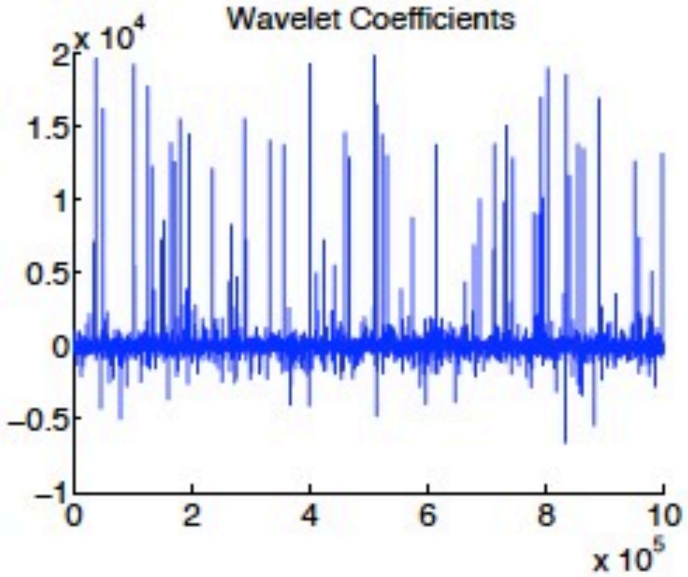
Statistical physics approach to compressed sensing

Marc Mézard
LPTMS, Université Paris Sud, CNRS

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

arXiv:1109.4424

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,+...)

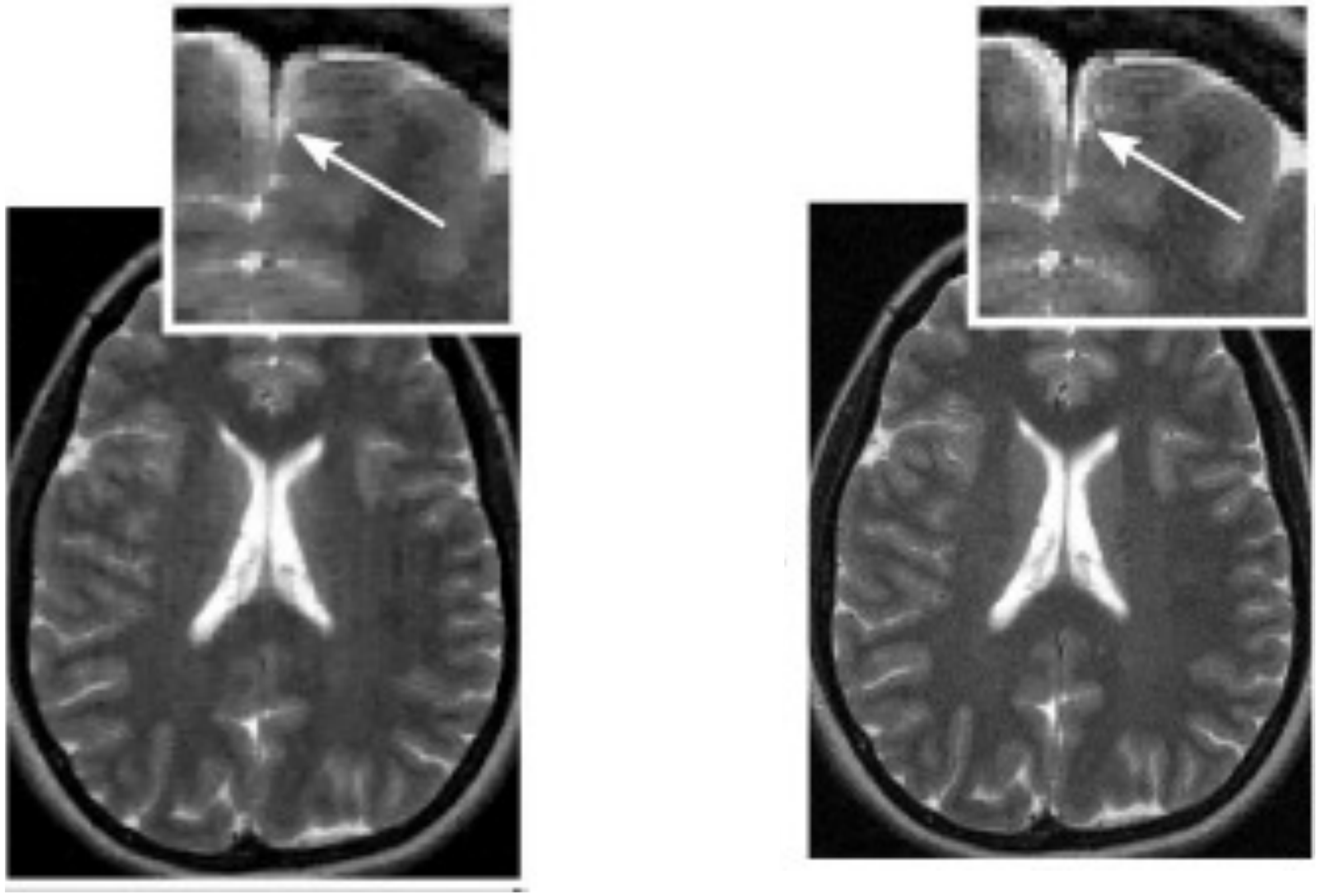
Compressed sensing

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

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

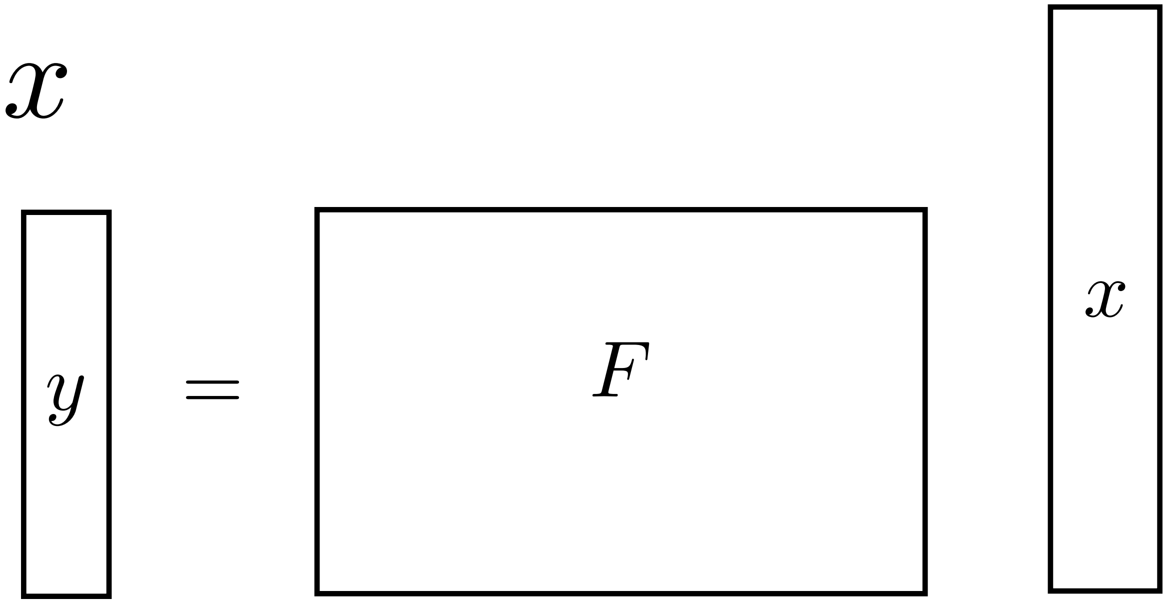
$F = M \times N$ matrix

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 $y = Fx$



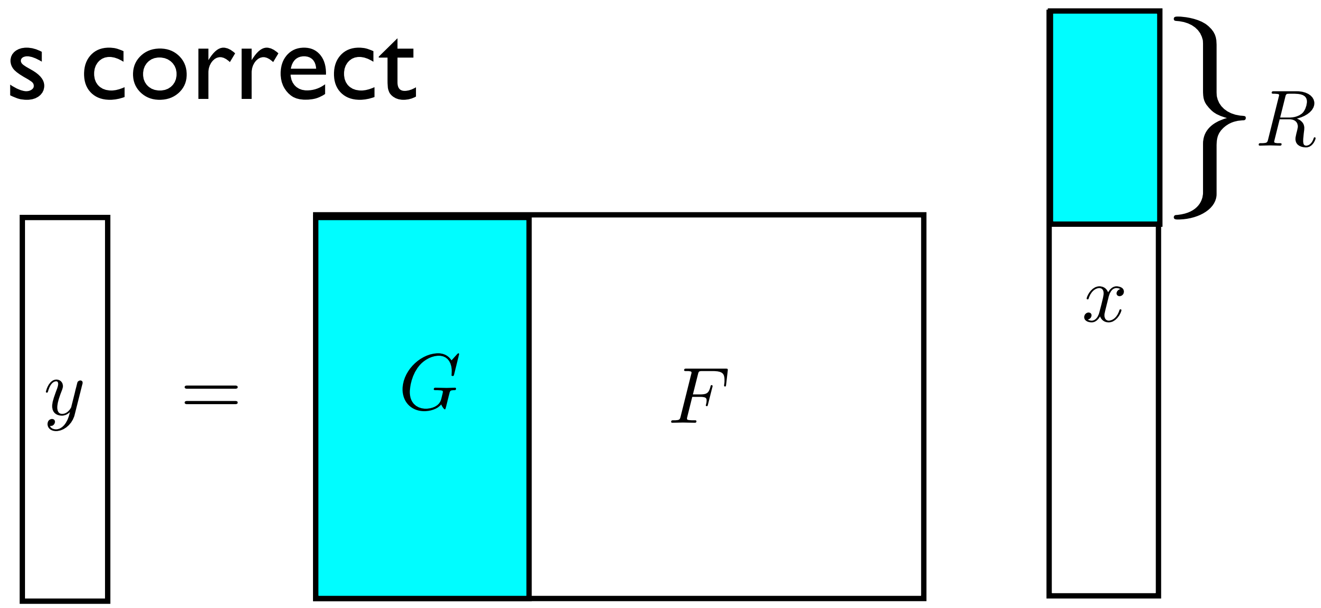
Exploit the sparsity of the original s

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. $x_1, \dots, x_R \neq 0$

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

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

$R < M$ ⇨ too many equations

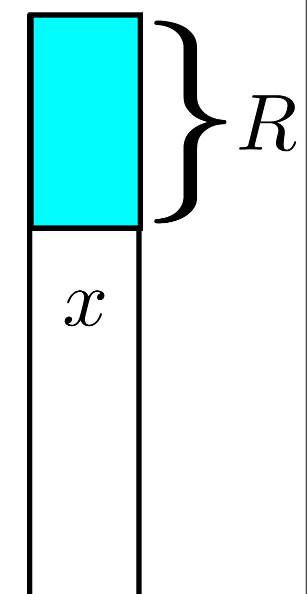
⇨ generically inconsistent (no solution), except if the guess of locations of $x_i \neq 0$ was correct

