ECOLE SUPÉRIEURE DE PHYSIQUE ET DE CHIMIE INDUSTRIELLES

A statistical physics approach to compressed sensing

or y=Ax revisited

Florent Krzakala
ESPCI, & CNRS

in collaboration with
Jean Barbier (ESPCI), Emmanuelle Gouillart (Saint-Gobain),
Marc Mézard (ENS), François Sausset (LPTMS)
Yifan Sun (ESPCI) and Lenka Zdeborová (IPhT Saclay)

European Research Council
Established by the European Commission

http://krzakala.org

Compressed sensing

or y=Ax revisited

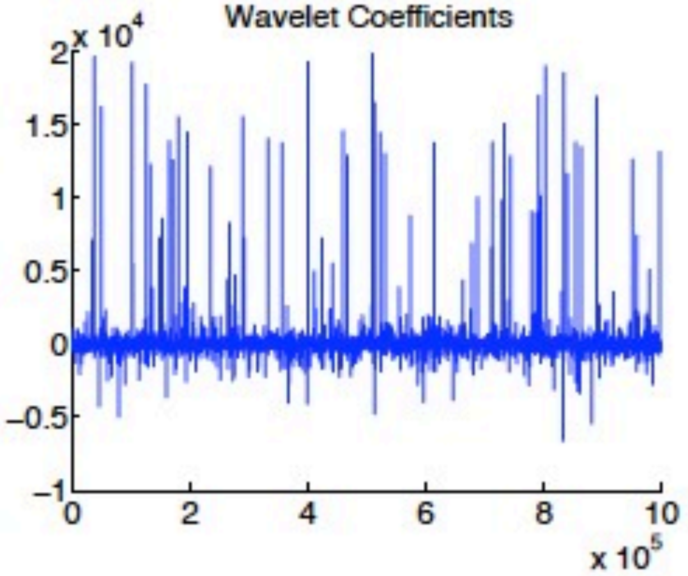
- What is compressed sensing?
- What is the link between statistical physics and compressed sensing?
- How can one use statistical physics to improve on compressed sensing technics?

Compressed sensing

or y=Ax revisited

- **What is compressed sensing?**
- What is the link between statistical physics and compressed sensing?
- How can one use statistical physics to improve on compressed sensing technics?

What is compressed sensing?

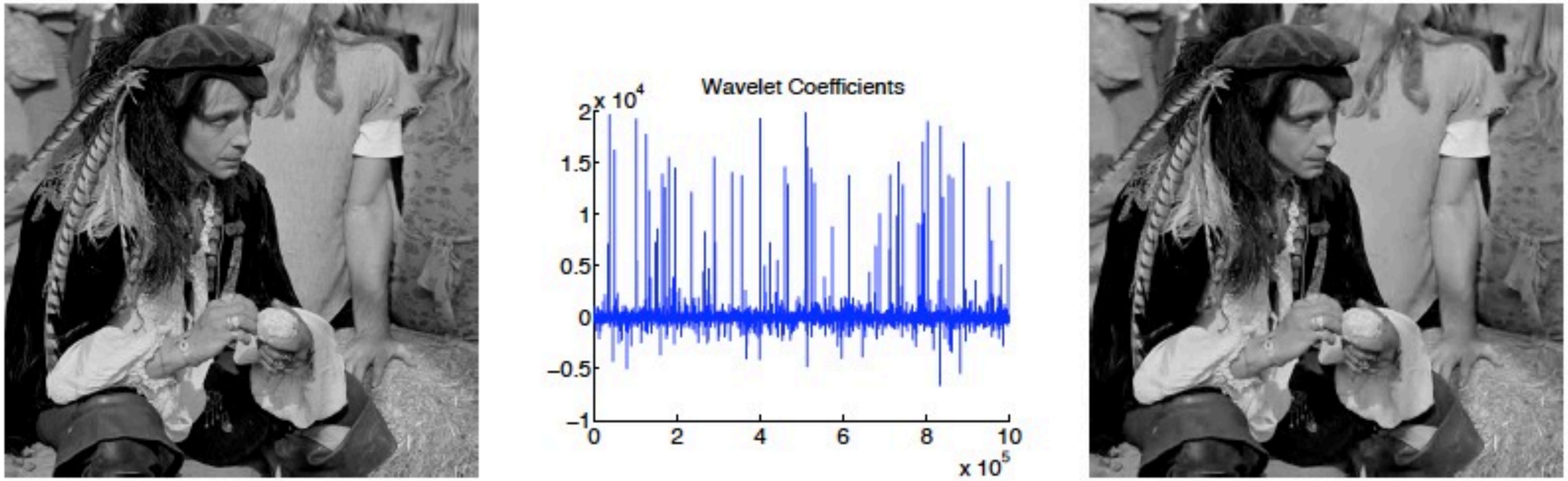


From 10^6 wavelet coefficients, keep 25.000

Most signal of interest are <u>sparse</u> in an appropriated basis

⇒Exploited for data compression (JPEG2000).

What is compressed sensing?



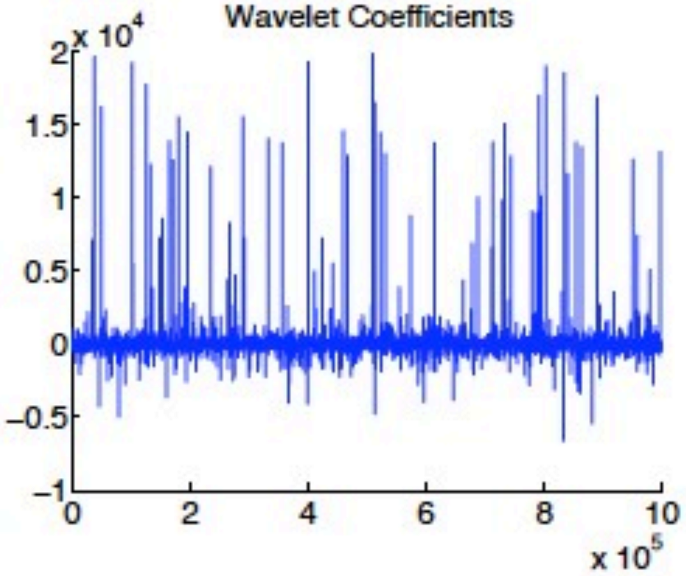
From 10^6 wavelet coefficients, keep 25.000

Most signal of interest are <u>sparse</u> in an appropriated basis
⇒Exploited for data compression (JPEG2000).

Why do we record a huge amount of data, and then keep only the important bits?

Couldn’t we record <u>only</u> the relevant information directly?

What is compressed sensing?



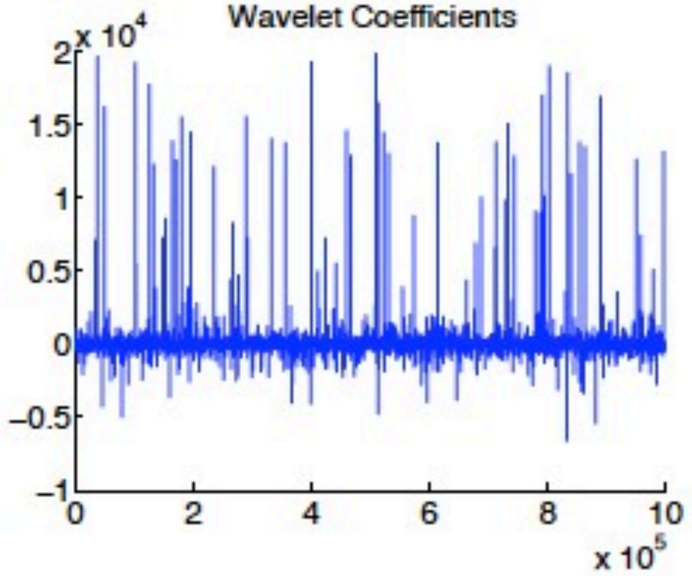
Why do we record a huge amount of data, and then keep only the important bits?

Couldn't we record only the relevant information directly?

Compressed Sensing

1) Record directly in compressed form (gain of time and storage)
2) Reconstruct the original signal afterwards

What is compressed sensing?



Why do we record a huge amount of data, and then keep

compressed sensing

● Rechercher sur le Web ○ Rechercher les pages en Français

date indifférente

inclure les citations

Conseil : Recherchez des résultats uniquement en **français**. Vous pouvez indiquer votre langue de recherche sur la page

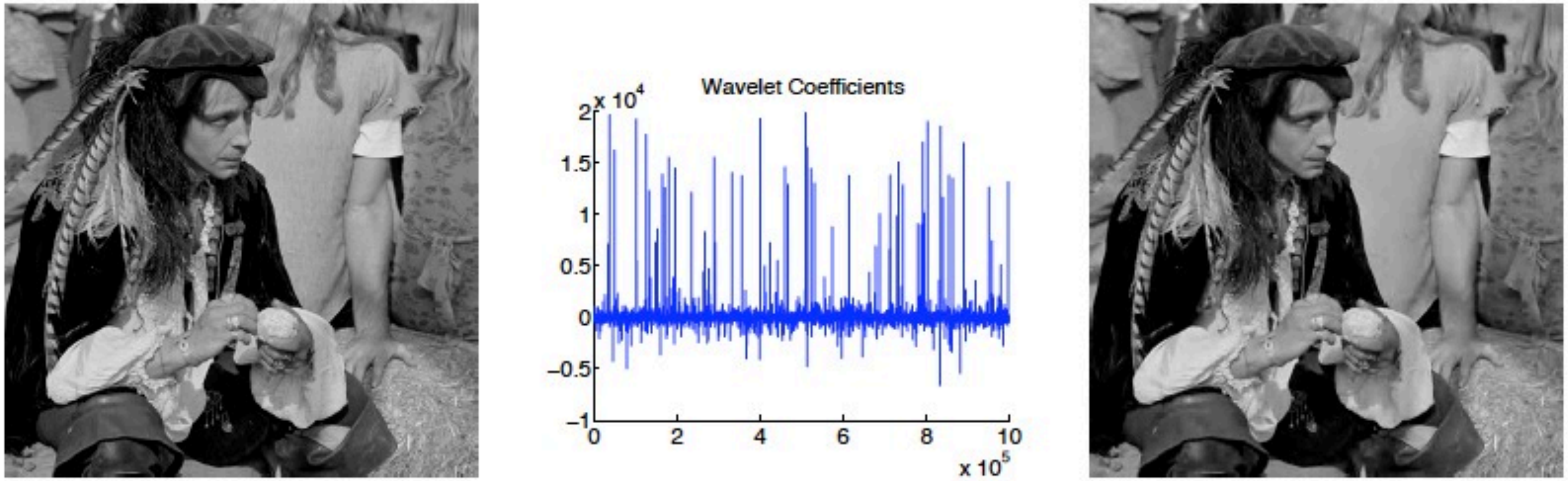
Compressed sensing

DL Donoho - Information Theory, IEEE Transactions on, 2006 - ieeexplore.ieee.org

Abstract Suppose x is an unknown vector in Ropf m (a digital image or signal); we plan to measure n general linear functionals of x and then reconstruct. If x is known to be compressible by transform coding with a known transform, and we reconstruct via the ...

