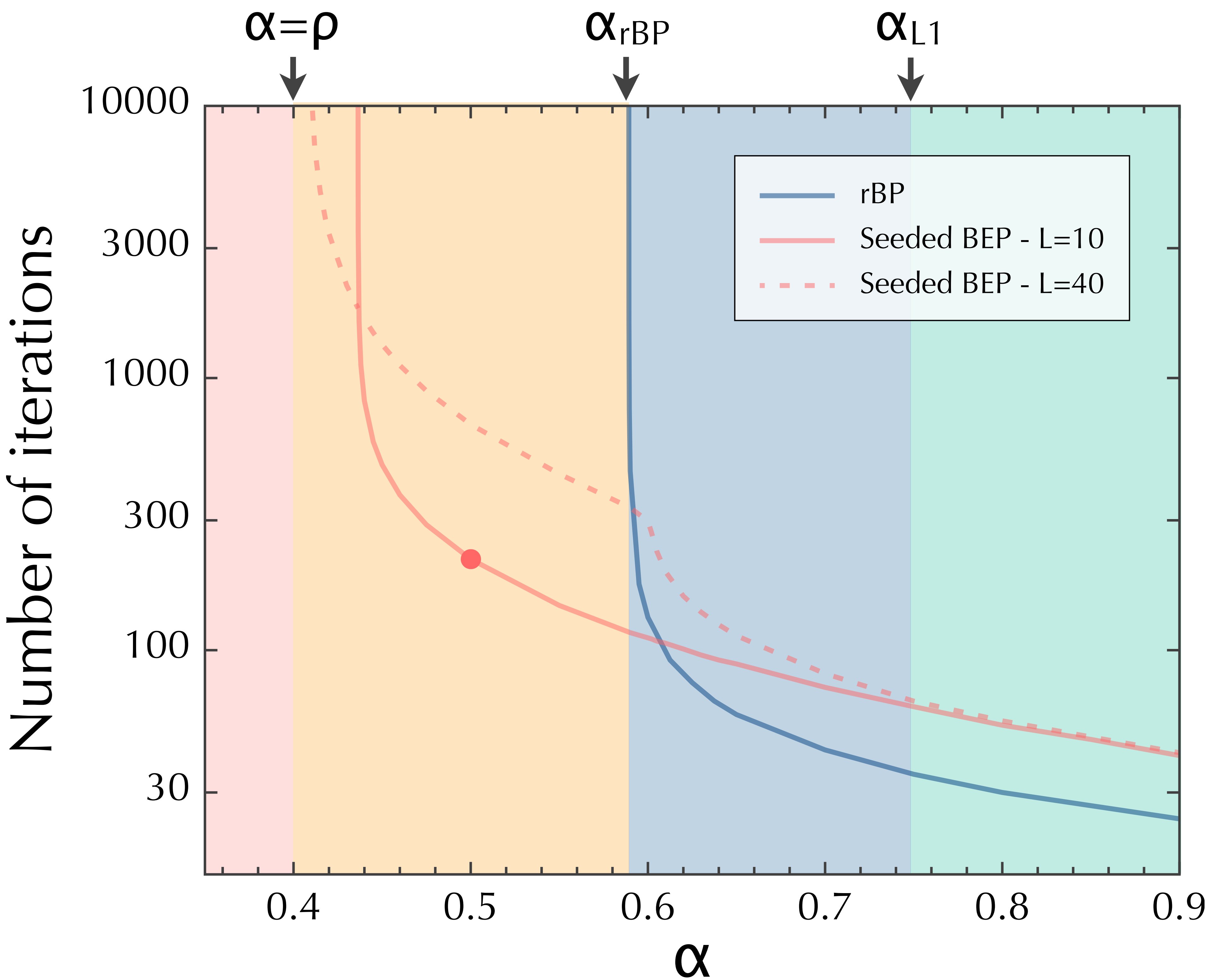
Numerical study of the dynamical system

optimize J_1, J_2 → there is no dynamical glass transition (in the large L limit)

$$\alpha = \rho_0 + \frac{c}{L} \qquad \rightarrow \alpha_c = \rho_0$$

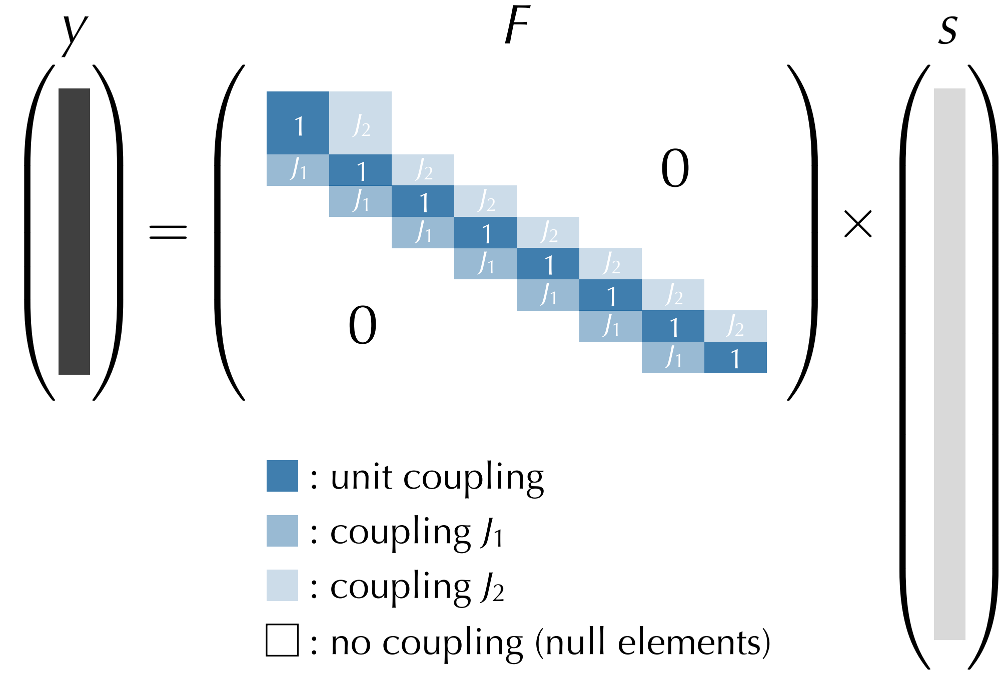
Recent proof: Donoho Javanmard Montanari