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

$R < M$

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

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

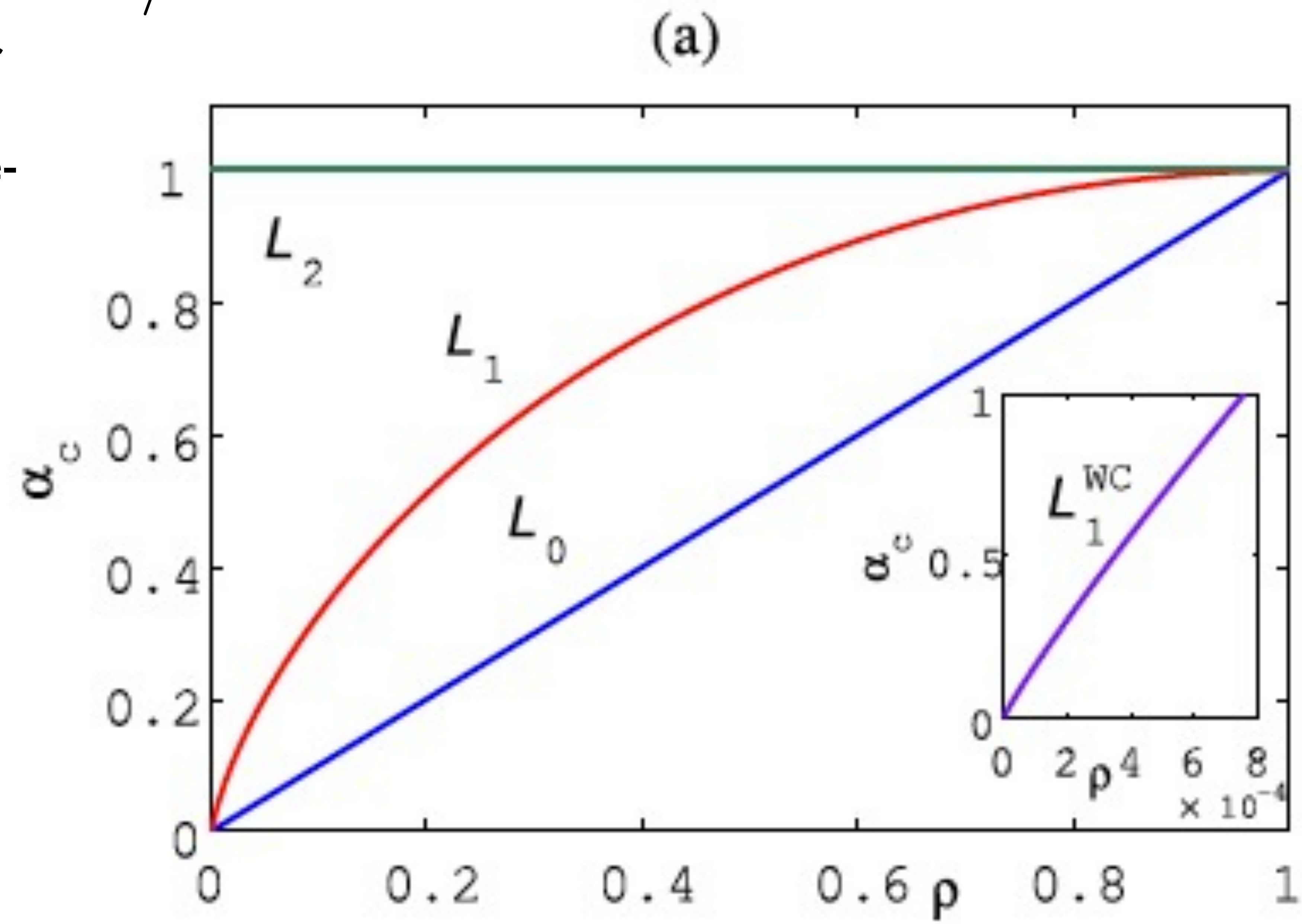
$M = \alpha N$ equations

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

Phase diagram

$\alpha = M/N$

Number of measurements per variable



Donoho 2006, Donoho Tanner 2005

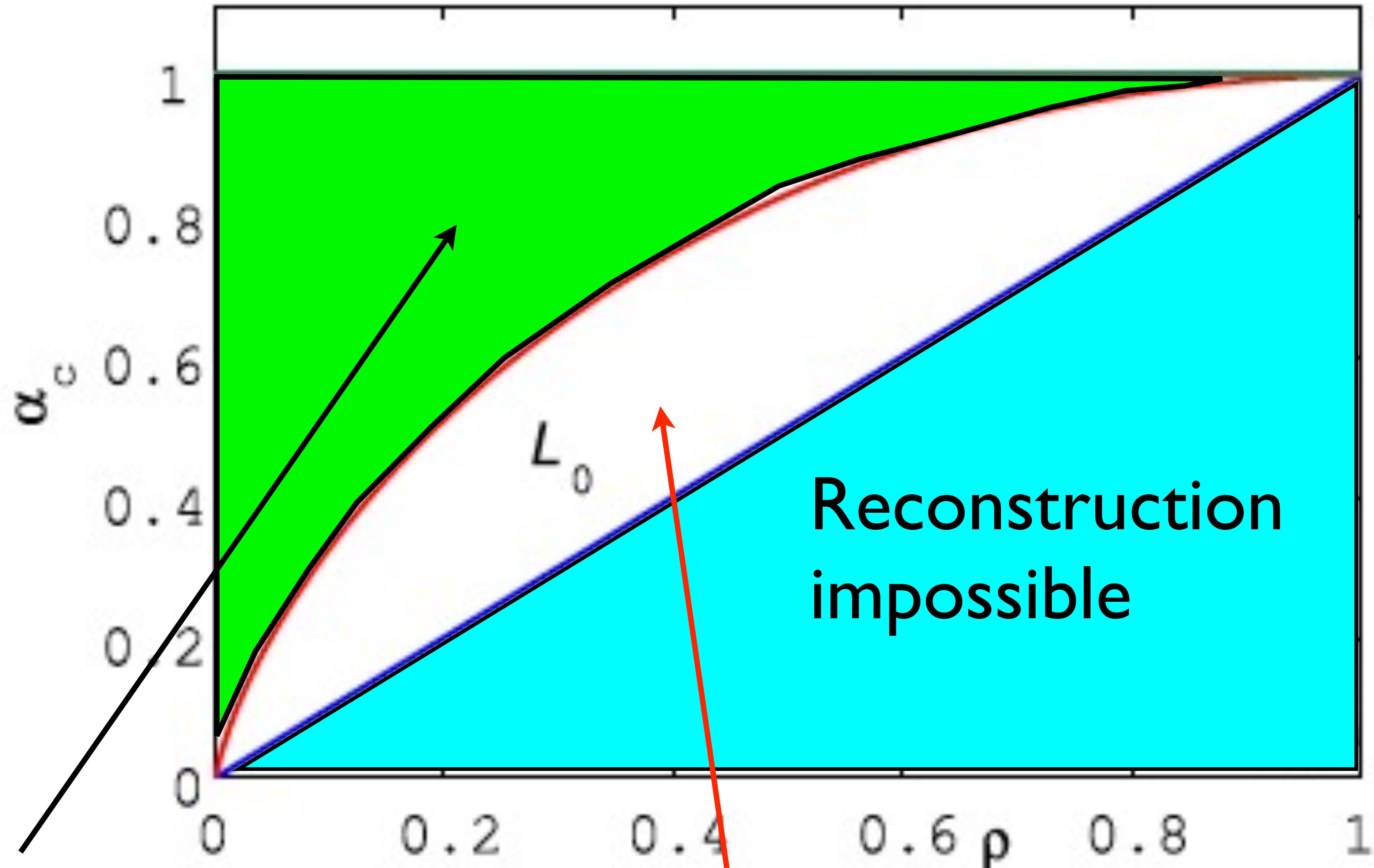
Kabashima, Wadayama and Tanaka, JSTAT 2009

$\rho = R/N$

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

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:

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

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 $\mathbf{x} = \mathbf{s}$

Also true for broader class of measurement matrices F

e.g. the seeding matrices to be used in the final design

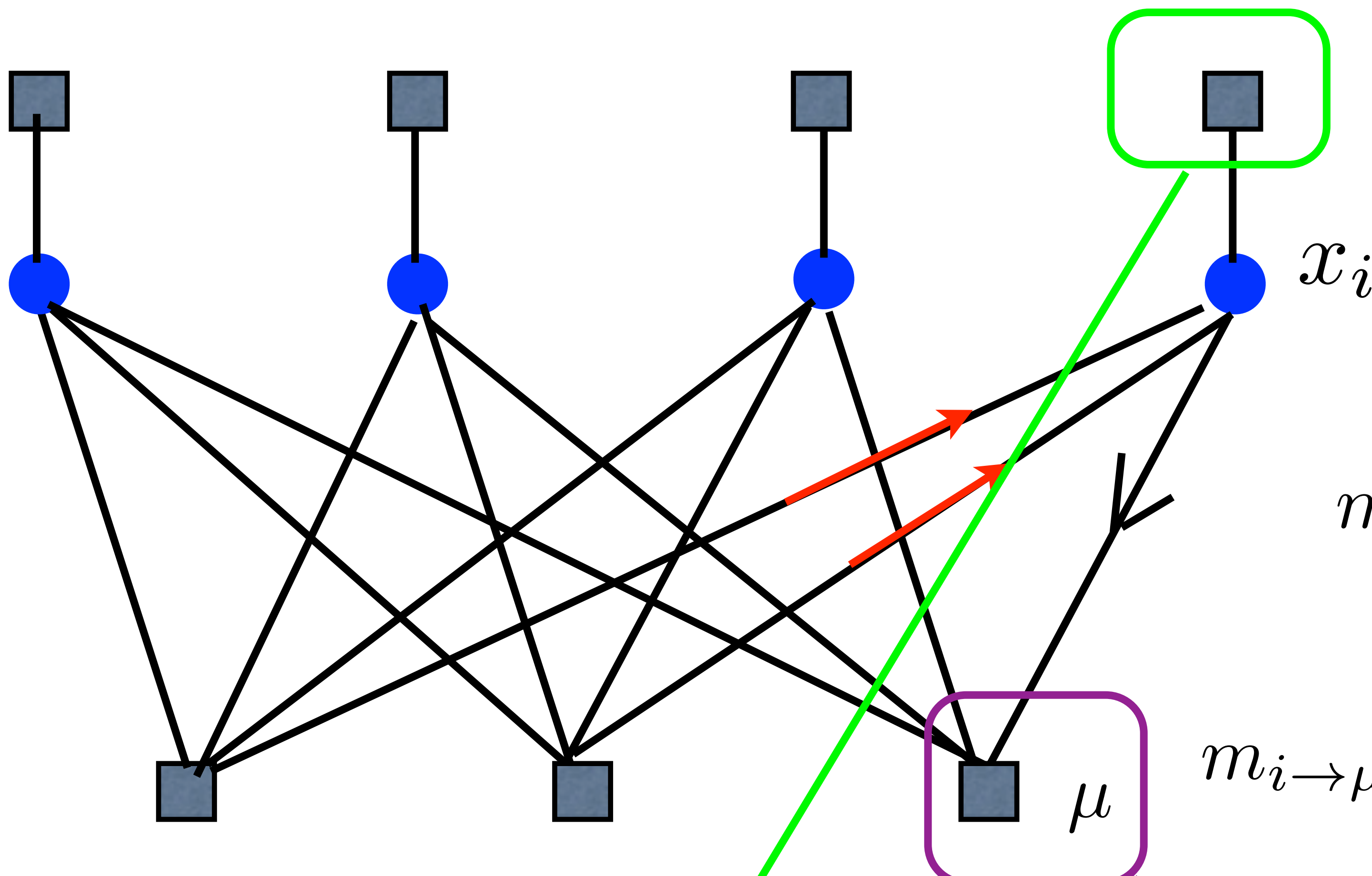
Step 2: belief propagation-based reconstruction with parameter learning

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

Message passing for compressed sensing



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

$$m_{i\to\mu} = f\left(\{m_{\nu\to i}\}, \nu \in \partial i \setminus \mu\right)$$

Useless as such !