Cité 4384 fois - Autres articles - findit@espci - Les 46 versions

What is compressed sensing?



Why do we record a huge amount of data, and then keep

compressed sensing

● Rechercher sur le Web ○ Rechercher les pages en Français

Conseil : Recherchez des résultats uniquement en **français**. Vous pouvez indiquer votre langue de recherche sur la page

Compressed sensing

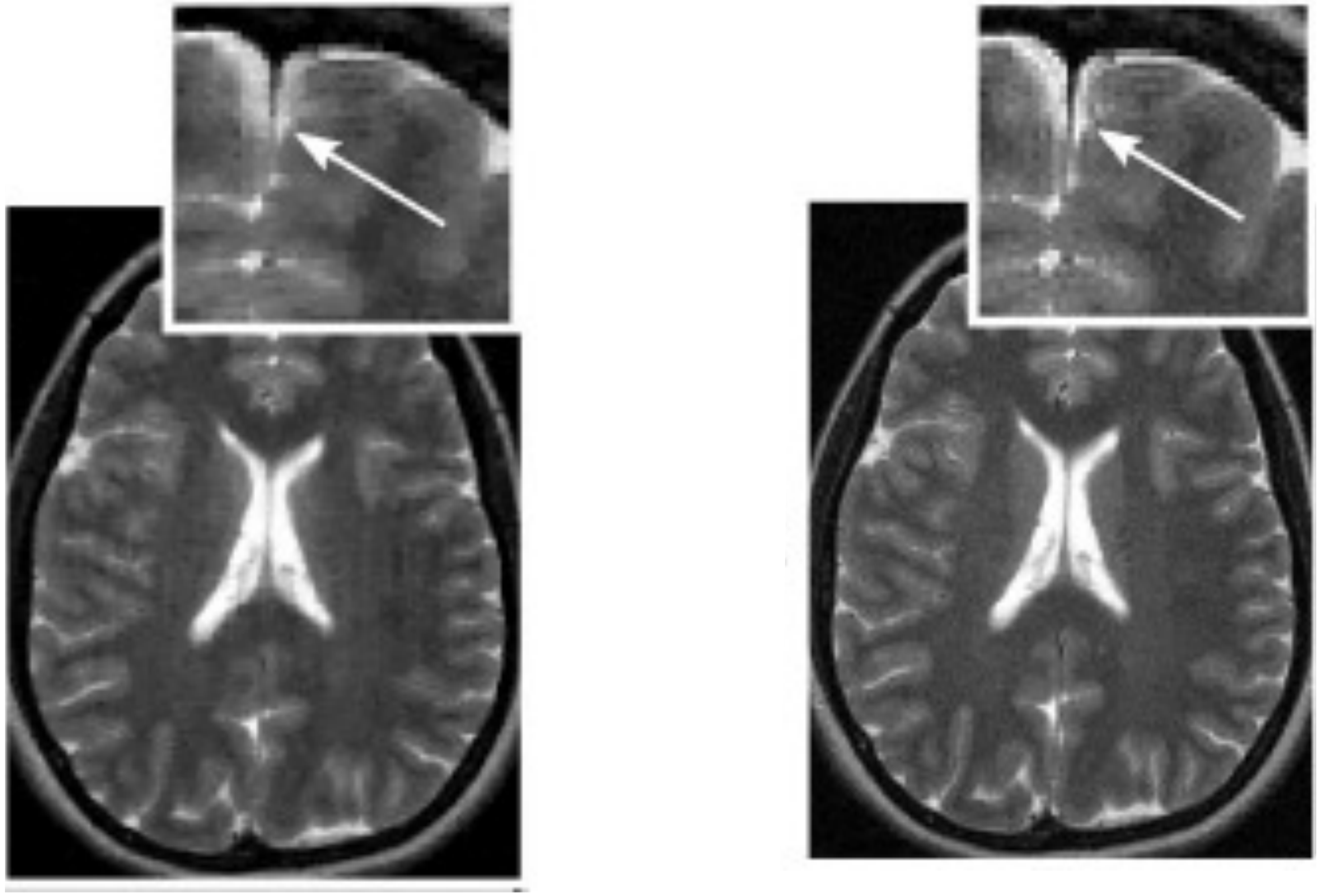
DL Donoho - Information Theory, IEEE Transactions on, 2006 - ieeexplore.ieee.org

Abstract Suppose x is an unknown vector in Ropf m (a digital image or signal); we plan to measure n general linear functionals of x and then reconstruct. If x is known to be compressible by transform coding with a known transform, and we reconstruct via the ...

Cité 4384 fois - Autres articles - findit@espci - Les 46 versions

Teaser:

An example from magnetic resonance imaging



Left: image acquired with CS

Acceleration by a factor 2.5

Lustig, Donoho, Pauly ’07

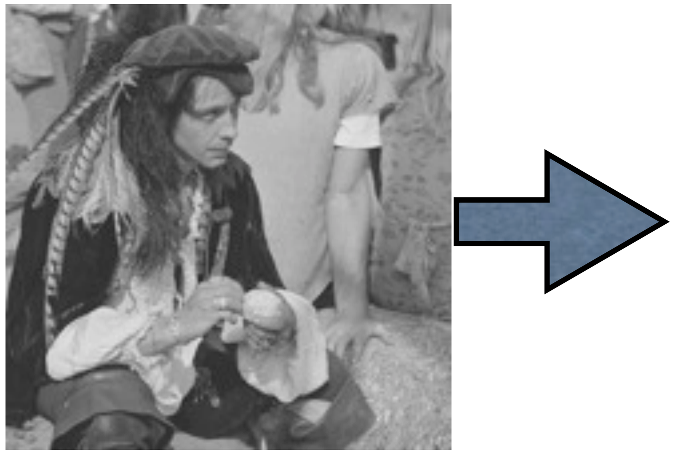
Possible applications

- Rapid Magnetic Resonance Imaging
- Image acquisition (single-pixel camera)
- DNA microarrays
- Group testing
- Fast data compression
- Herschel spacial telescope
- Compressed Sensing Microscopes
- Sparse Principal Component Analysis
- Compressed quantum state tomography
- ...

How does compressed sensing work?

How does compressed sensing work?

Image I



nxn pixels

vector of size N=n×n

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

How does compressed sensing work?

M measurements

=

M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

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

How does compressed sensing work?

M measurements

=

M linear operations on the vector

Image I

nxn pixels

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

vector of size N=n×n

$$\vec{y} = G\vec{I}$$

G=M×N matrix

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

vector of size M

How does compressed sensing work?

M measurements

=

M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

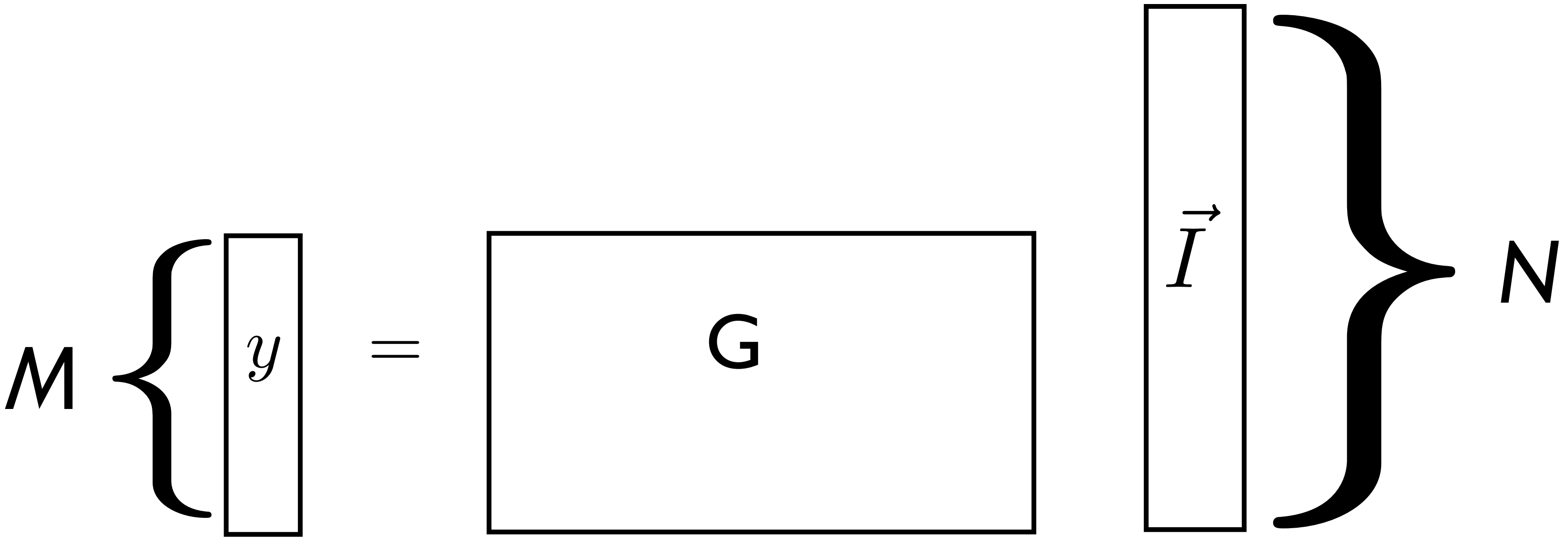
$$\vec{I} = \begin{pmatrix} I^1 \\ . \\ . \\ . \\ . \\ I^N \end{pmatrix}$$

$$\vec{y} = G\vec{I}$$

G=M×N matrix

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

vector of size M



M×N matrix

How does compressed sensing work?