Numerical study



Analytic study: cavity equations, density evolution, replicas

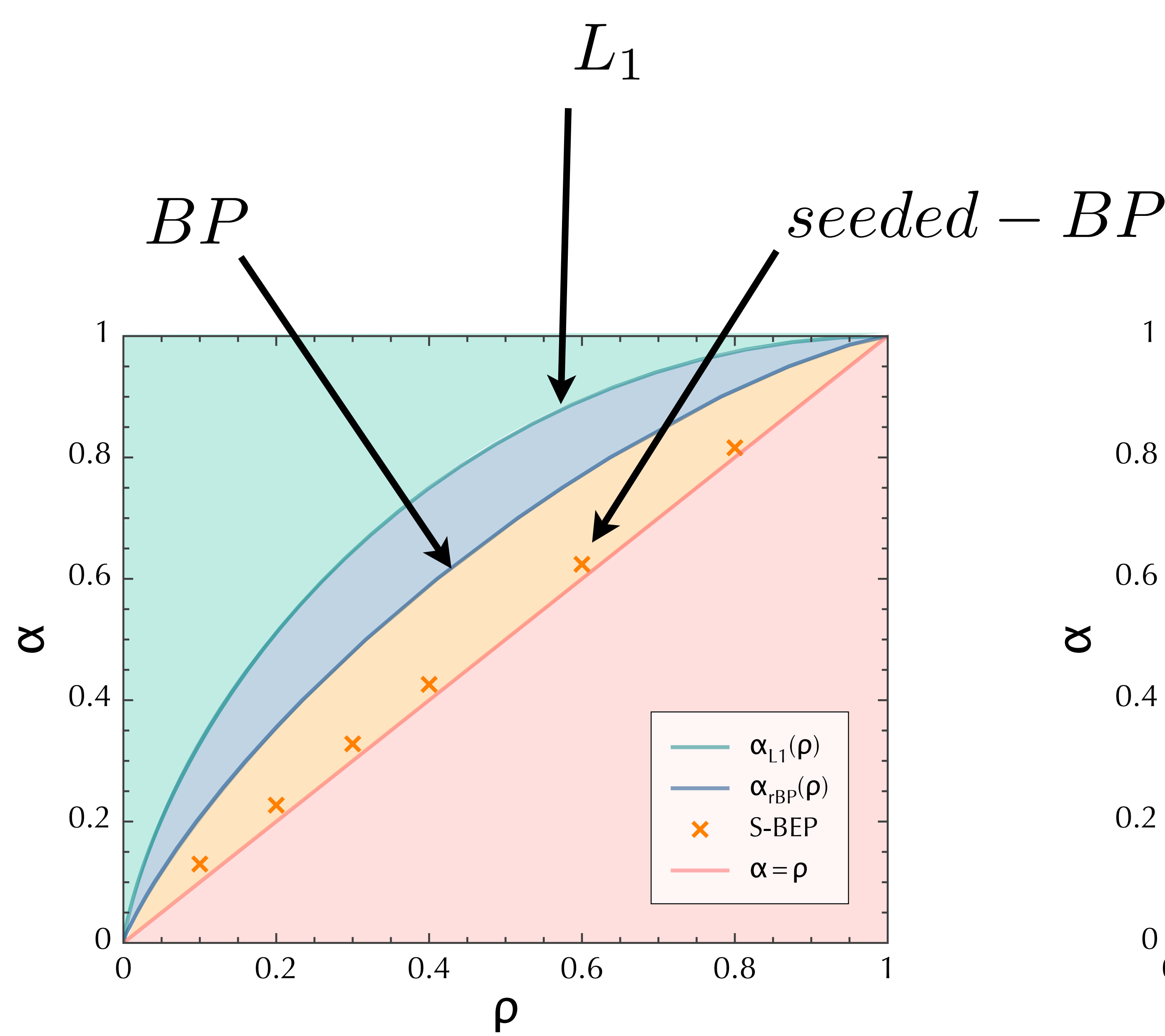
Replica study of the seeding measurement matrix : in some regimes of $\alpha_1,\ J_1,\ J_2$