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

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

NB : Possible further simplification: «Approximate Message Passing» (TAP-form)

Donoho-Montanari

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

$\sum_{\gamma\neq\mu} A_{\gamma\to i}$ depends only weakly on μ

Expansion to first order in the correction (Onsager's reaction term). Messages: two real numbers on each vertex

$$\omega_\mu = \sum_i F_{\mu i} a_{i\to\mu} \qquad \gamma_\mu = \sum_i F_{\mu i}^2 v_{i\to\mu}$$

$$U_i = \sum_\mu A_{\mu\to i} \qquad V_i = \sum_\mu B_{\mu\to i}$$

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:

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

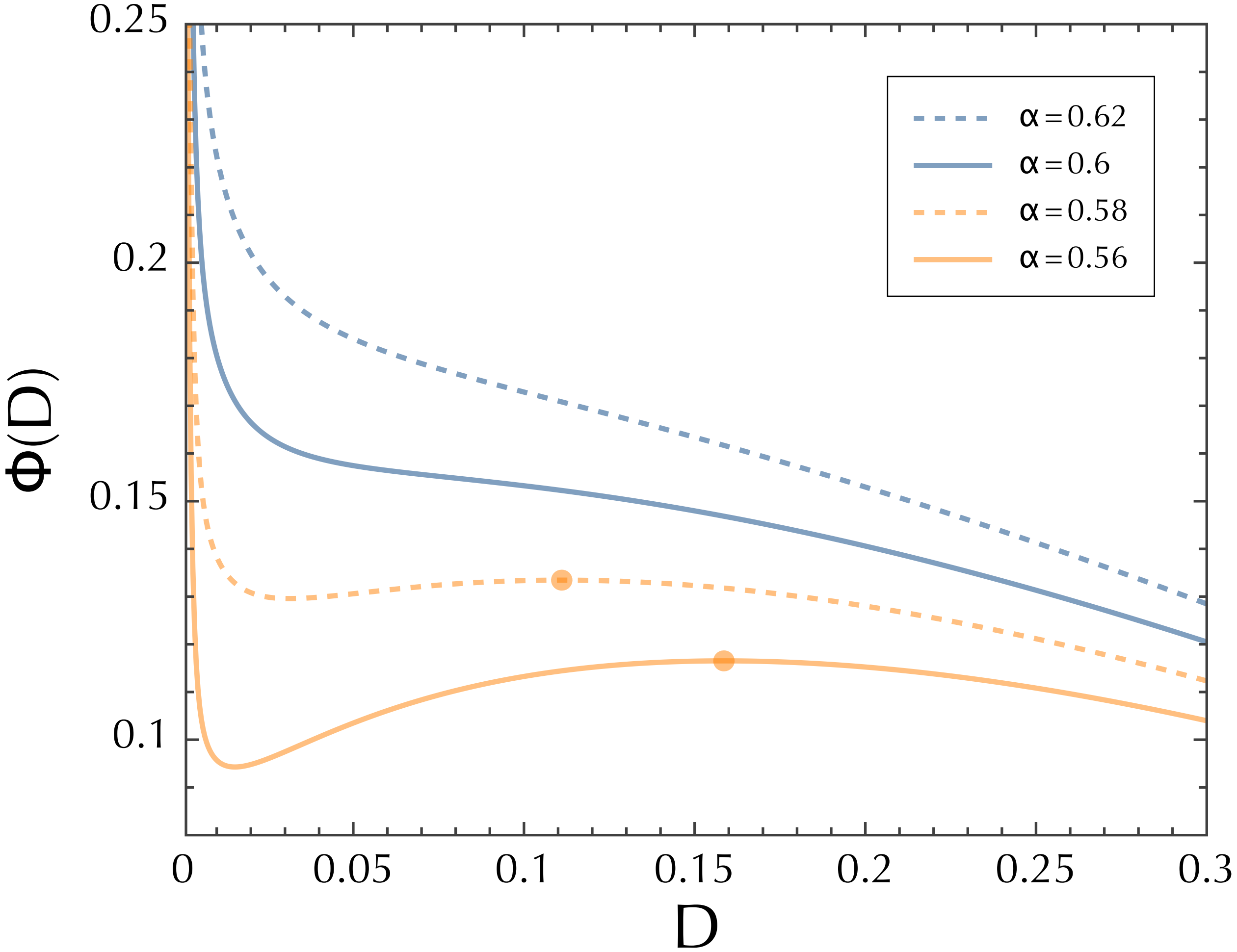
Cavity approach shows that the order parameters of the BP iteration flow according to the gradient of the replica free entropy Φ