M measurements
=
M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

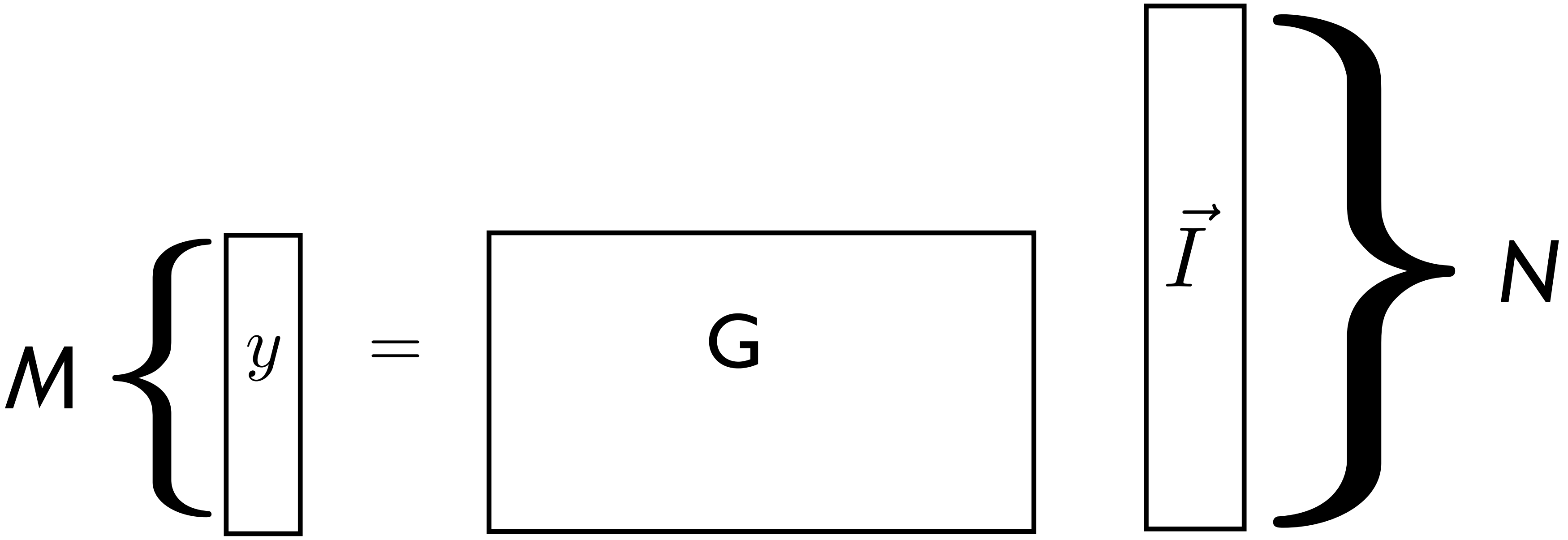
$$\vec{I} = \begin{pmatrix} I^1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ I^N \end{pmatrix}$$

$$\vec{y} = G\vec{I}$$

G=M×N matrix

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

vector of size M



$$M\left\{ \boxed{y} \right. = \boxed{\;G\;} \; \left. \boxed{\vec{I}} \right\} N$$

M×N matrix

How does compressed sensing work?

M measurements

=

M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

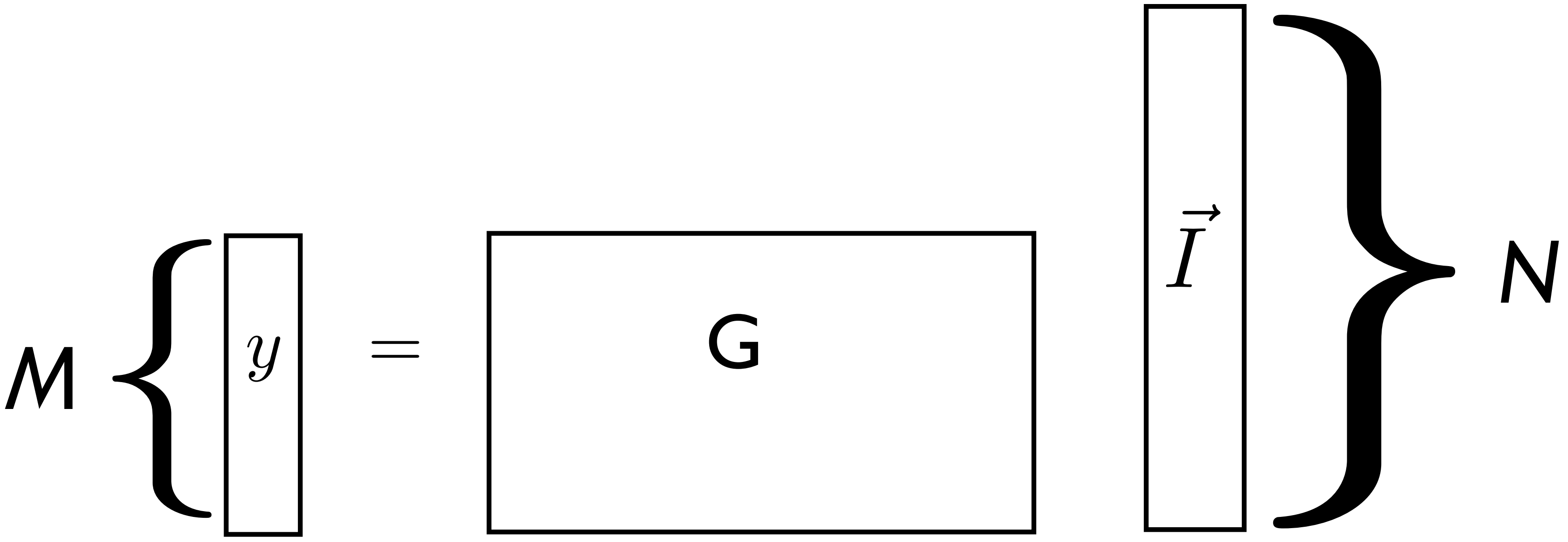
$$\vec{I} = \begin{pmatrix} I^1 \\ . \\ . \\ . \\ . \\ I^N \end{pmatrix}$$

$$\vec{y} = G\vec{I}$$

G=M×N matrix

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

vector of size M



M×N matrix

Problem: you know y and G, how to reconstruct I ?

How does compressed sensing work?

M measurements

=

M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

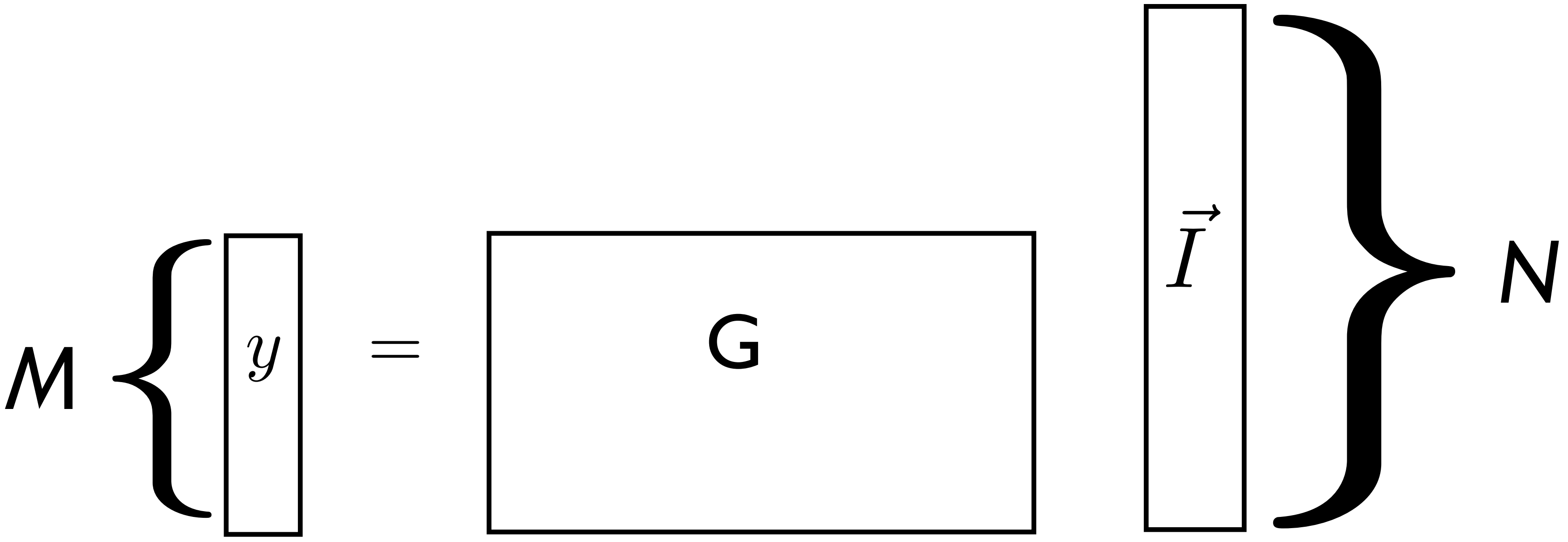
$$\vec{I} = \begin{pmatrix} I^1 \\ . \\ . \\ . \\ . \\ I^N \end{pmatrix}$$

$$\vec{y} = G\vec{I}$$

G=M×N matrix

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

vector of size M



$$M\left\{\boxed{y}\right. = \boxed{G}\;\left.\boxed{\vec{I}}\right\}N$$

M×N matrix

Problem: you know y and G, how to reconstruct I ?

If M=N ☞ easy, just use: $I = G^{-1}y$

How does compressed sensing work?

M measurements
=
M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

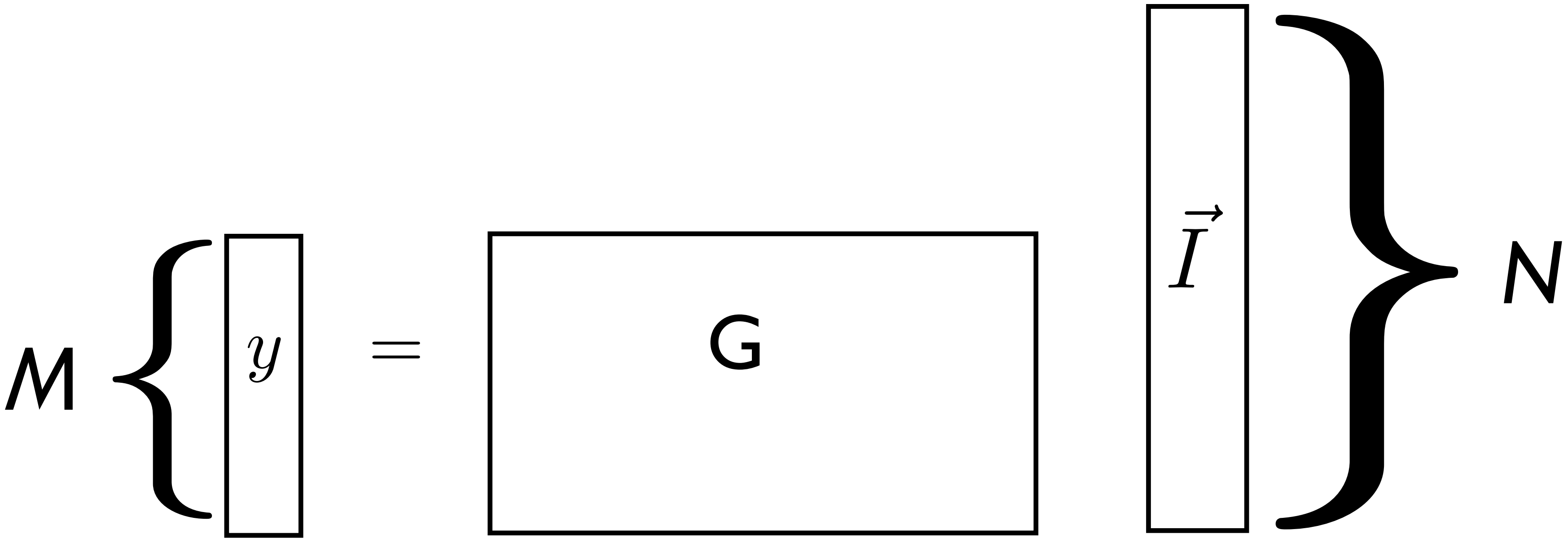
$$\vec{I} = \begin{pmatrix} I^1 \\ . \\ . \\ . \\ . \\ I^N \end{pmatrix}$$

$$\vec{y} = G\vec{I}$$

G=M×N matrix

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

vector of size M



$$M \left\{ \boxed{y} \right. = \boxed{\quad G \quad} \left. \boxed{\vec{I}} \right\} N$$

M×N matrix

Problem: you know y and G, how to reconstruct I ?