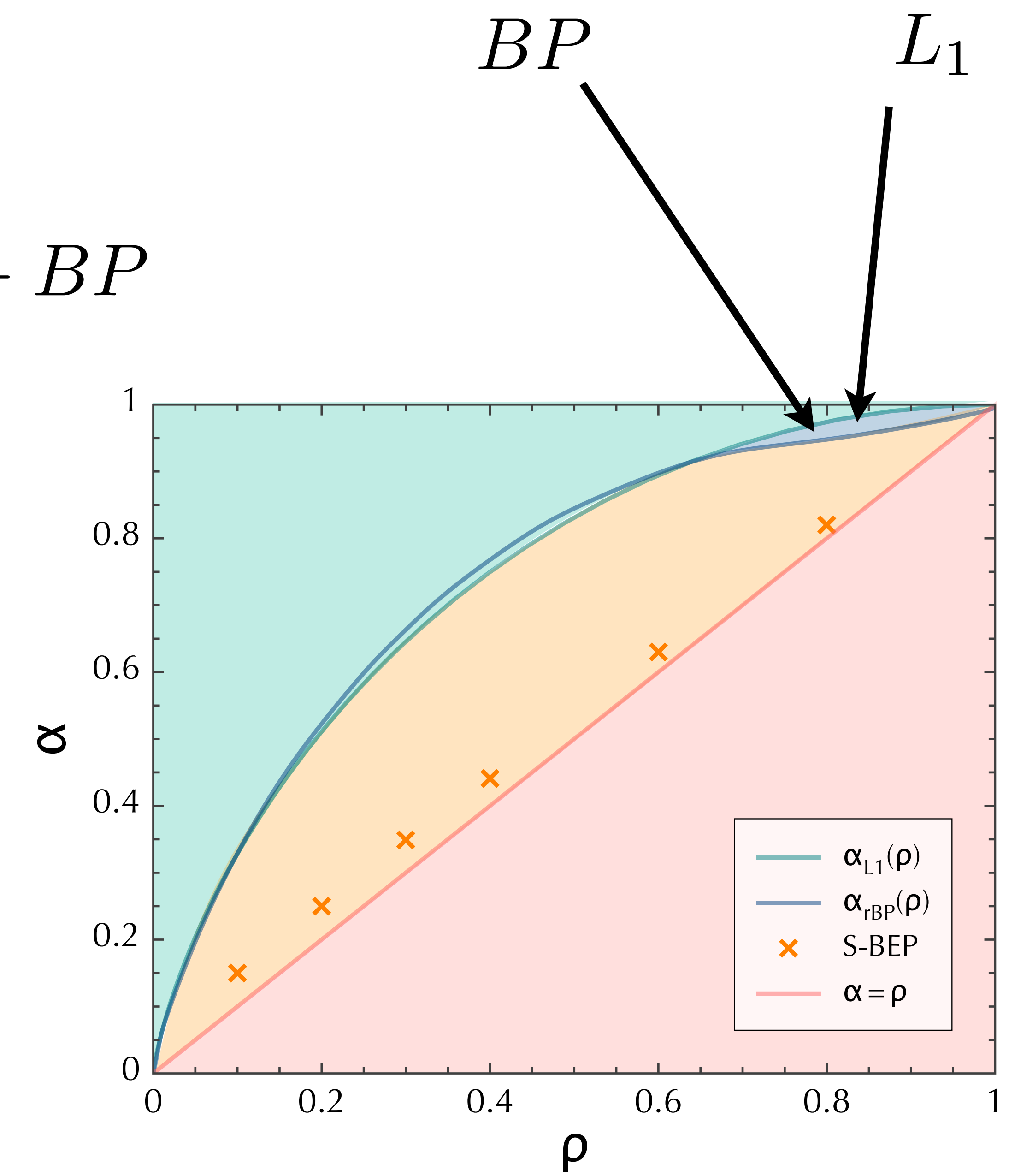


there is no dynamical glass transition (in the large L limit)