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

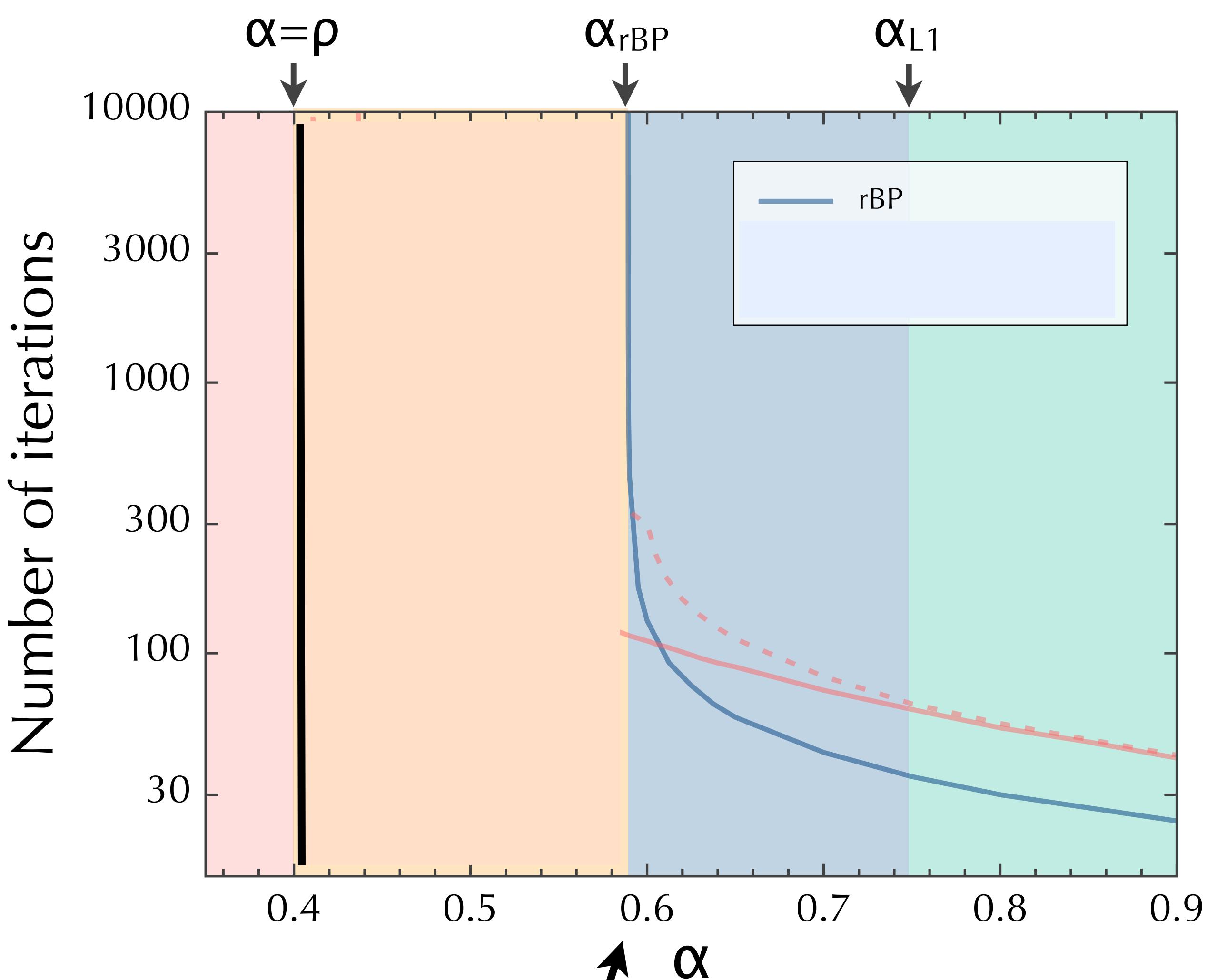
Free entropy

$\rho_0 = .4$

BP convergence time



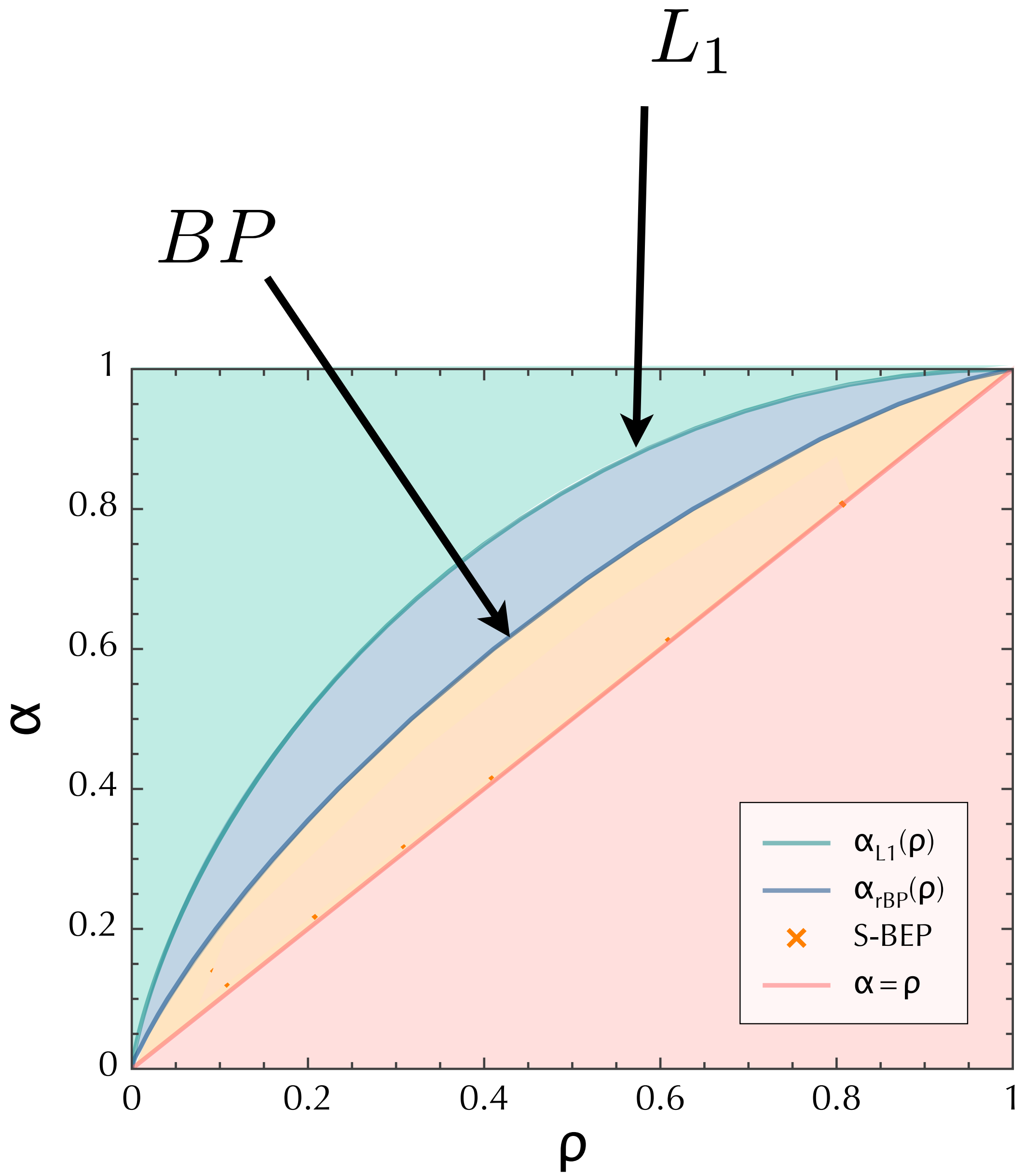
distance to native state



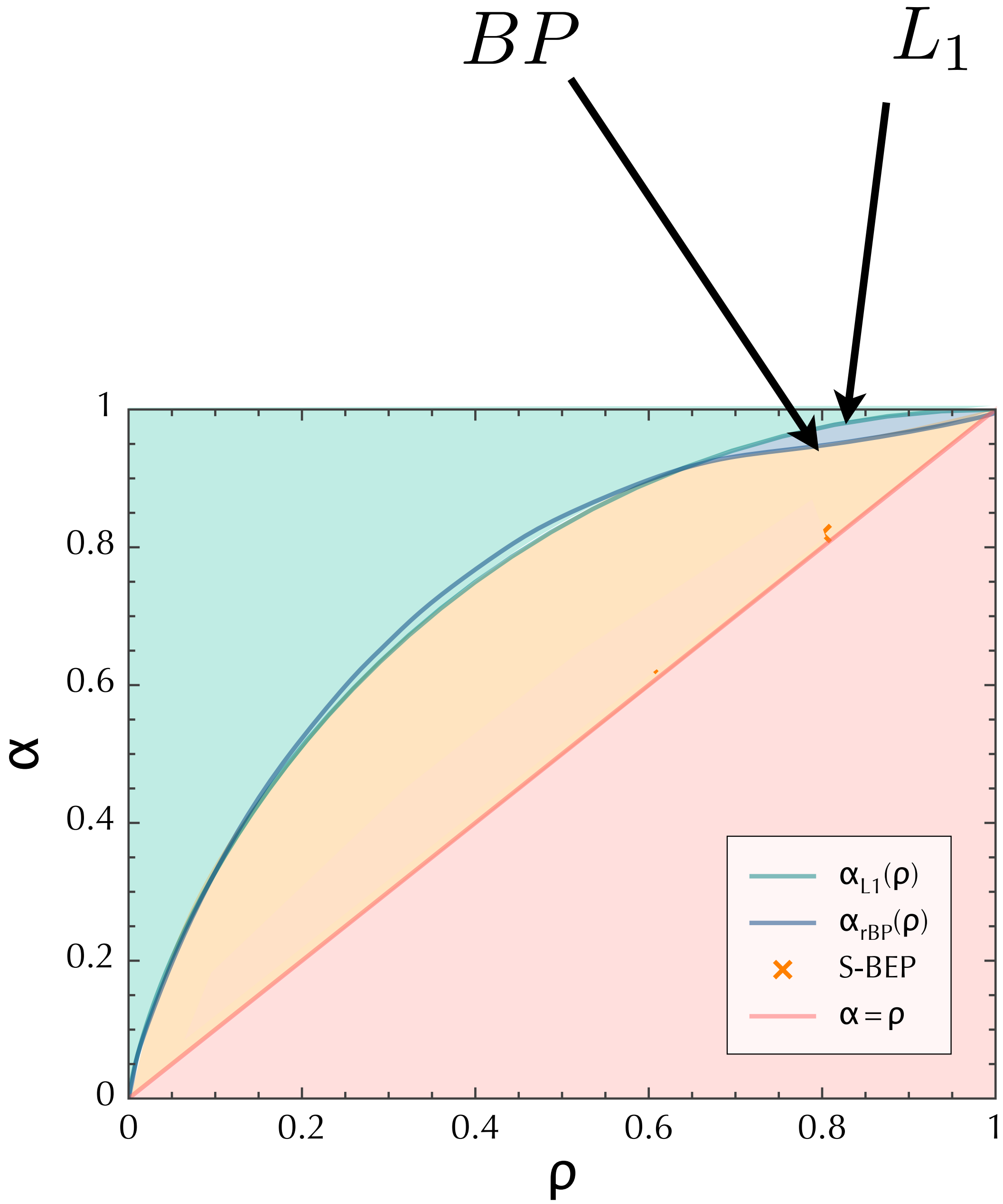
Dynamic glass transition

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

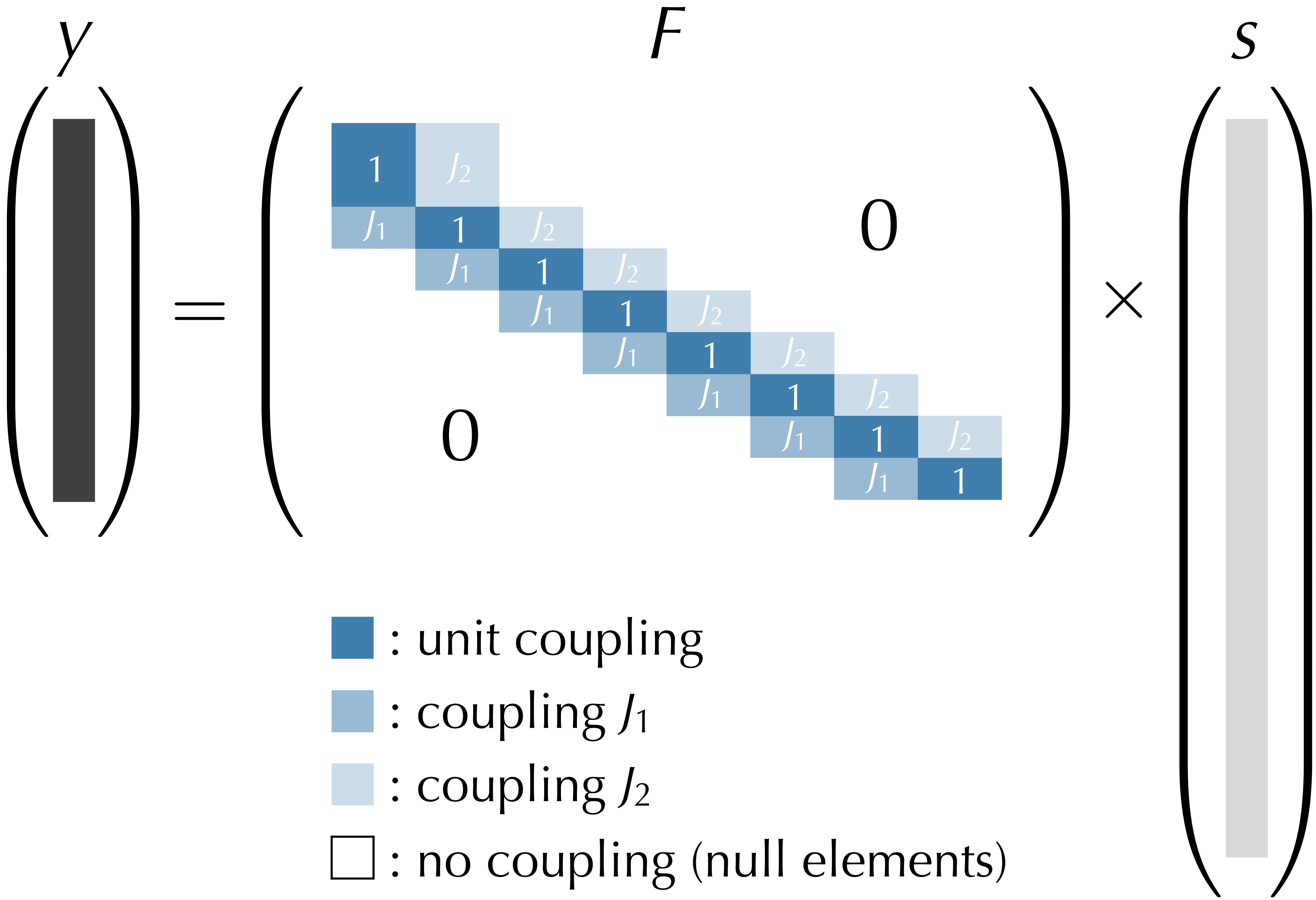
Getting around the glass trap: design the matrix F so that one nucleates the naive state (crystal nucleation idea, borrowed from error correcting codes!)