If M<N ☞ under-constrained system of equations

How does compressed sensing work?

M measurements
=
M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

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

$$\vec{y} = G\vec{I}$$

G=M×N matrix

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

vector of size M

M×N matrix

How does compressed sensing work?

M measurements

=

M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

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

$$\vec{y} = G\vec{I}$$

G=M×N matrix

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

vector of size M

M×N matrix

Compressed sensing input:

The signal is sparse in an appropriate basis

How does compressed sensing work?

M measurements
=
M linear operations on the vector

Image I

nxn pixels

vector of size N=n×n

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

$$\vec{y} = G\vec{I}$$

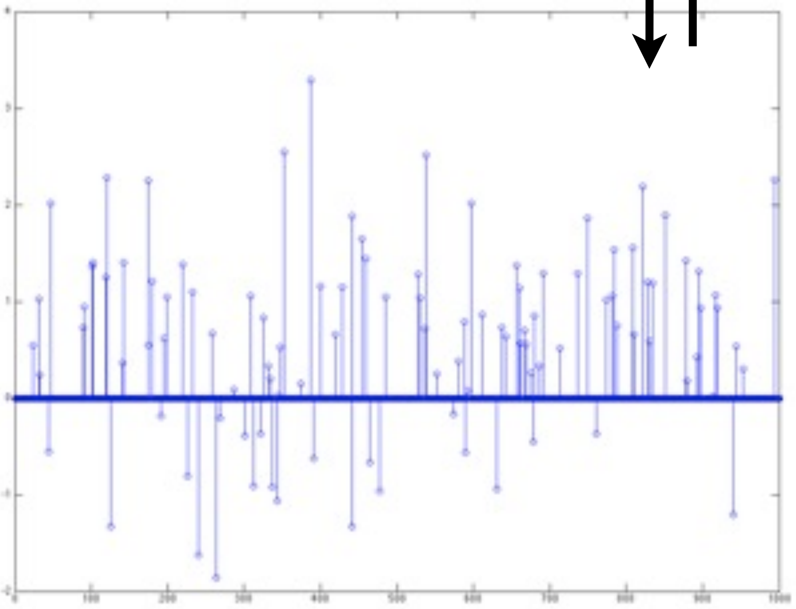
G=M×N matrix

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

vector of size M

N×N matrix
Direct and inverse
Wavelet transforms

ψ^{-1} ψ



How does compressed sensing work?

M measurements

=

M linear operations on the vector

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

vector of size M

$$\vec{y} = G\vec{I}$$

G=M×N matrix

Image I

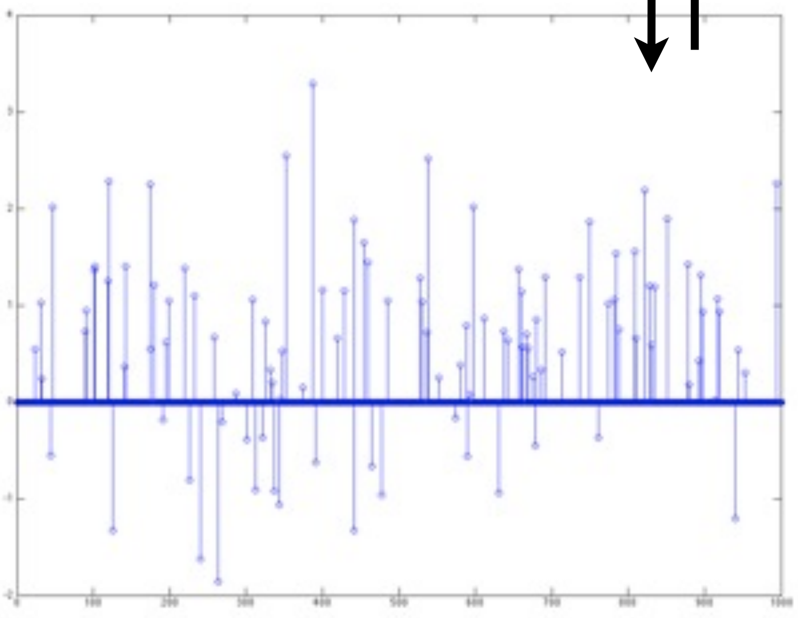
nxn pixels

vector of size N=n×n

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

N×N matrix
Direct and inverse
Wavelet transforms

ψ^{-1} ψ



Sparse vector
of size N=n×n

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

How does compressed sensing work?

M measurements

=

M linear operations on the vector

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

vector of size M

$$\vec{y} = G\vec{I}$$

G=M×N matrix

Image I

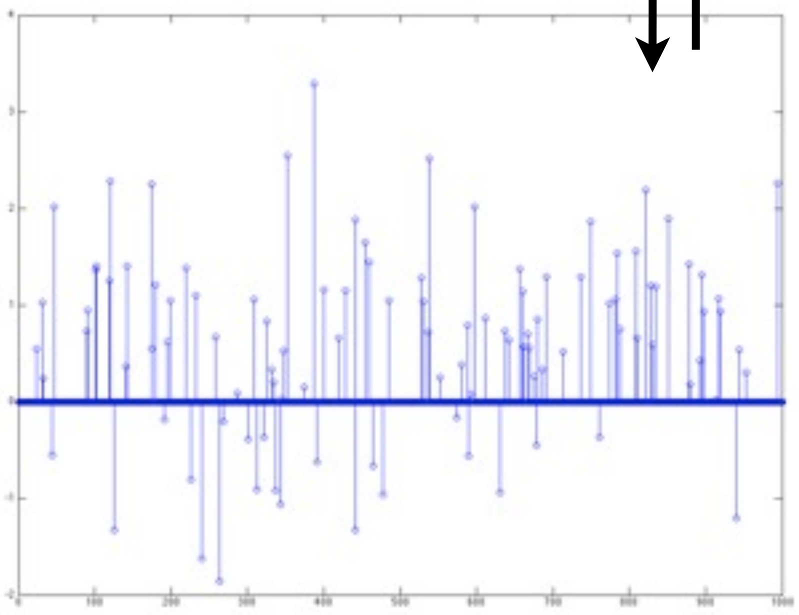
nxn pixels

vector of size N=n×n

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

N×N matrix
Direct and inverse
Wavelet transforms

ψ^{-1} ψ



Sparse vector
of size N=n×n

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

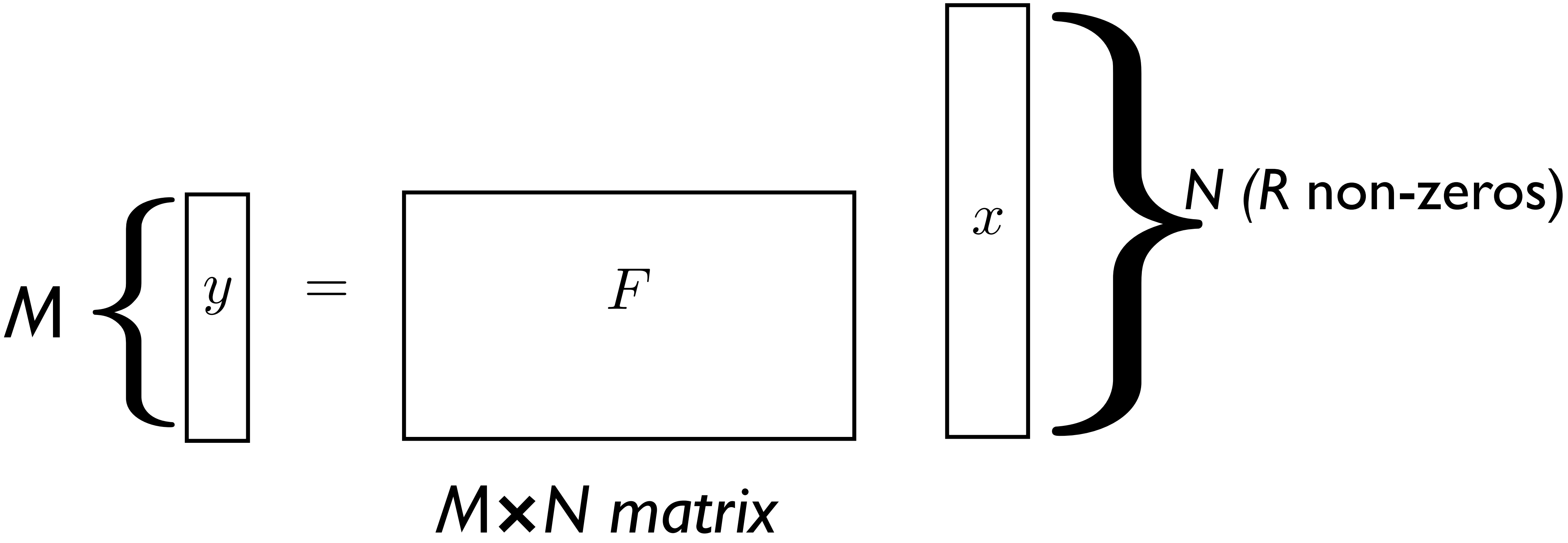
The problem to solve is now

$$\vec{y} = F\vec{x}$$

with $F = G\psi$

F=M×N matrix

How does compressed sensing work?



The problem to solve is now

$$\vec{y} = F\vec{x}$$

with $F = G\psi$

F=M×N matrix

- Need to find a sparse solution of an under-constrained set of linear equations
- Ideally works as long as M>R
- Robust to noise

The reconstruction problem:

Inverting an underconstrained linear system

Consider a system of linear measurements

$$y = Fx$$

Measurements

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

Signal

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

$F = M \times N$ matrix

The problem: $y = Fx$, find x

Generically: ● if $M = N$

Unique solution obtained by inversion $x = F^{-1}y$

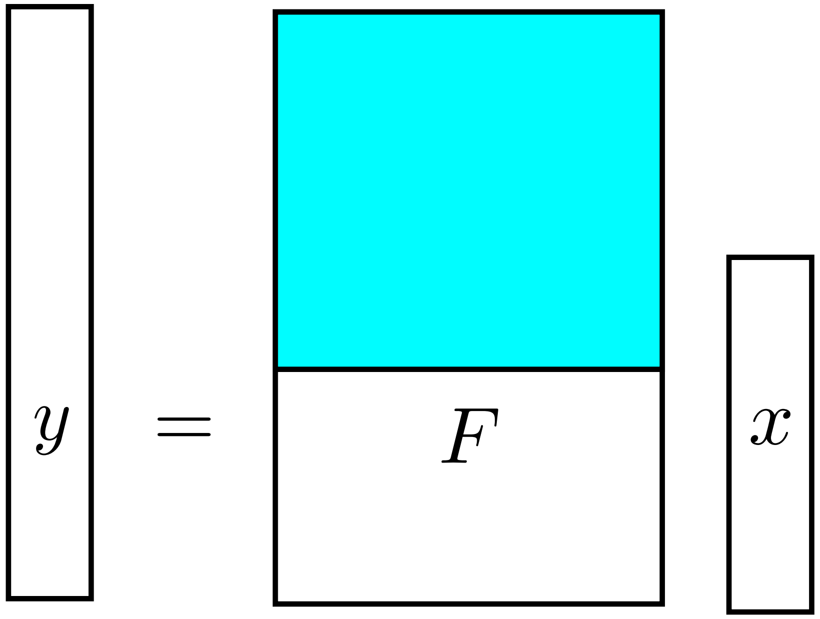
The problem: $y = Fx$, find x

Generically:
- if $M = N$

Unique solution obtained by inversion $x = F^{-1}y$

- if $M > N$ solution obtained from the

inversion of a $N \times N$ submatrix of F with full rank

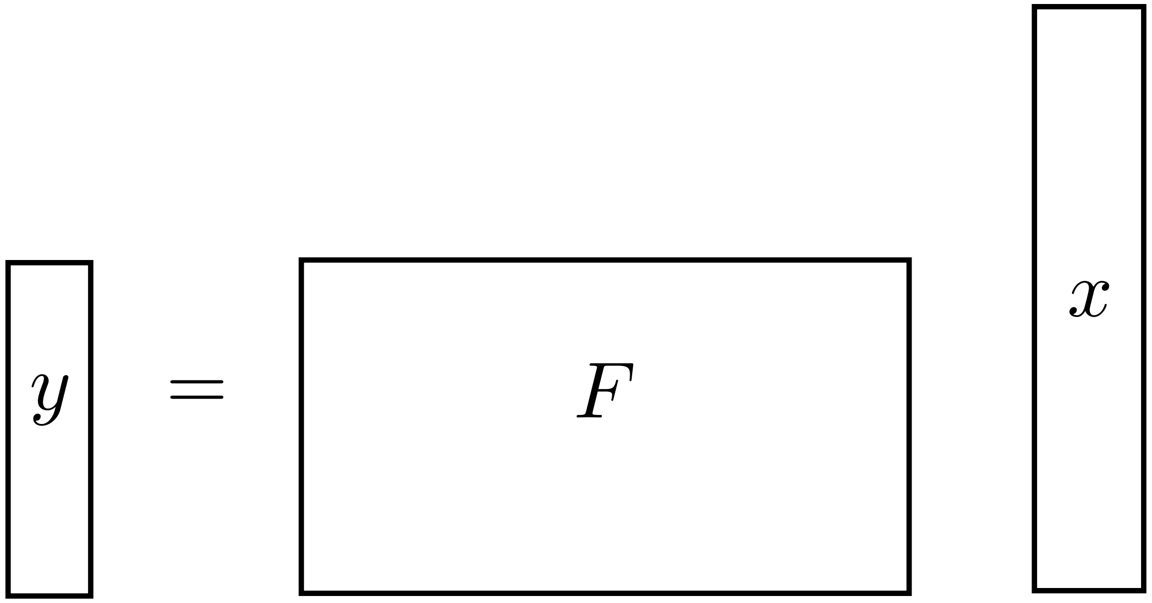


NB: too many equations, redundant system, **<u>but</u>** consistent

because the y measurements are obtained as $y = Fx$

The problem: $y = Fx$, find x

Generically: ● if $M < N$

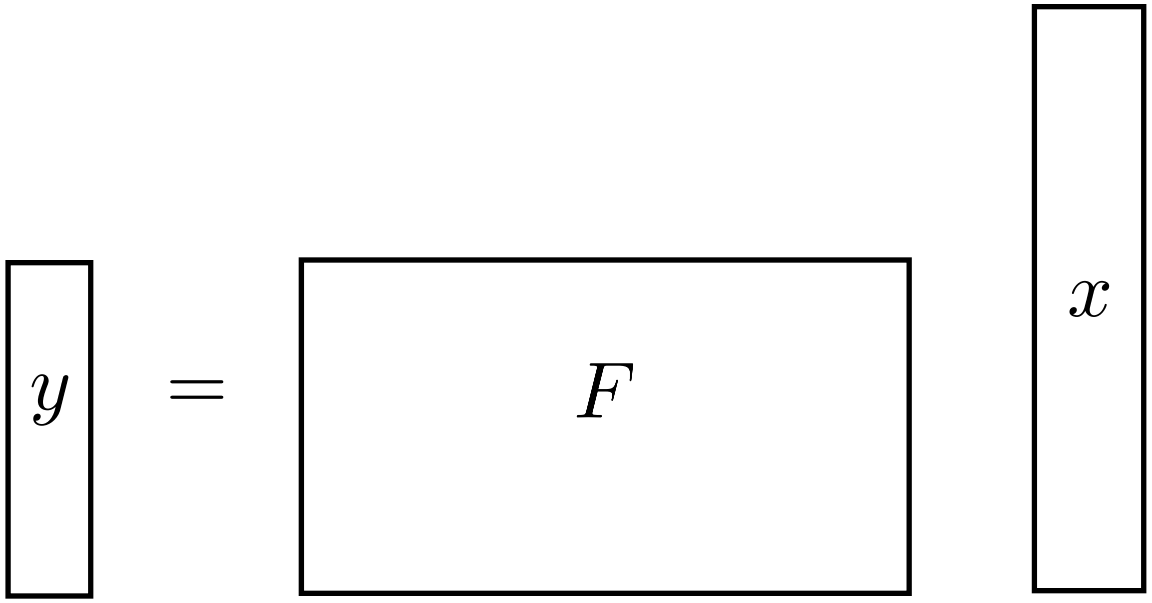


Not enough measurements to determine the signal x from its linear transform y

The problem: $y = Fx$, find x

Generically:

- if $M < N$

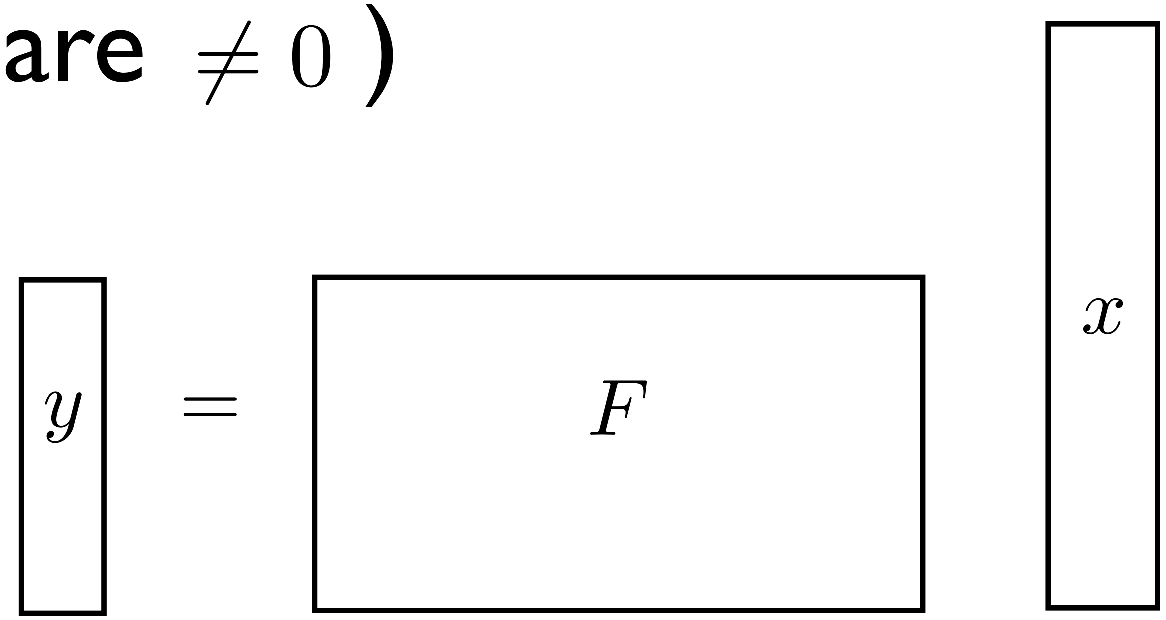


Not enough measurements to determine the signal x from its linear transform y

To invert, you need as many measurements (M) as number of unknowns (N)

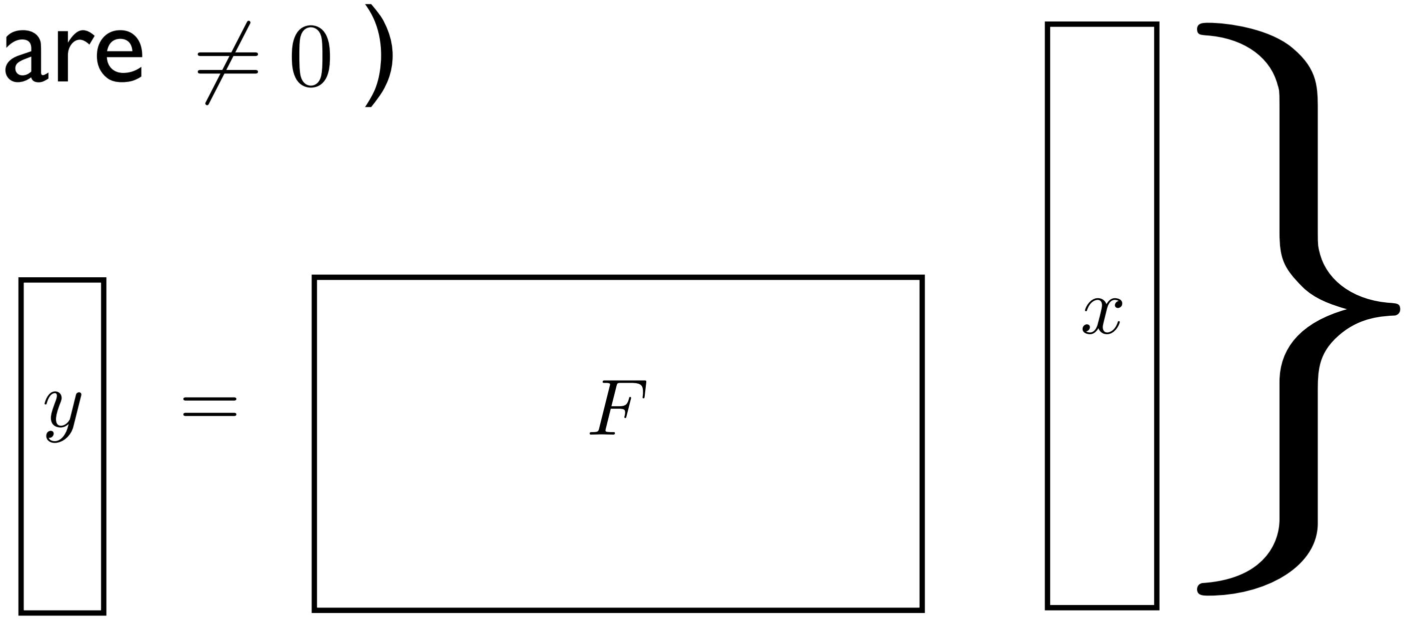
The problem: $y = Fx$, find x

- if $M < N$ but x is sparse (only R of its components are $\neq 0$)