possible to reach the optimal compressed sensing limit $\alpha = \rho$

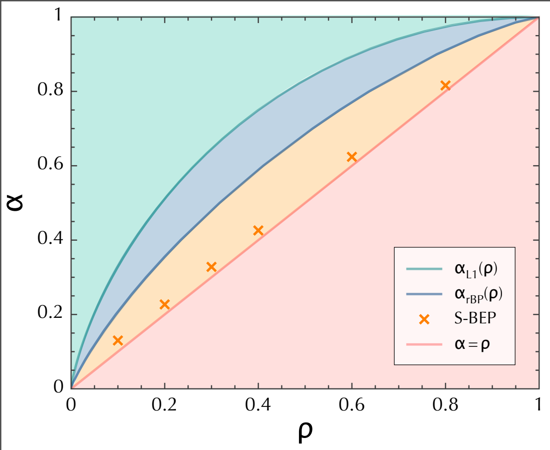
Gaussian signal   Binary signal





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

L_1 phase transition line moves up when using seeding F

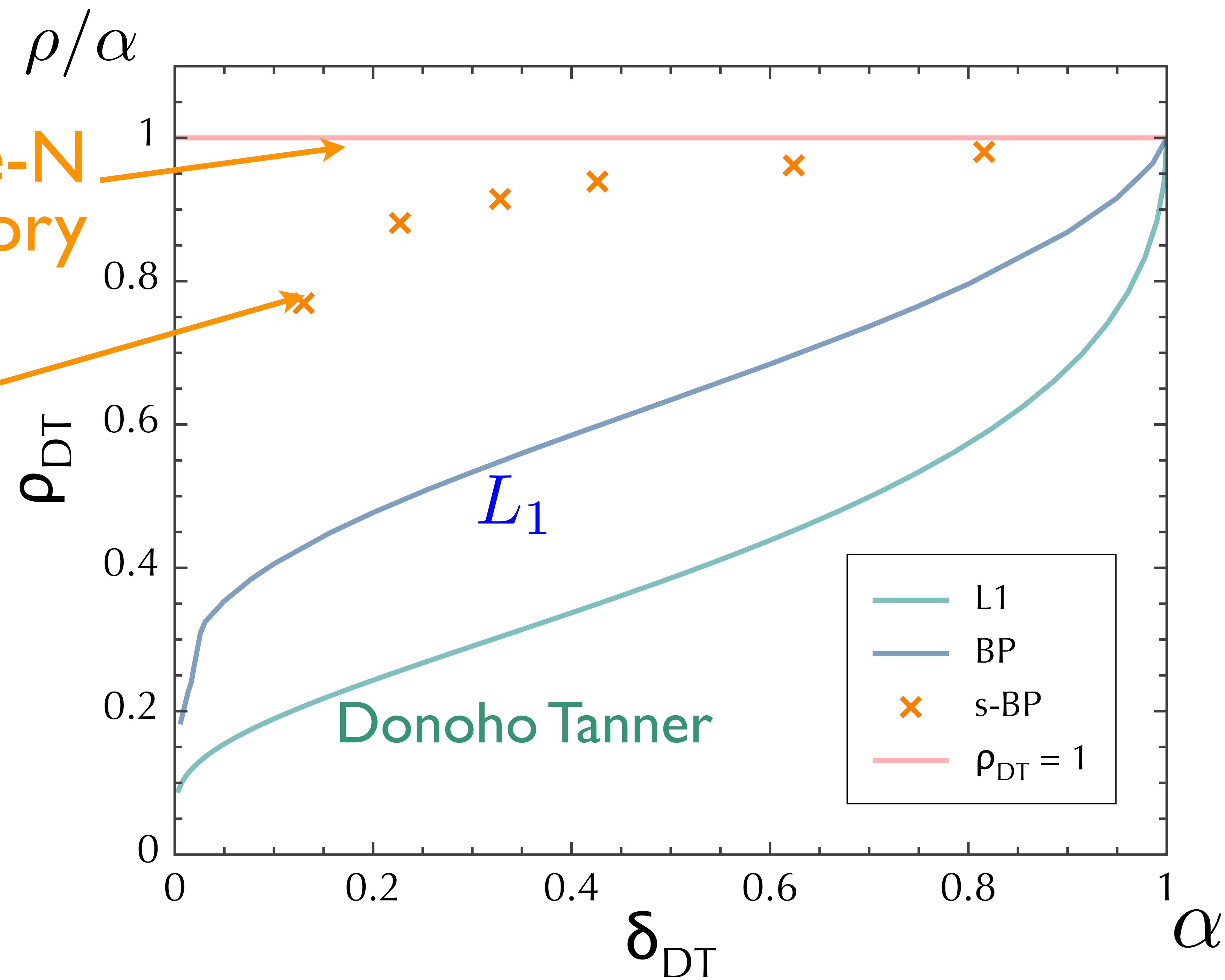


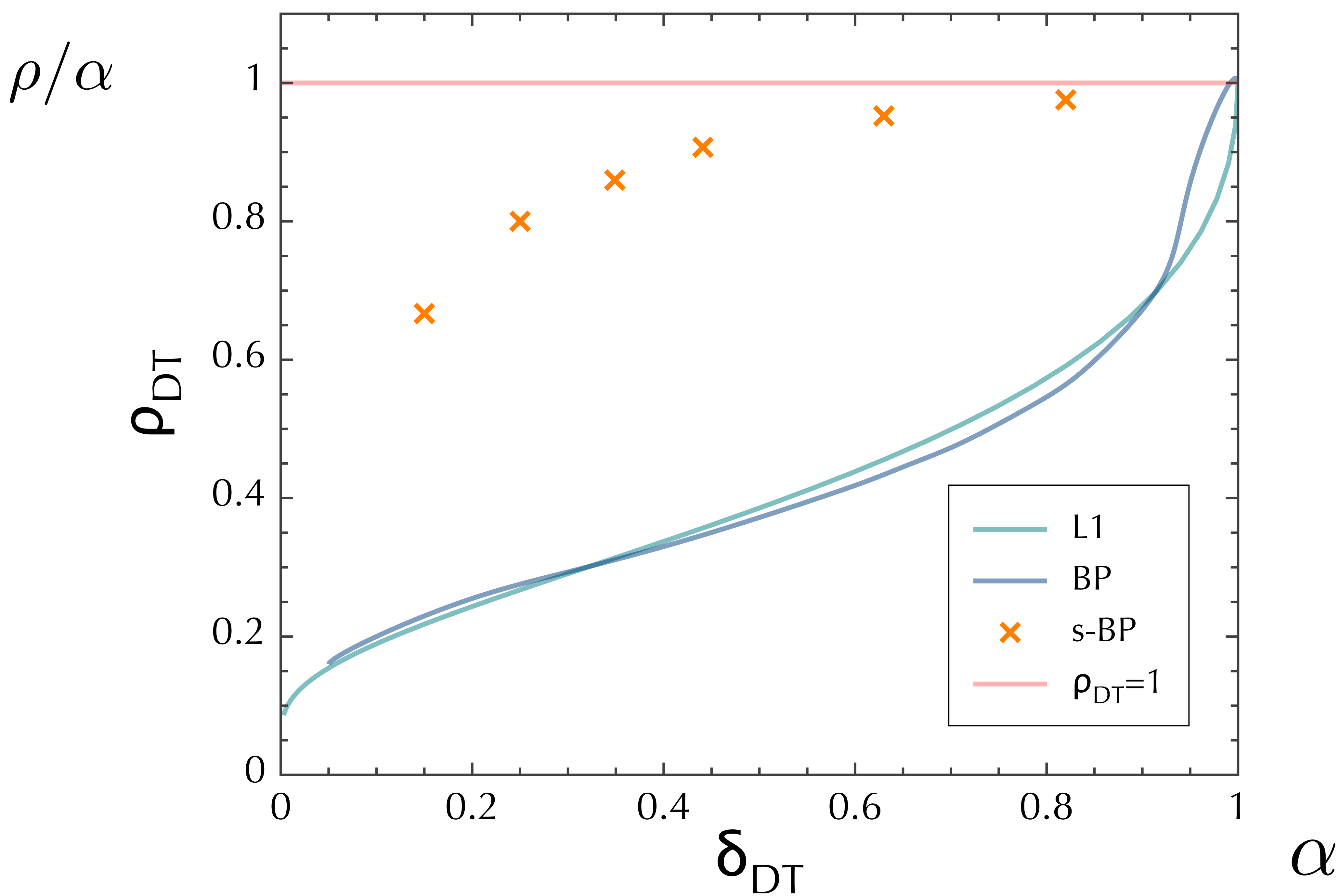
with US notation:

Our s-BP algorithm:

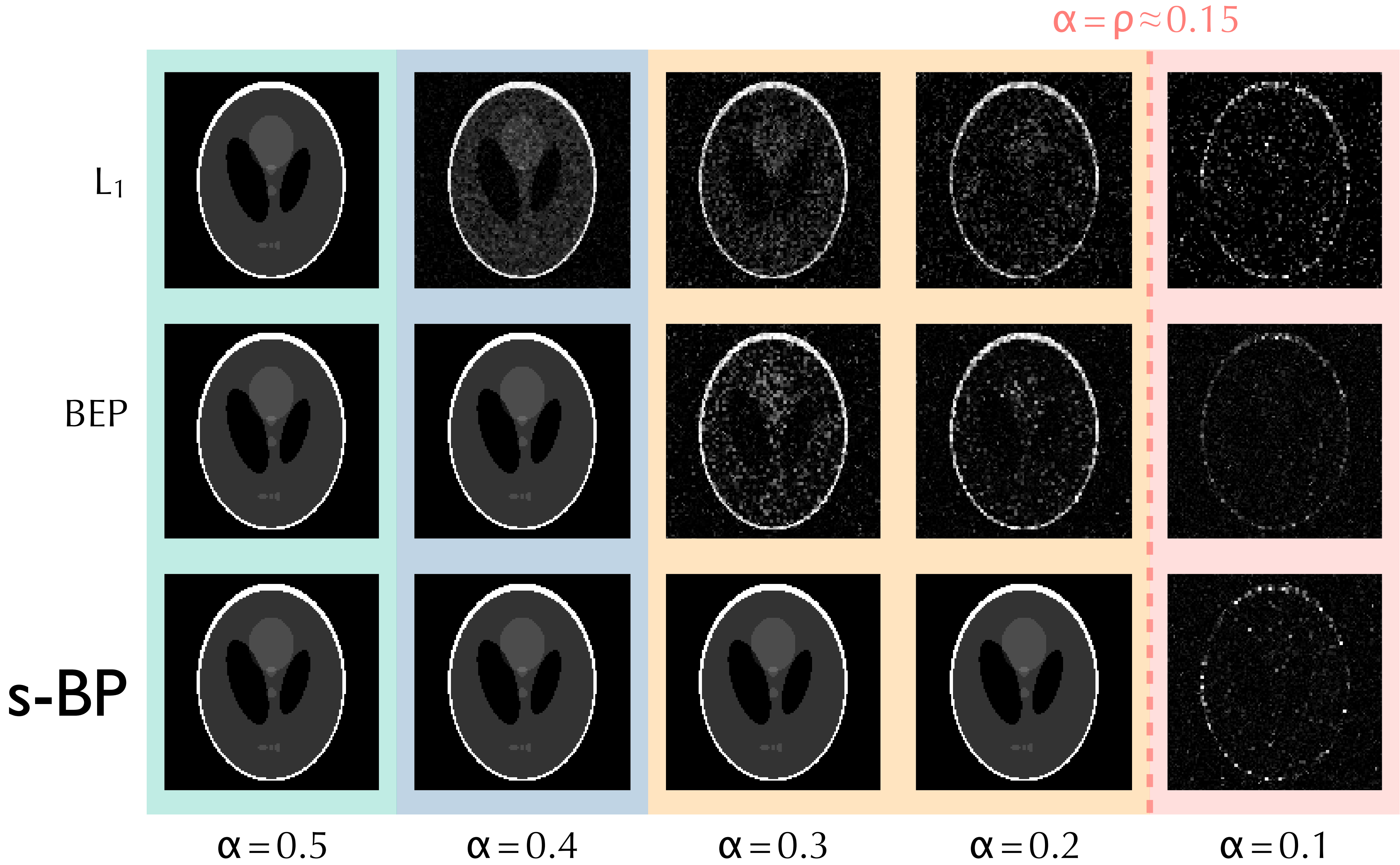
large-N theory

simulations

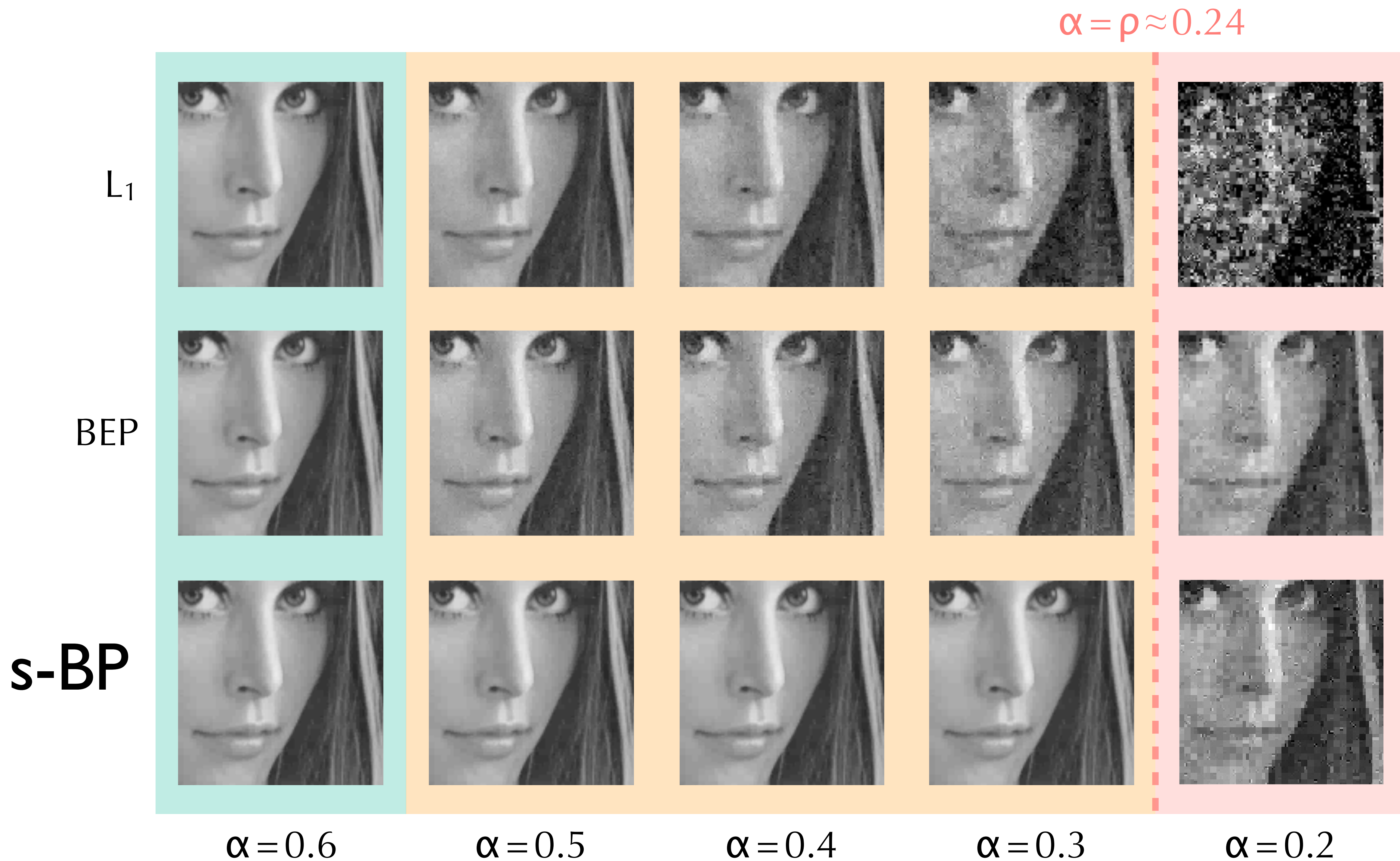




ρ/α
ρ_{DT}
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
δ_{DT}
α
L1
BP
s-BP
ρ_{DT}=1



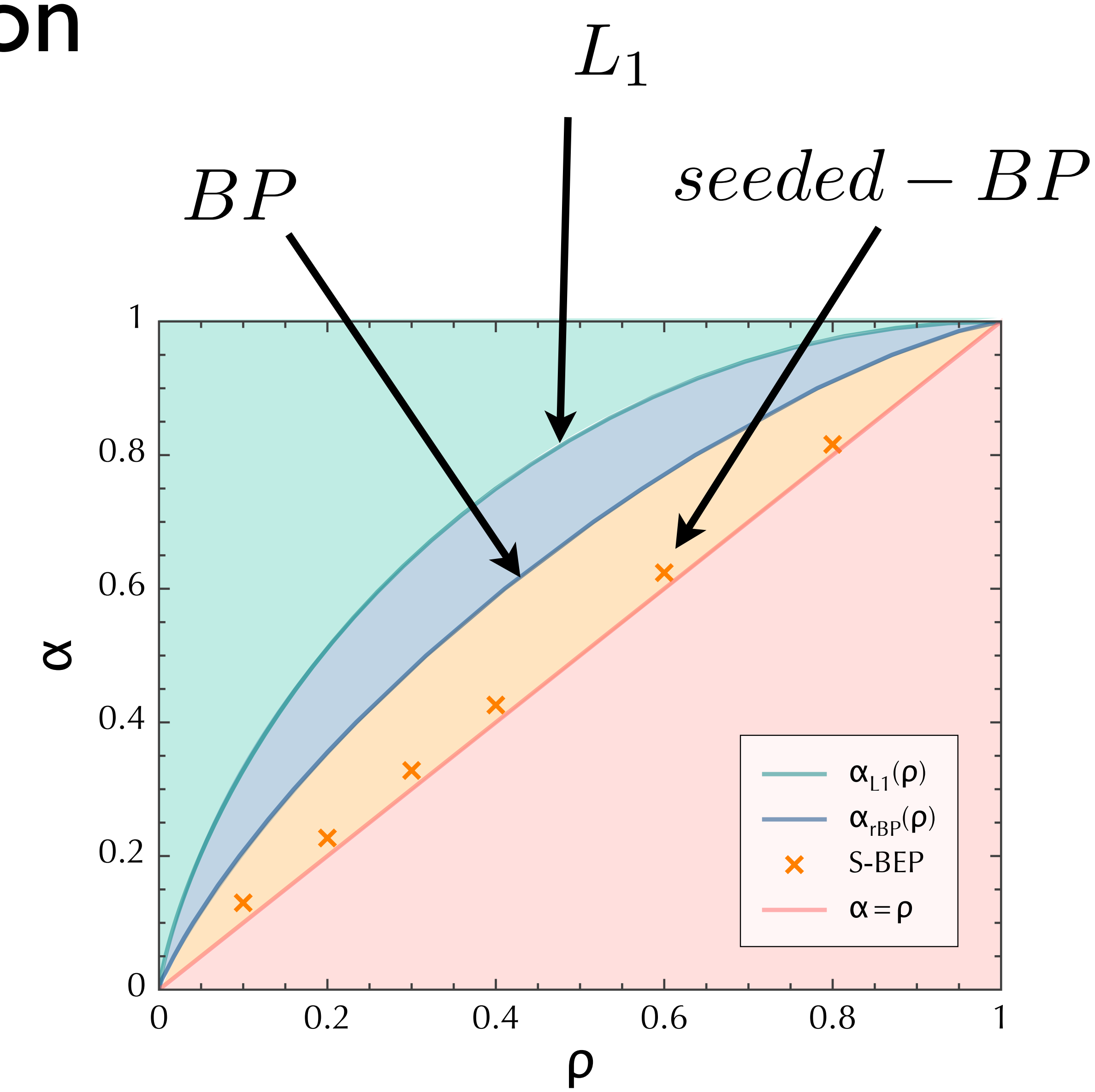
Shepp-Logan phantom, in the Haar-wavelet representation


$\alpha = \rho \approx 0.24$
L_1
BEP
s-BP
$\alpha = 0.6$
$\alpha = 0.5$
$\alpha = 0.4$
$\alpha = 0.3$
$\alpha = 0.2$

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

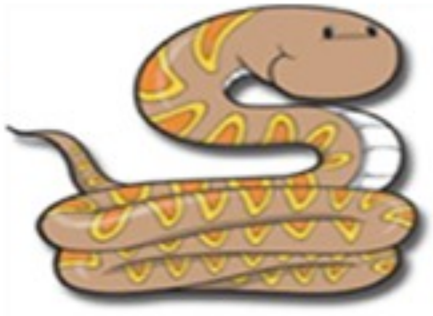


Very powerful CS solver, available on our «ASPICS» webpage:

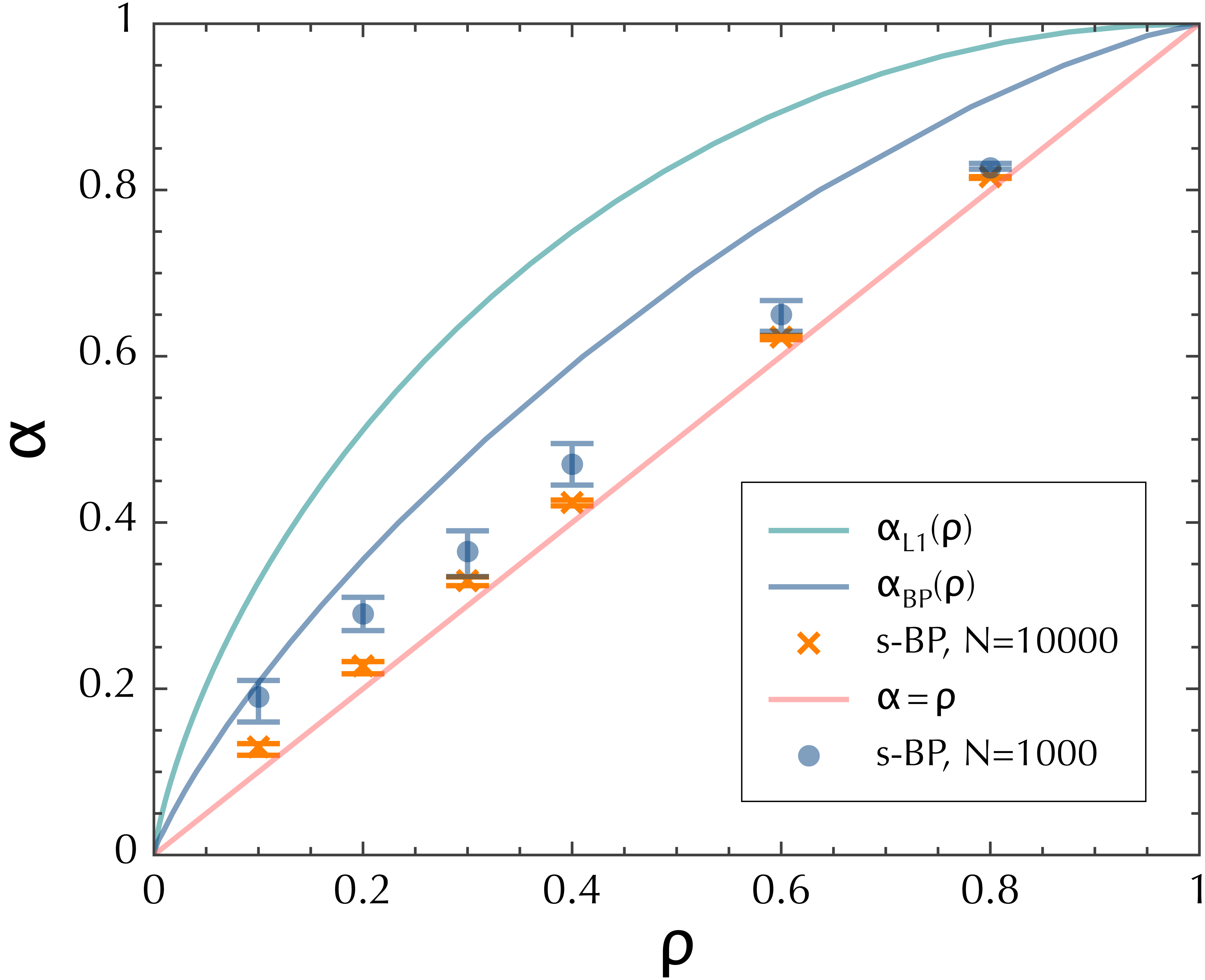
www.

ASPICS: Applying Statistical Physics to Inference in Compressed Sensing

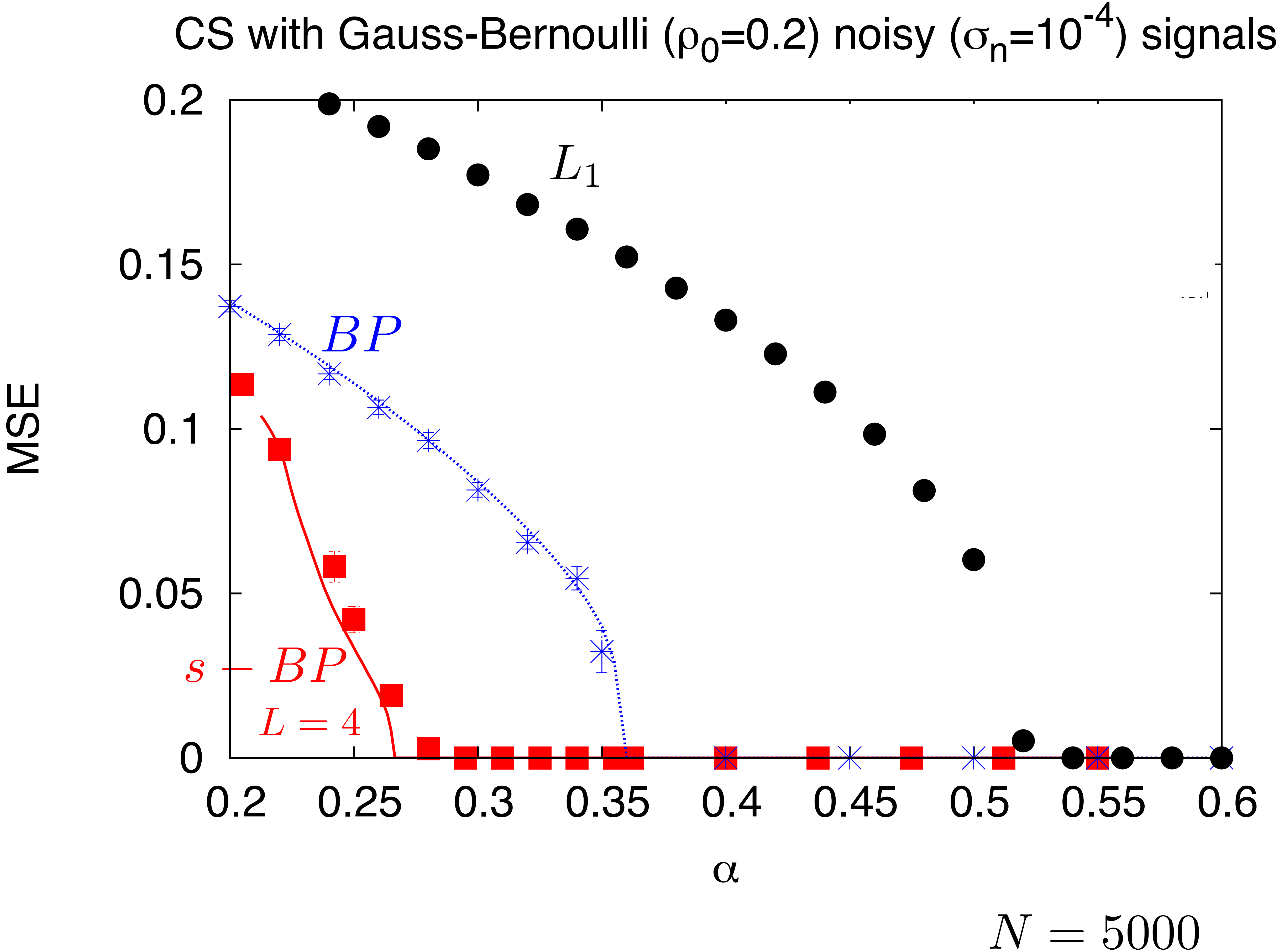
This page contains codes, papers and data on the algorithms we have been developing using statistical physics for inference in interdisciplinary applications, with a special focus on compressed sensing.