Hassani Macris Urbanke

$\longrightarrow$ Seeded BP

Group the variables and the measurements into L blocks

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



$L = 8$

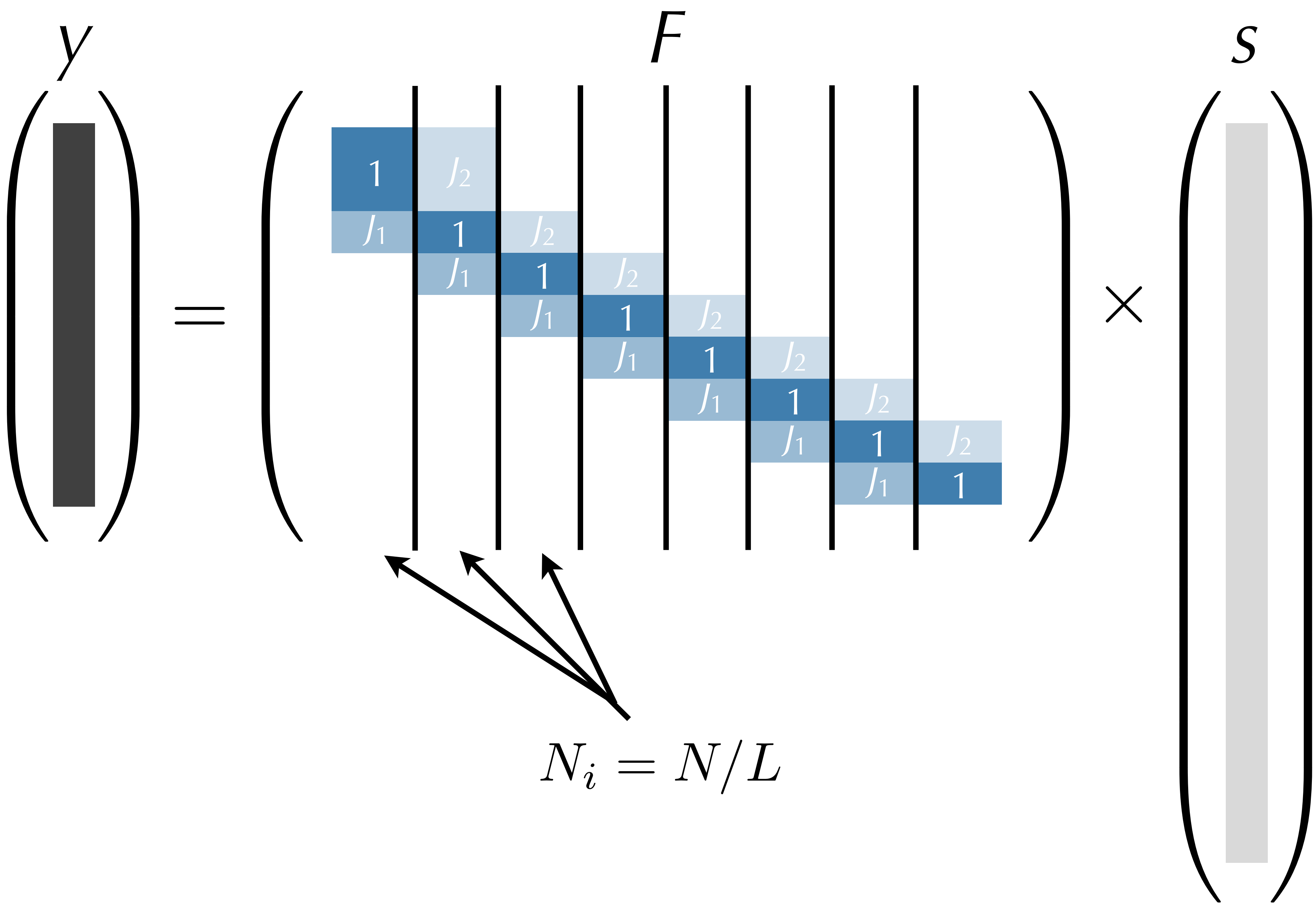
$N_i = N/L$

$M_i = \alpha_i N/L$

$\alpha_1 > \alpha_{BP}$

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

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



$L = 8$

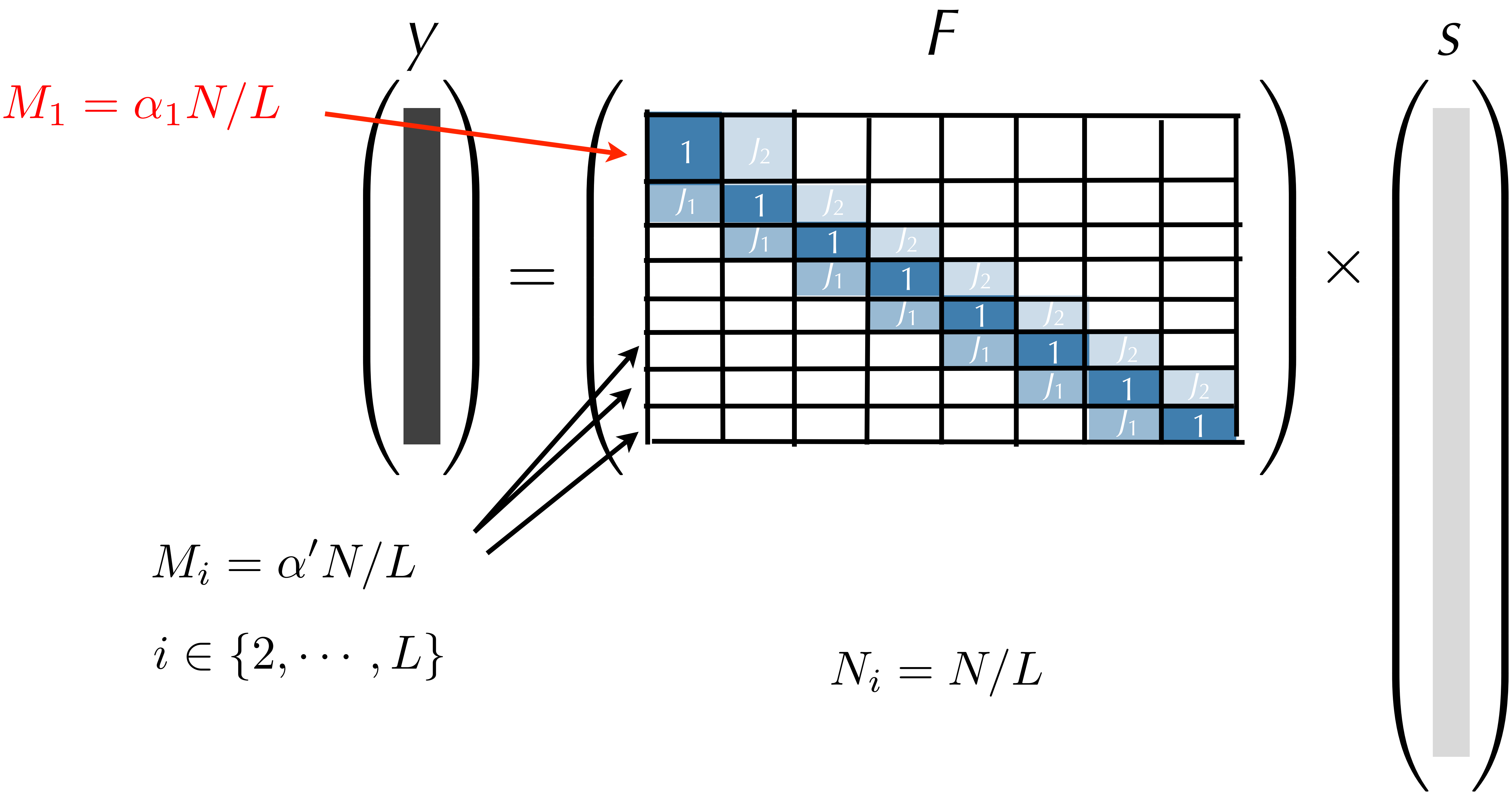
$N_i = N/L$

$M_i = \alpha_i N/L$

$\alpha_1 > \alpha_{BP}$

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

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



$L = 8$

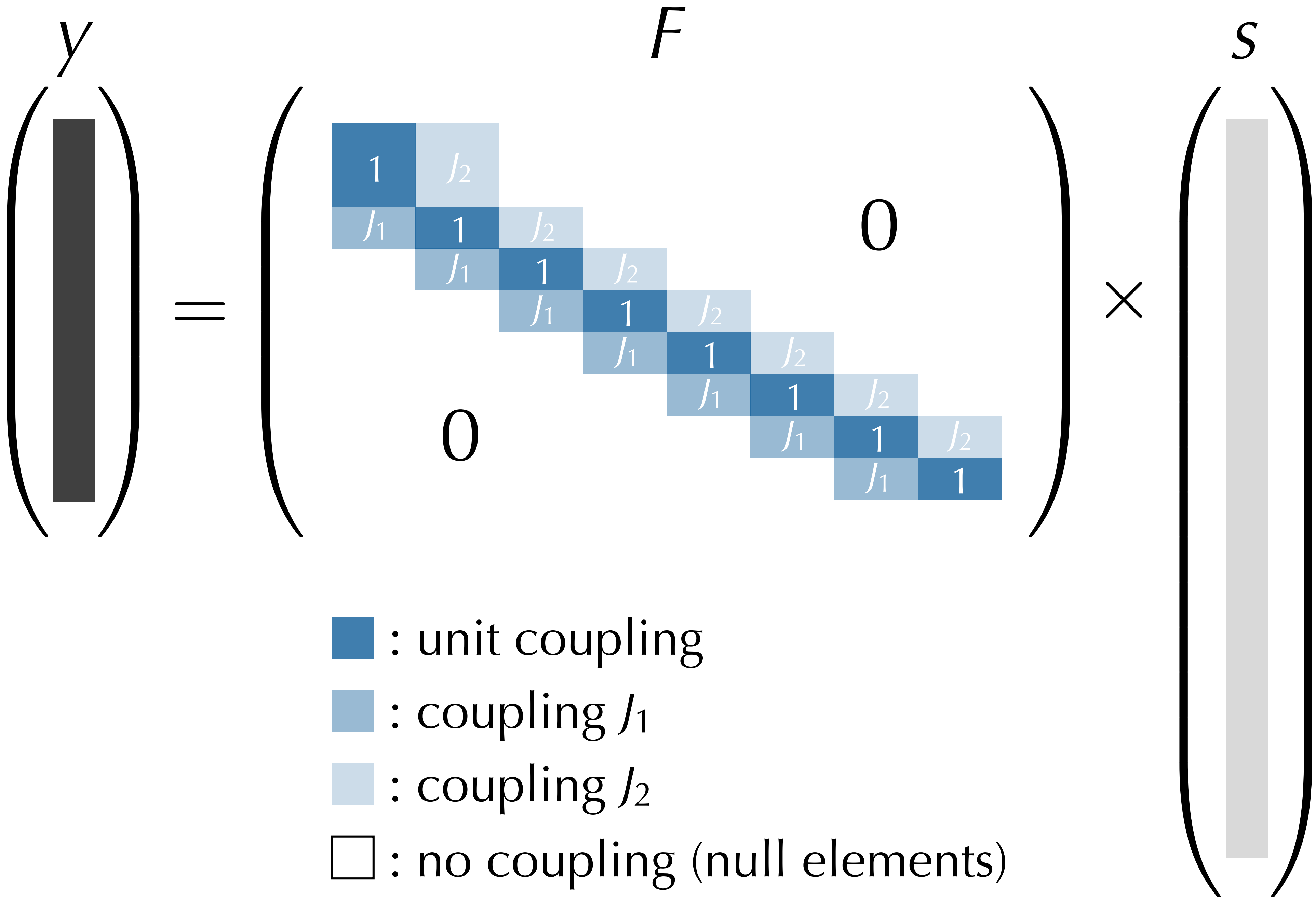
$N_i = N/L$

$M_i = \alpha_i N/L$

$\alpha_1 > \alpha_{BP}$

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

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



$$L = 8$$

$$N_i = N/L$$

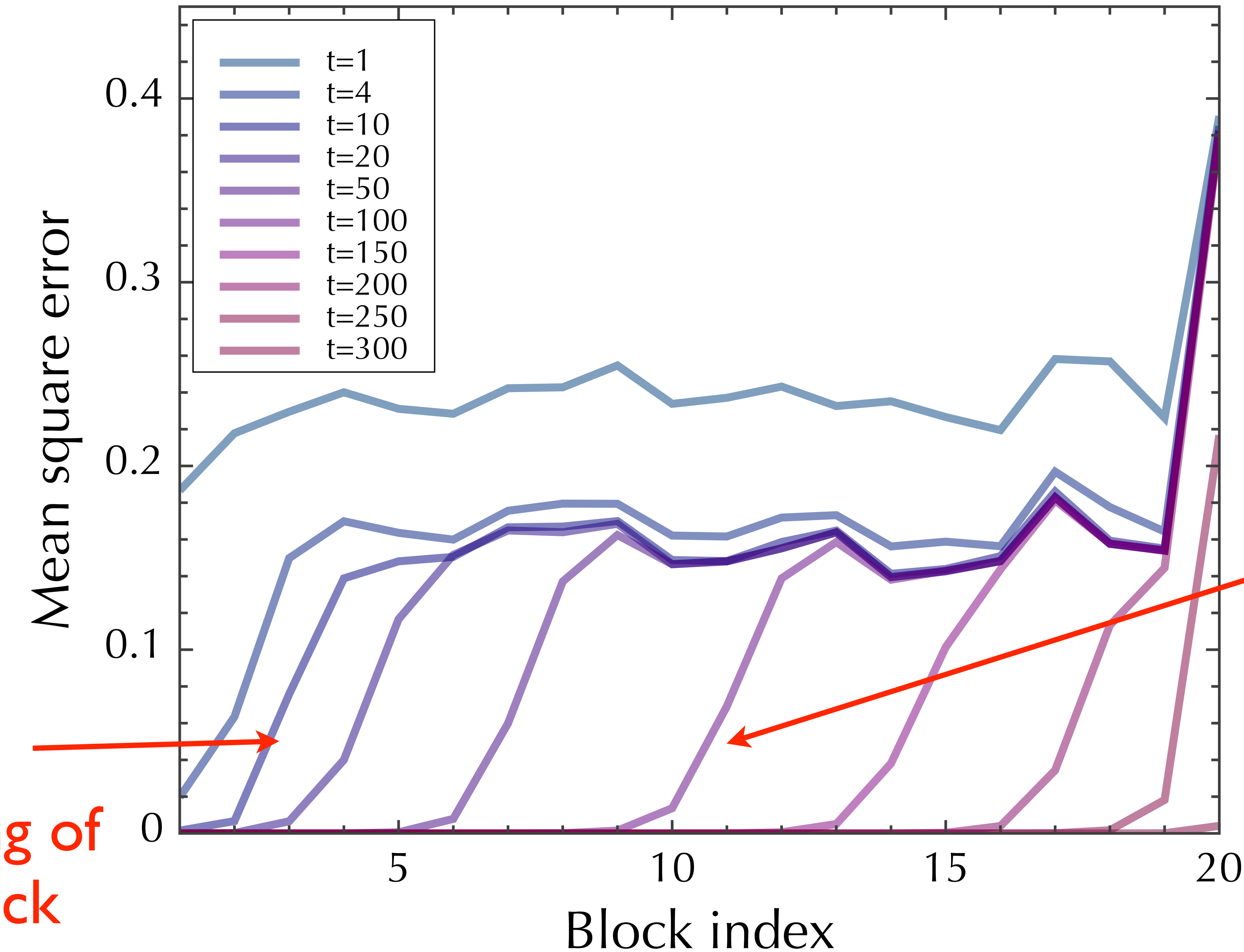
$$M_i = \alpha_i N/L$$