The problem: $y = Fx$, find x

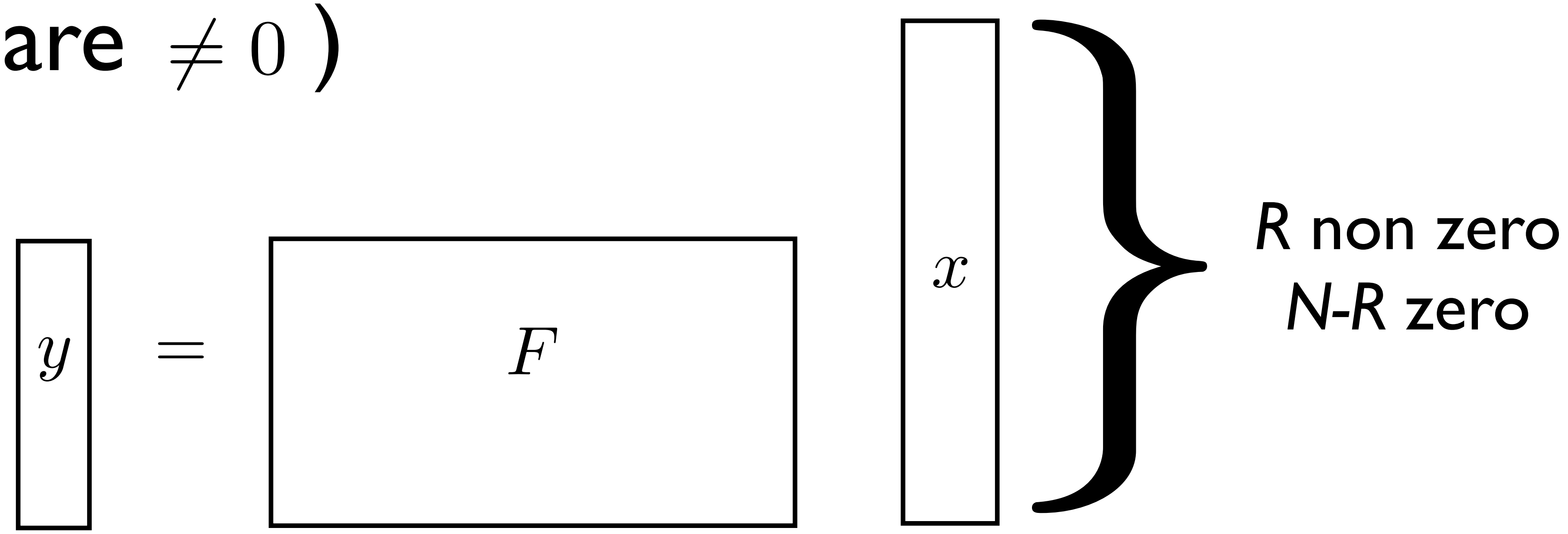
- if $M < N$ but x is sparse (only R of its components are $\neq 0$)



R non zero
N-R zero

The problem: $y = Fx$, find x

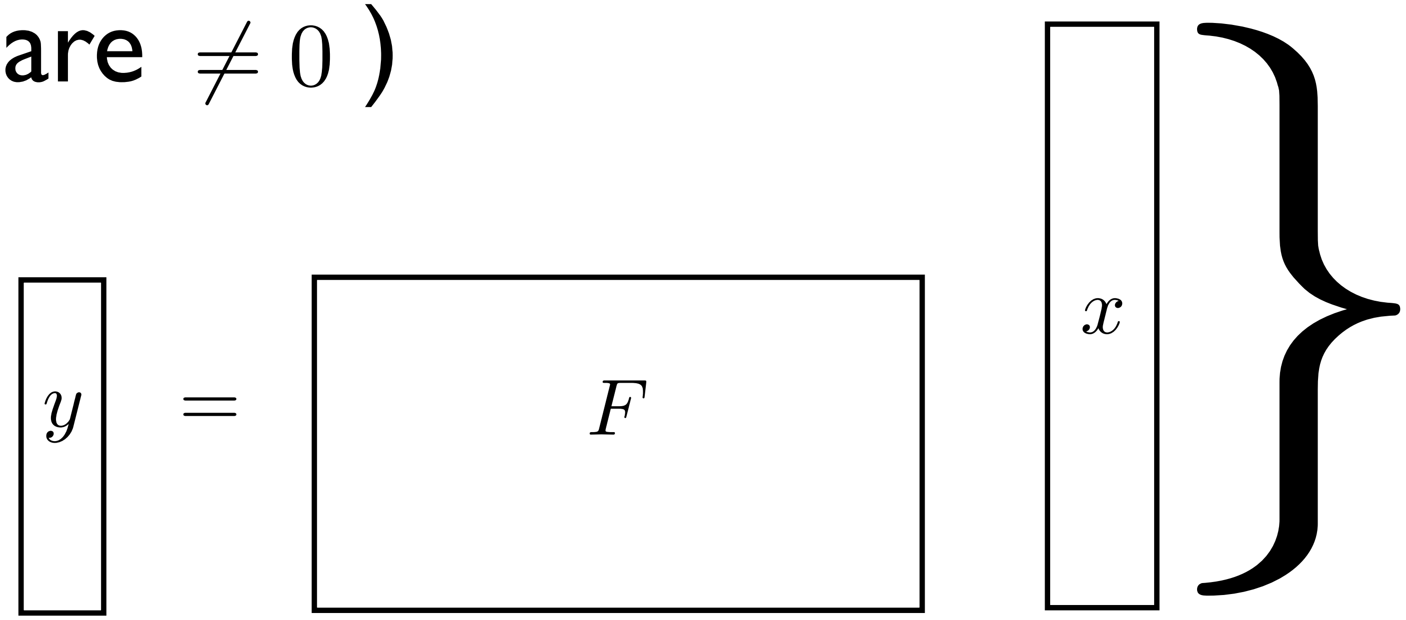
- if $M < N$ but x is sparse (only R of its components are $\neq 0$)



CLAIM:To invert, you need as many measurements (M) as number of unknown (R)

The problem: $y = Fx$, find x

- if $M < N$ but x is sparse (only R of its components are $\neq 0$)



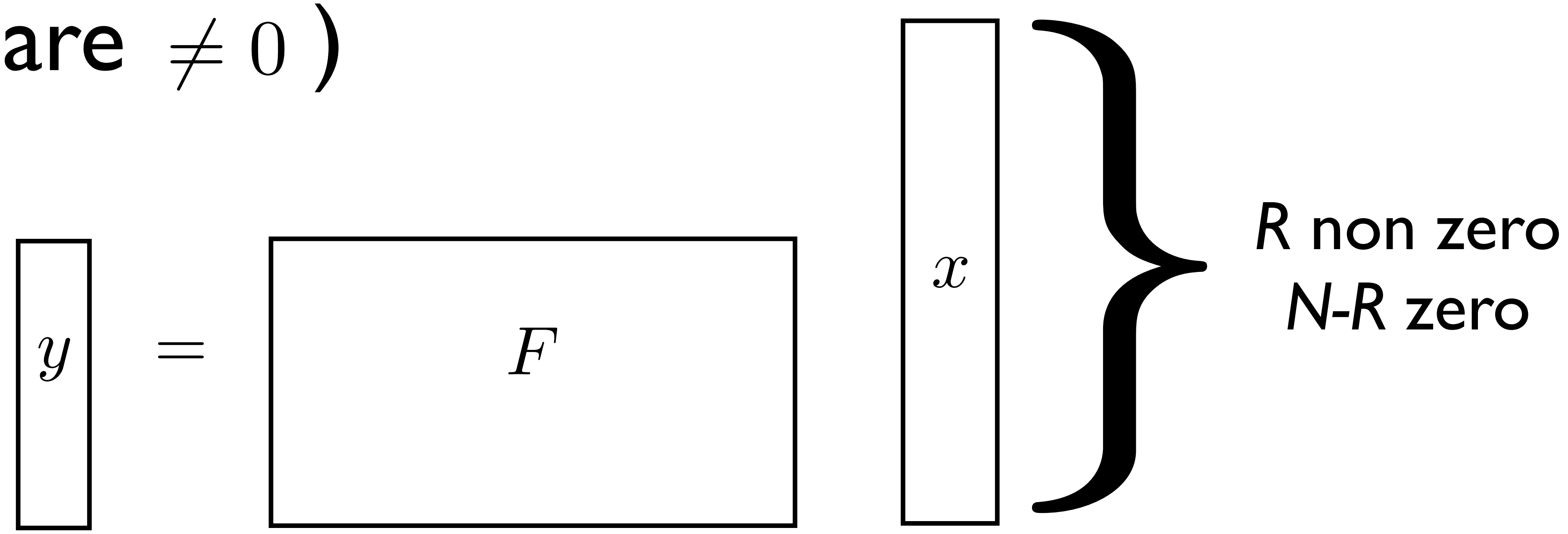
R non zero
N-R zero

CLAIM:To invert, you need as many measurements (M) as number of unknown (R)

If $R < M < N$: the reconstruction of the signal x from the measurement y is possible

The problem: $y = Fx$, find x

- if $M < N$ but x is sparse (only R of its components are $\neq 0$)

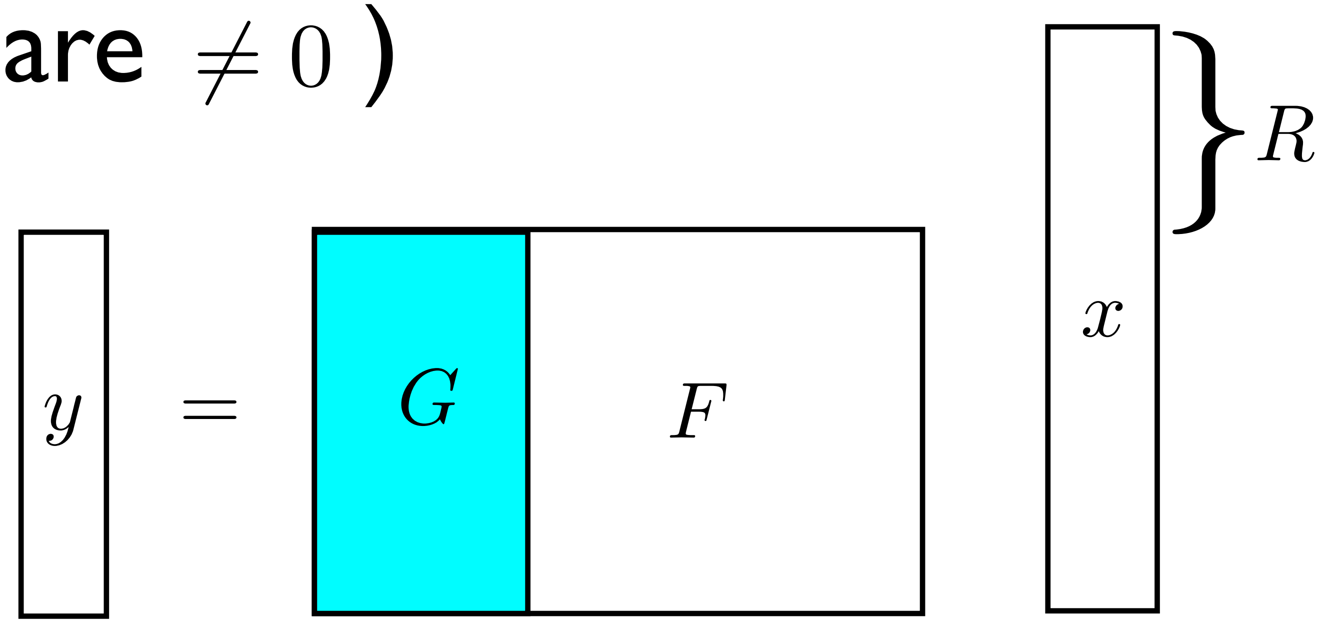


R non zero
N-R zero

A ‘simple’ solution: guess the positions where $x_i \neq 0$

The problem: $y = Fx$, find x

- if $M < N$ but x is sparse (only R of its components are $\neq 0$)



e.g.