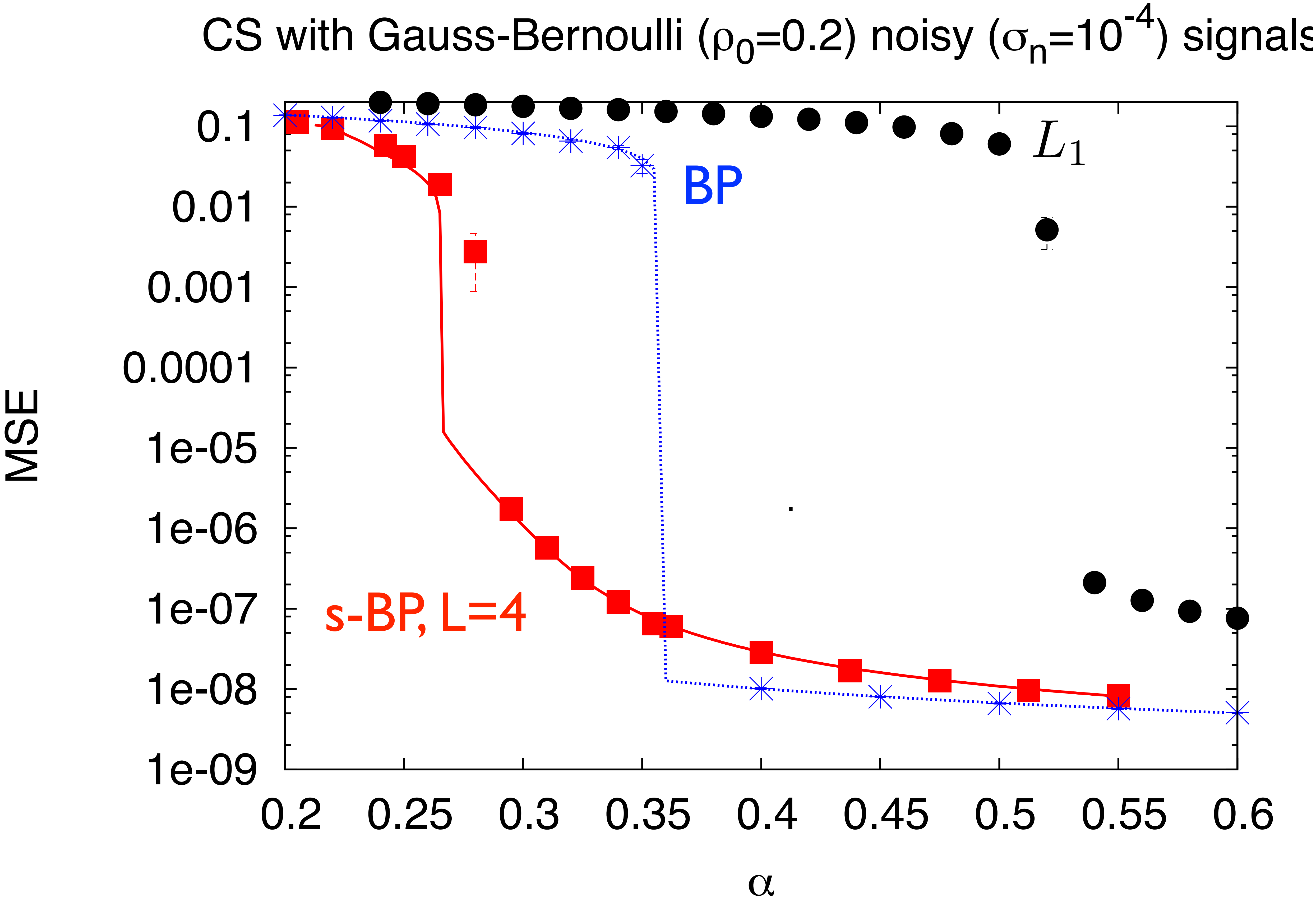


Finite-size effects



Robustness to noise





CS with Gauss-Bernoulli (ρ_0=0.2) noisy (σ_n=10^{-4}) signals
MSE
α
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
L_1
BP
s-BP, L=4