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

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

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

Numerical study



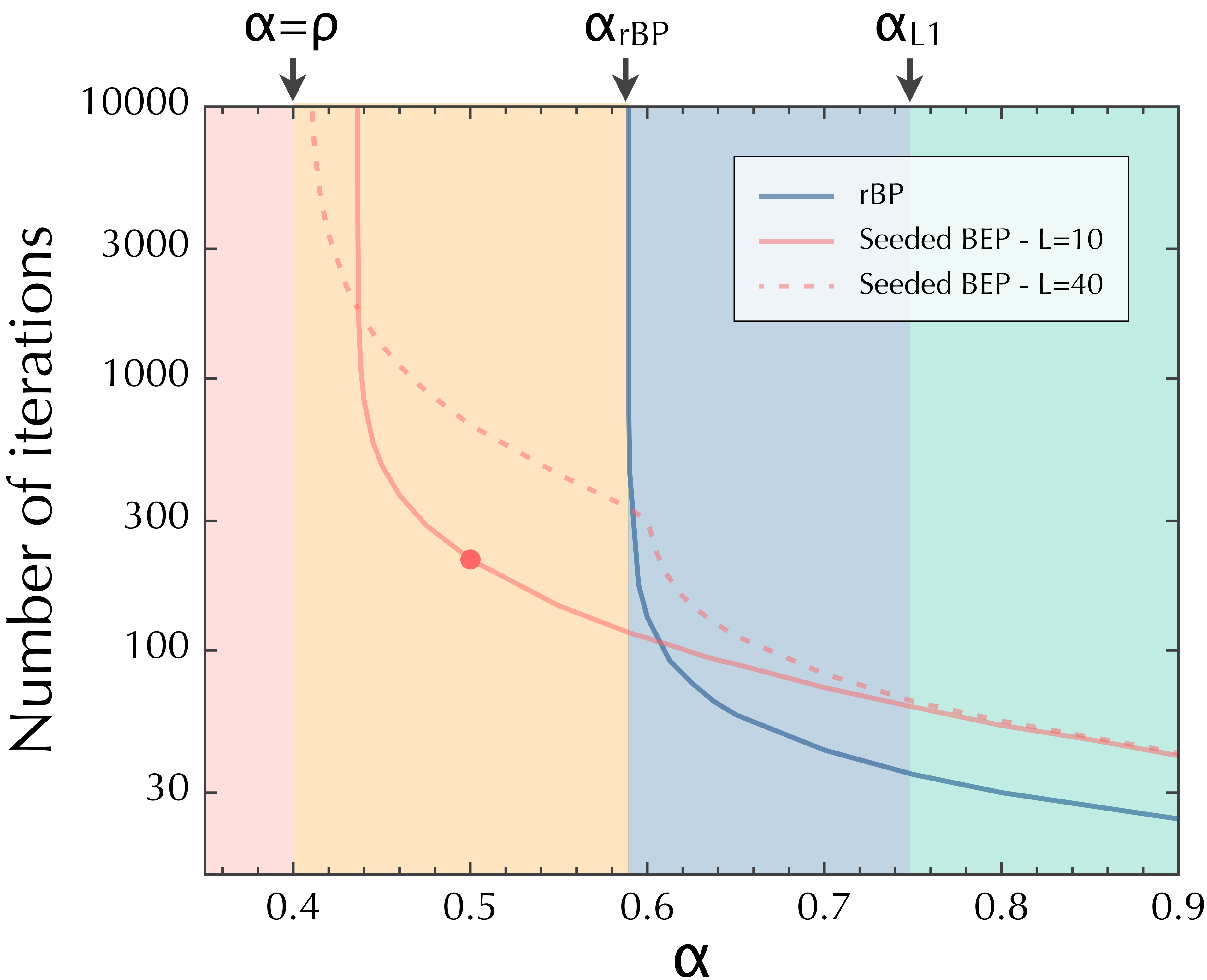
$t = 100$ decoding of blocks 1 to 9

$t = 10$ decoding of first block

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

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

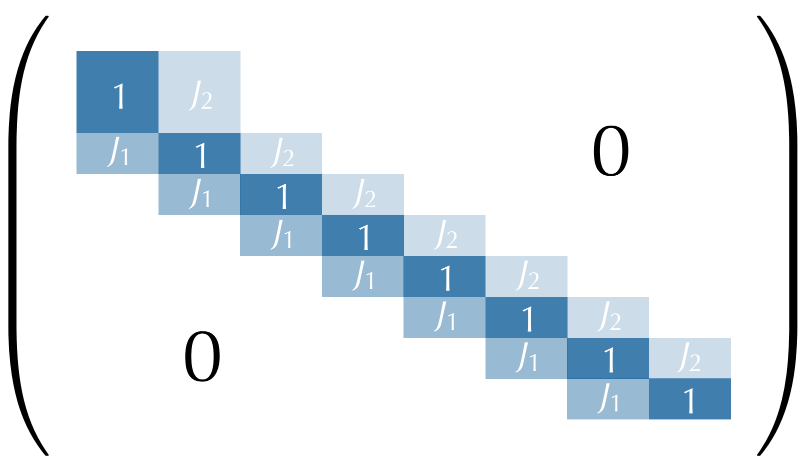
Numerical study



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

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

F



: unit coupling

: coupling J_1

: coupling J_2

: no coupling (null elements)

- 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 :

$$(\{D_r, V_r\}, \rho, \overline{x}, \sigma^2)^{(t+1)} = f\left((\{D_r, V_r\}\rho, \overline{x}, \sigma^2)^{(t)}\right)$$

$\rightarrow$ optimize J_1, J_2

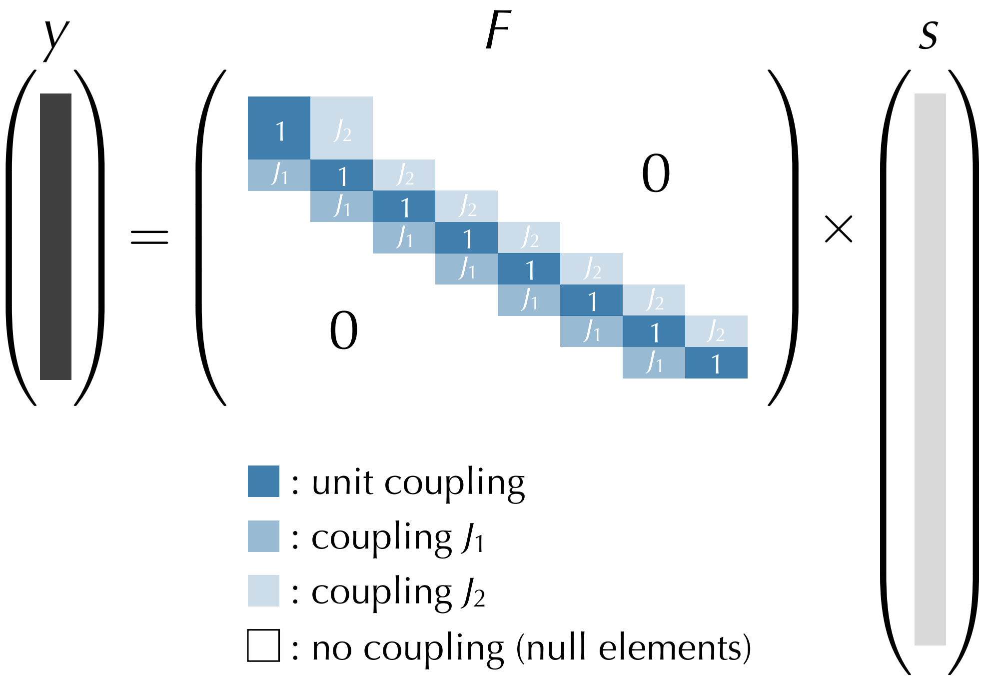
Known mapping f, depends on α_i, J_1, J_2