$x_1, \ldots, x_R \neq 0$

A ‘simple’ solution: guess the positions where $x_i \neq 0$

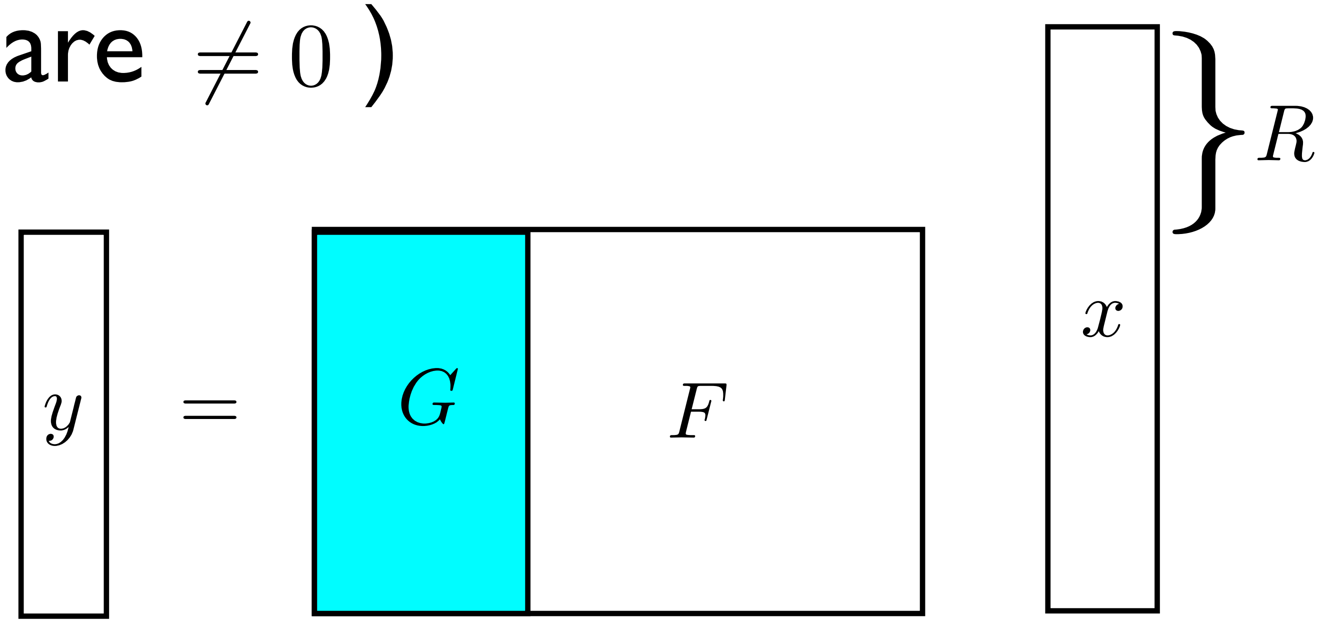
Solve : $y^\mu = \sum_{i=1}^{R} G^{\mu i} x_i \qquad \mu = 1, \ldots, M$

$R < M$ ⇨ too many equations

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

The problem: $y = Fx$, find x

- if $M < N$ but x is sparse (only R of its components are $\neq 0$)



e.g.

$x_1, \dots, x_R \neq 0$

A $\neq 0$

$R <$

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

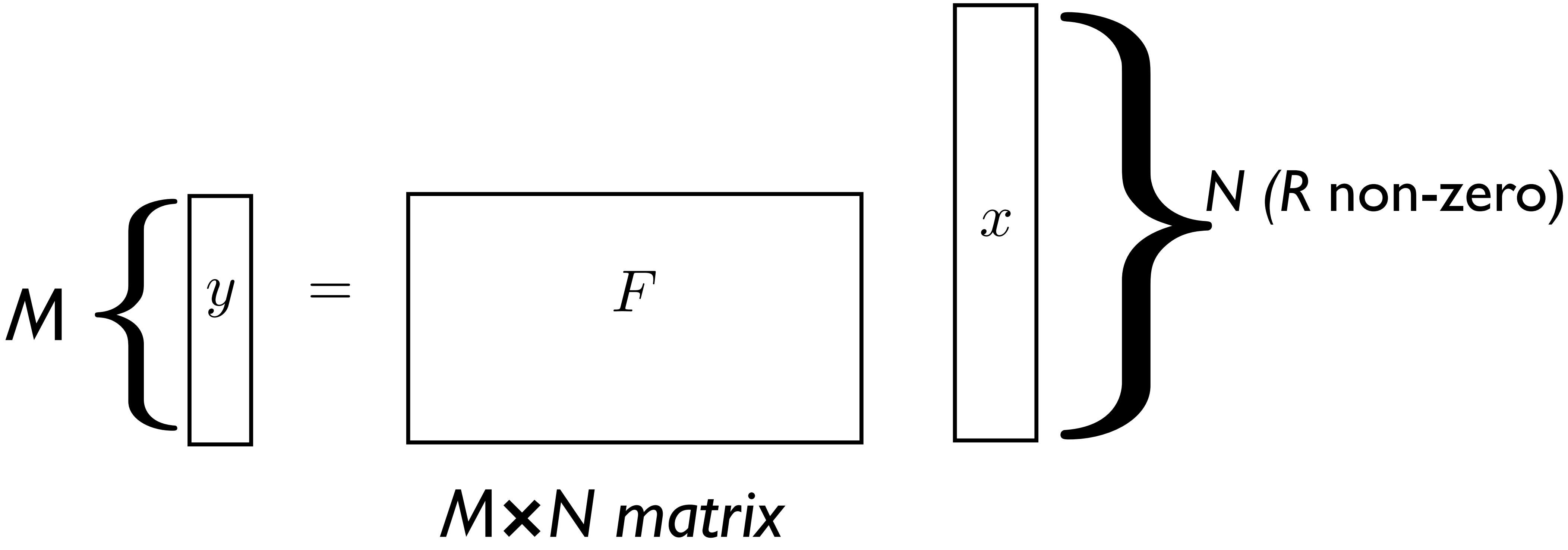
Long, but finite time...

except if

the guess of locations of $x_i \neq 0$ was correct

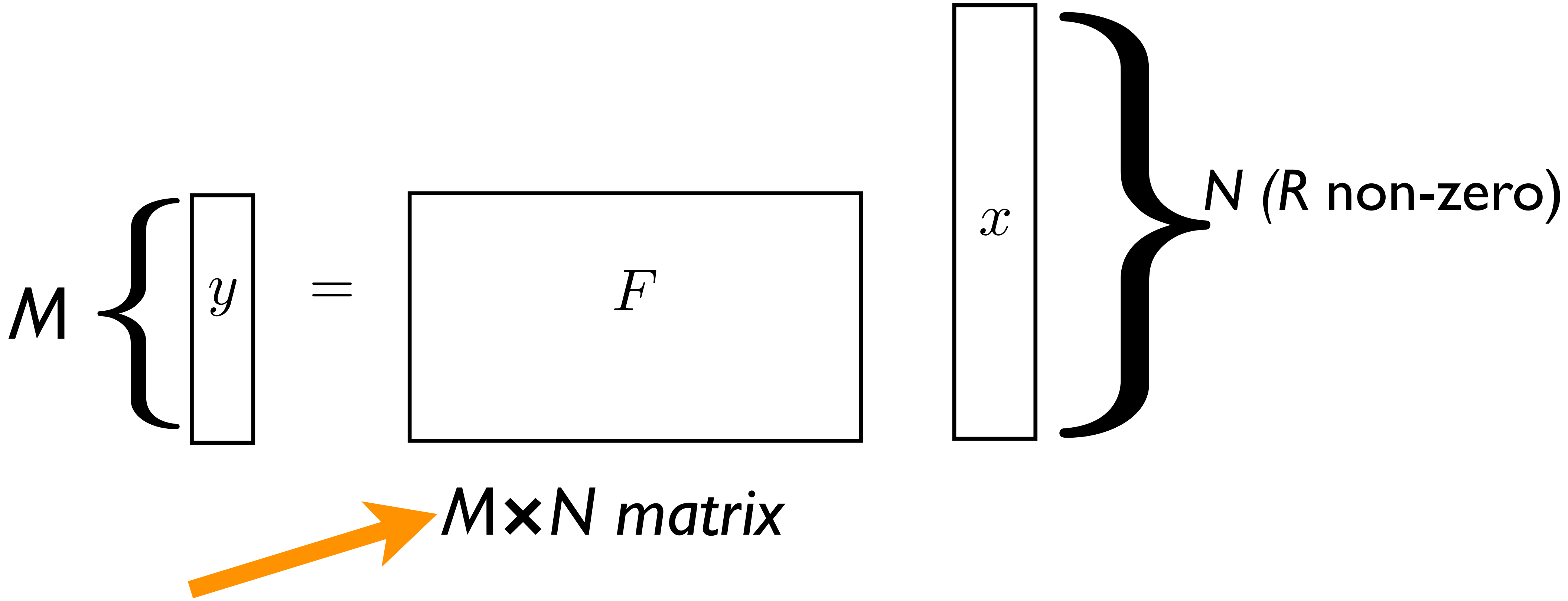
Compressed Sensing

One can reconstruct a N-dimensional sparse signal with R non-zero components from N>M>R measurements



Compressed Sensing

One can reconstruct a N-dimensional sparse signal with R non-zero components from N>M>R measurements