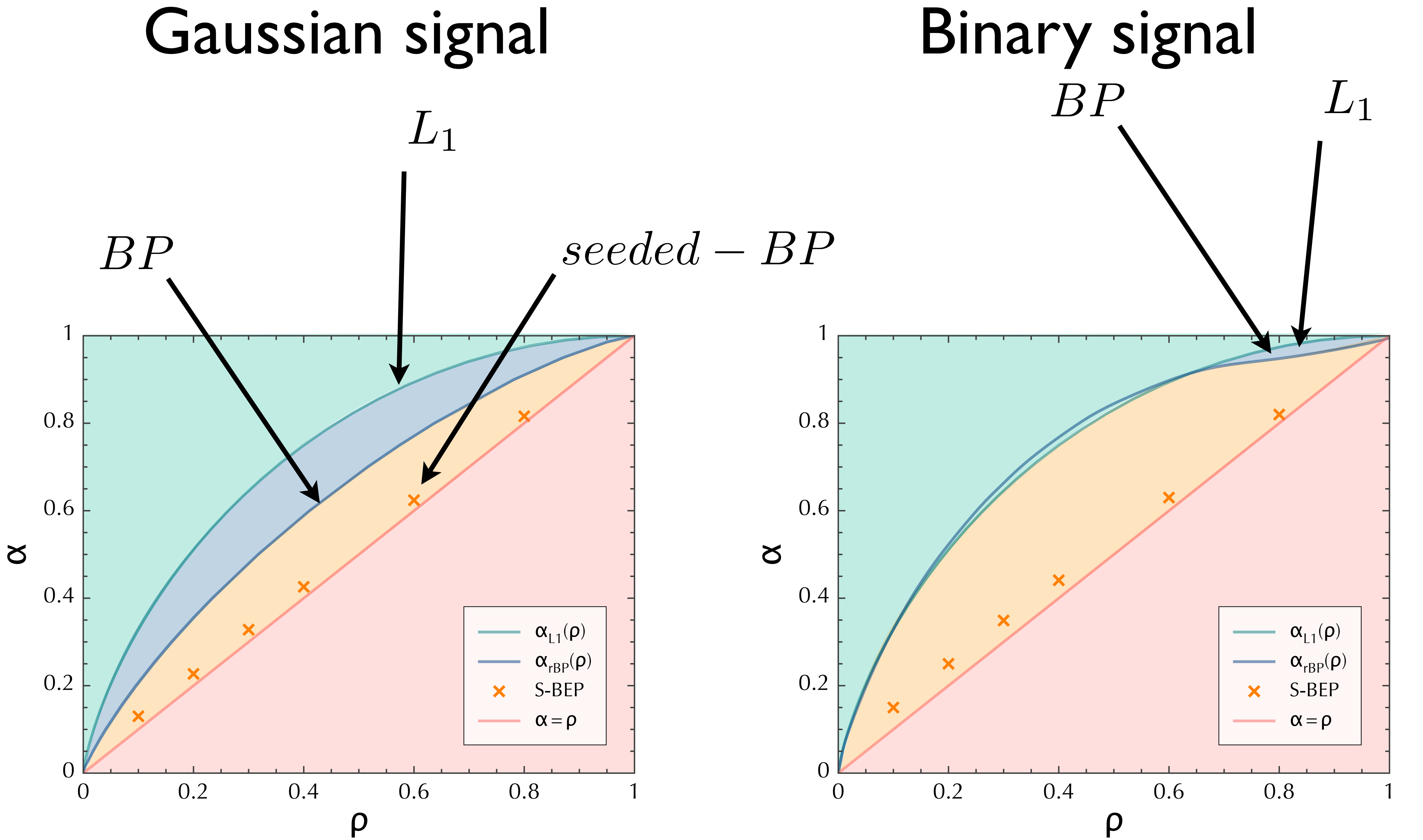
Analytic study: cavity equations, density evolution, replicas

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 $\alpha = \rho$

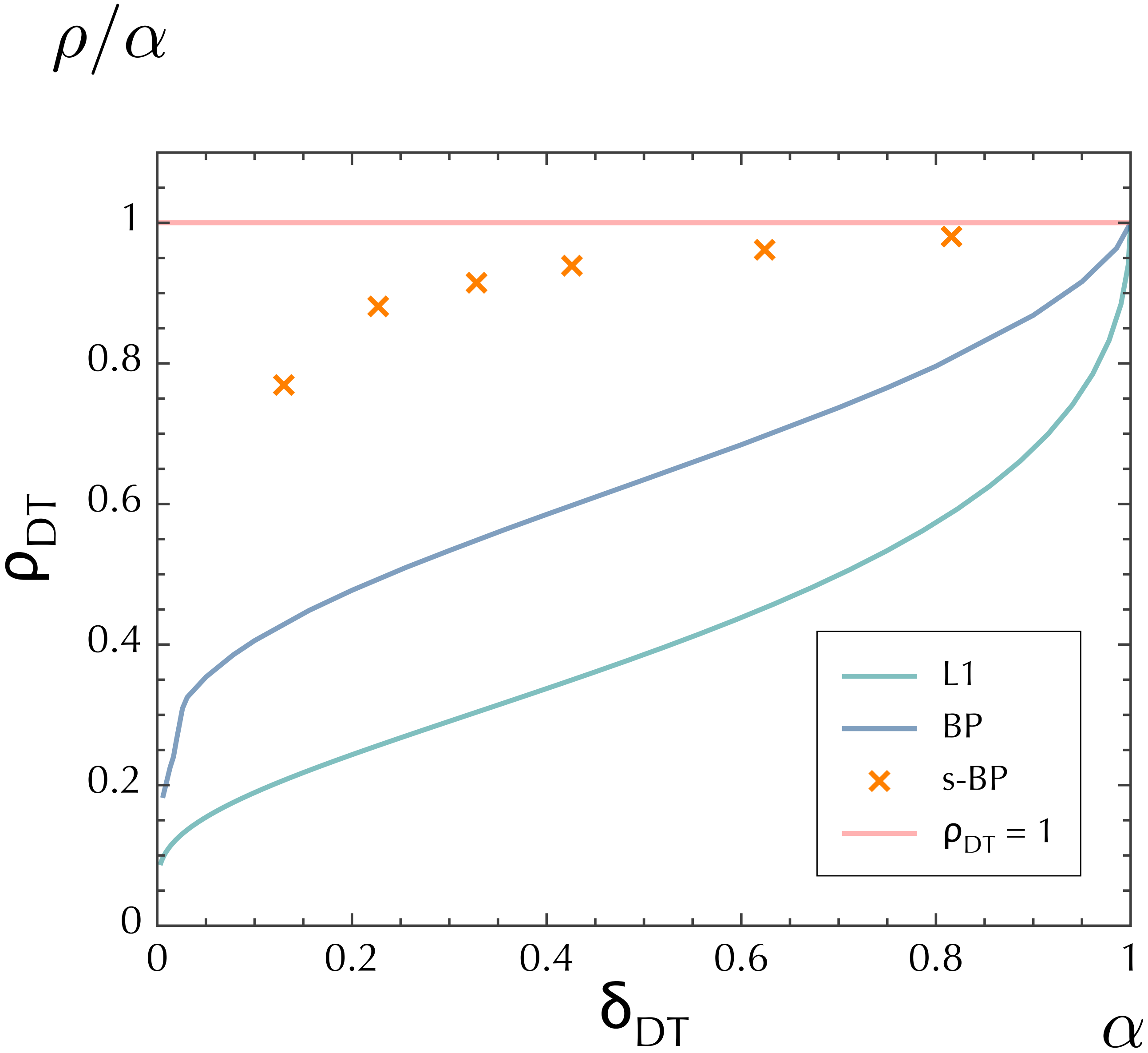




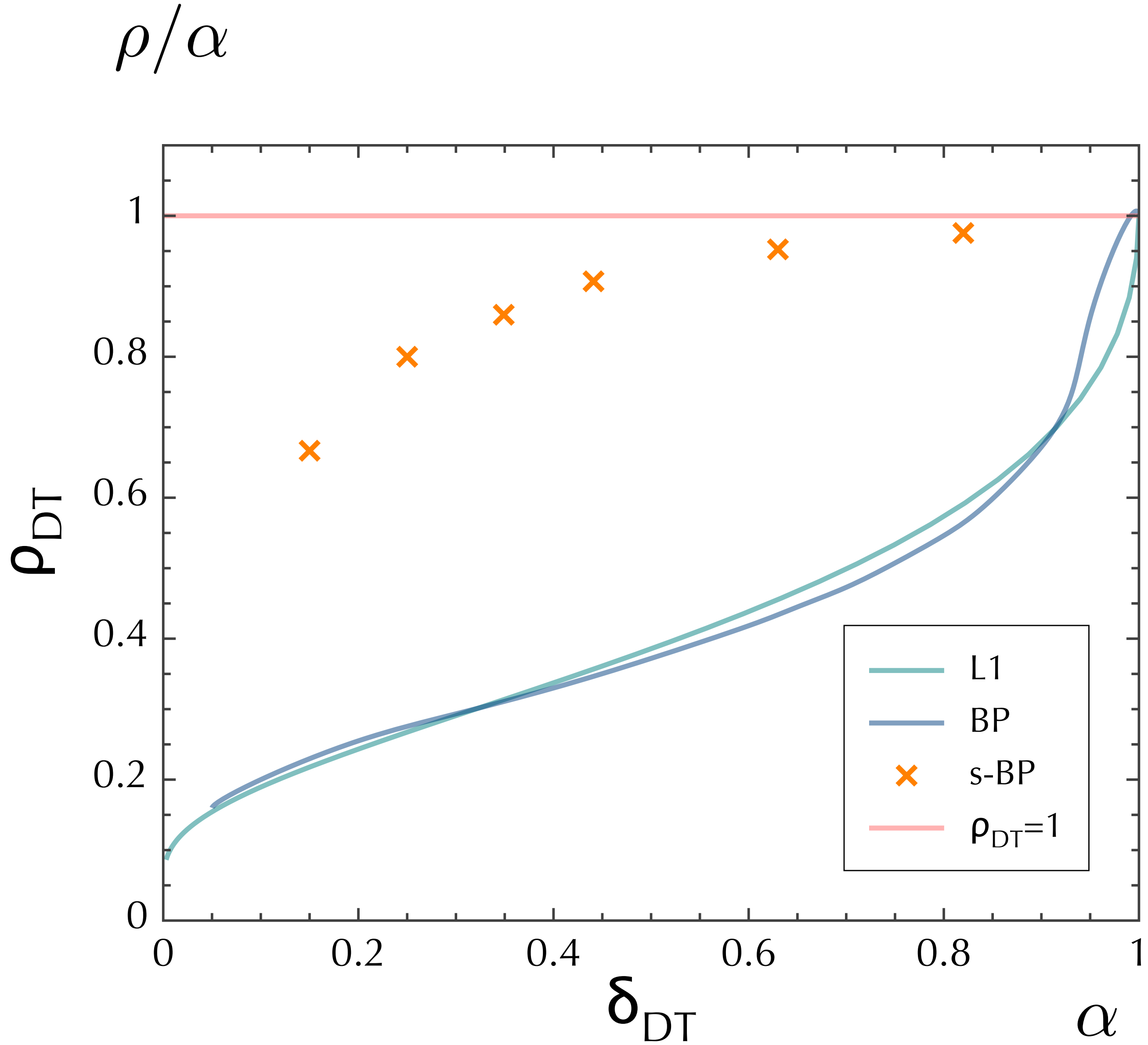
Theory: seeded-BP threshold at $\alpha=\rho$ when $L\to\infty$

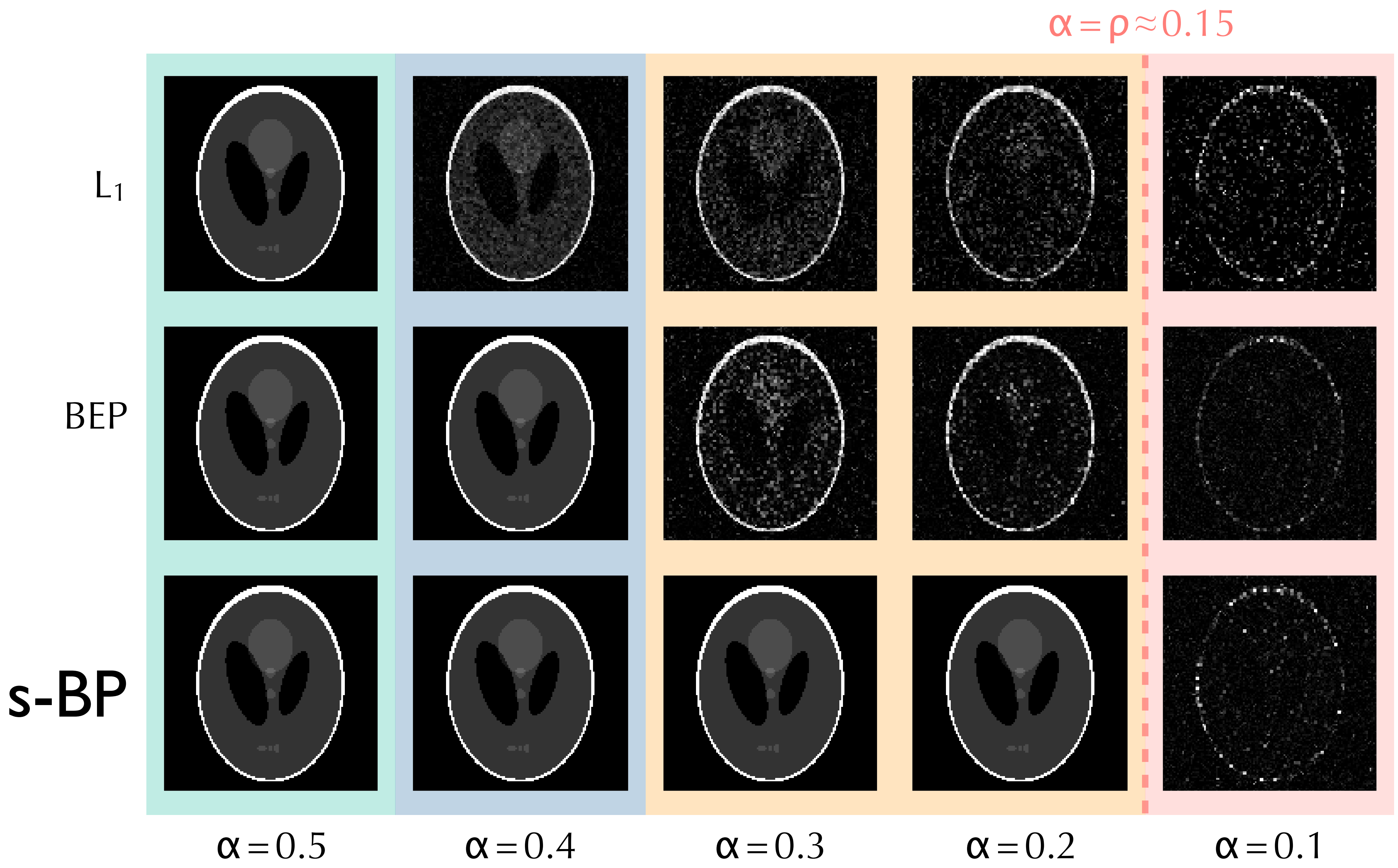
L_1 phase transition line moves up when using seeding F

Gaussian signal

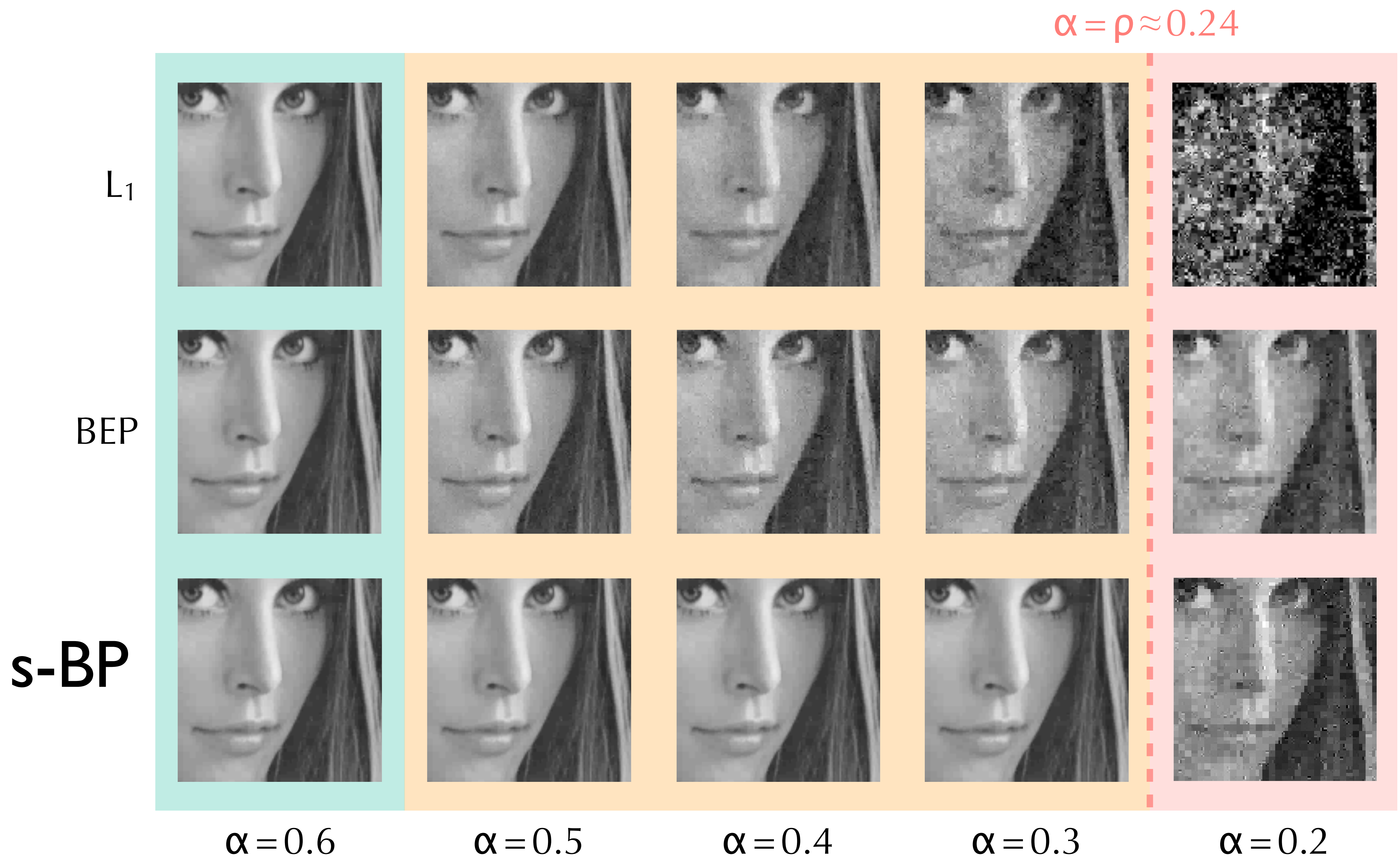


Binary signal





Shepp-Logan phantom, in the Haar-wavelet representation

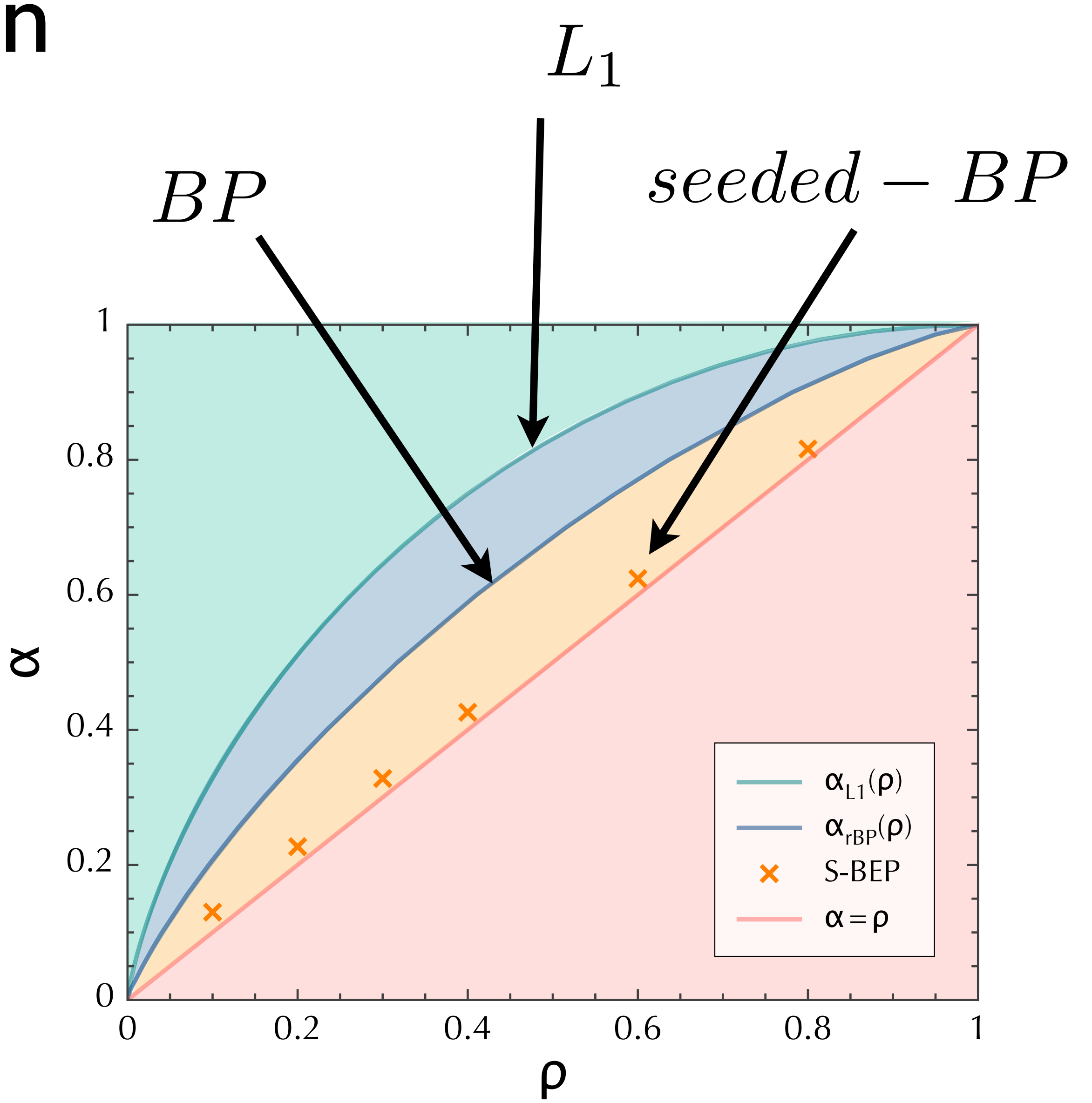


$\alpha = \rho \approx 0.24$
L_1
BEP
s-BP
$\alpha = 0.6$
$\alpha = 0.5$
$\alpha = 0.4$
$\alpha = 0.3$
$\alpha = 0.2$

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



Many things to be done

- Rigorous version of the analytic study (see Montanari's recent works)
- Rigorous study of analytic equations (choice of J_1, J_2, convergence, degradation of L_1 by seeding, etc.)
- Design of F for applications, taking into account constraints on the possible measurements
- Full study of the performance in presence of noise
- Non-linear versions
- ...