M

$y = F x$

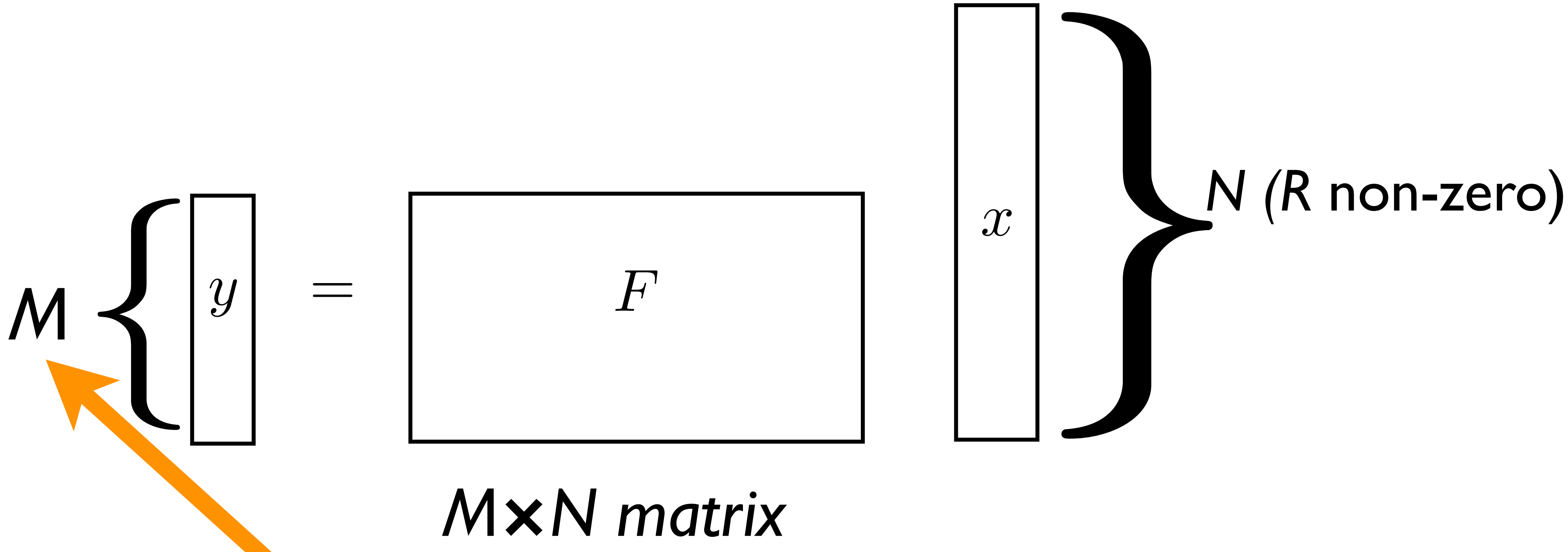
N (*R* non-zero)

M×*N matrix*

- Less measurements (gain of time and precision)

Compressed Sensing

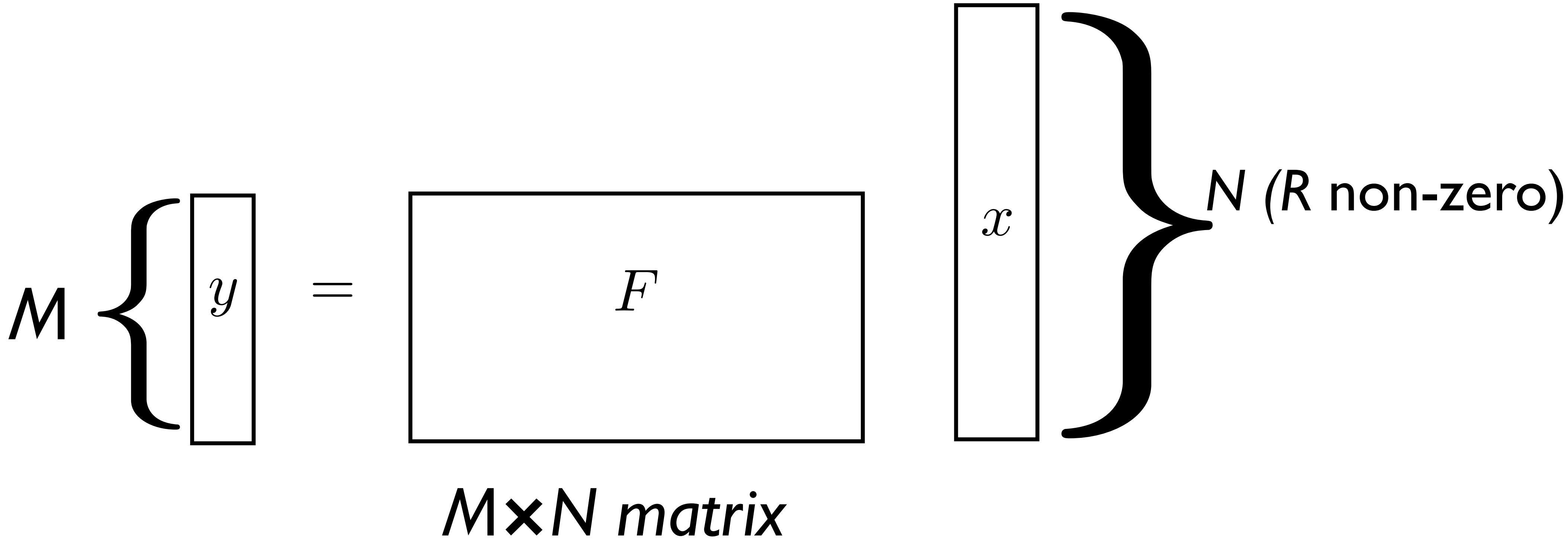
One can reconstruct a N-dimensional sparse signal with R non-zero components from N>M>R measurements



- Less measurements (gain of time and precision)
- Data already compressed (gain of memory storage)

Compressed Sensing

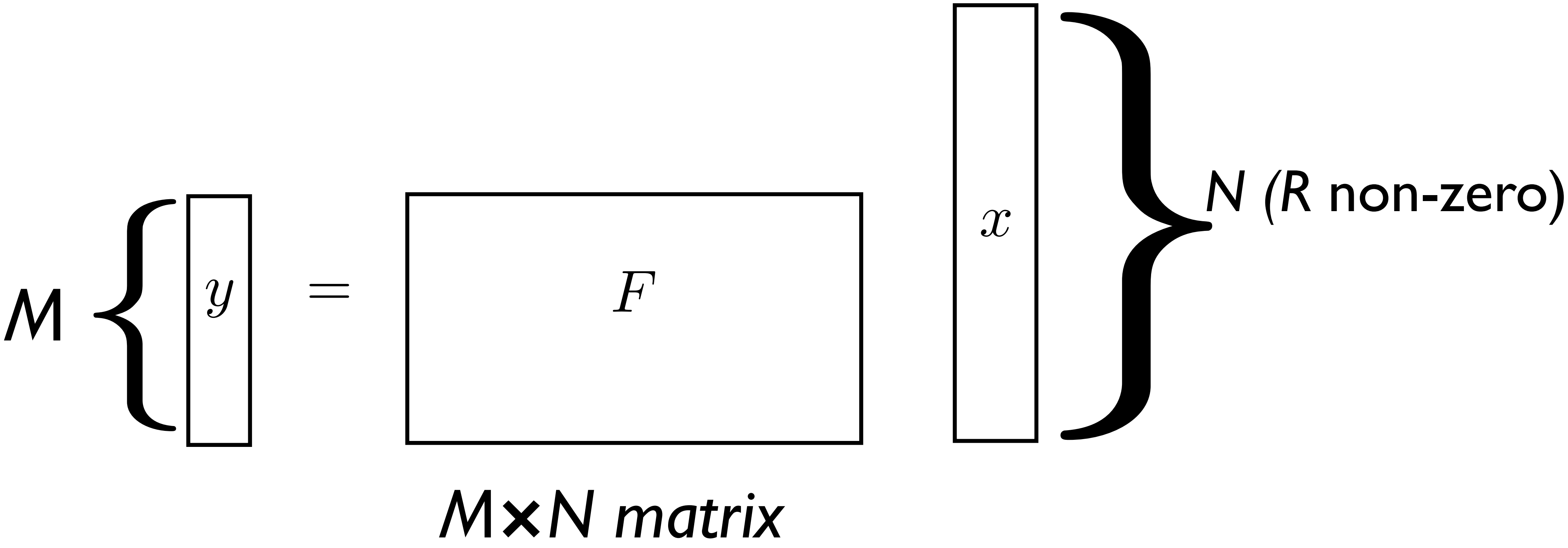
One can reconstruct a N-dimensional sparse signal with R non-zero components from N>M>R measurements



- Less measurements (gain of time and precision)
- Data already compressed (gain of memory storage)
- Price to pay: a reconstruction algorithm is needed

Compressed Sensing

One can reconstruct a N-dimensional sparse signal with R non-zero components from N>M>R measurements

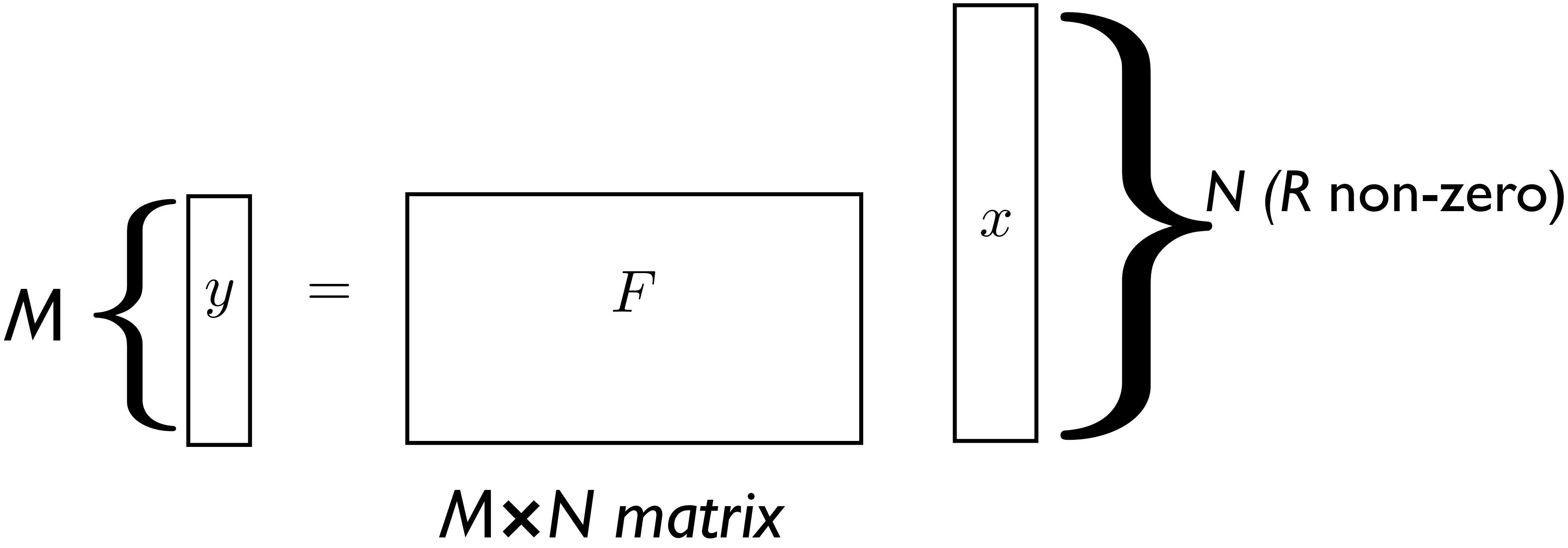


- Less measurements (gain of time and precision)
- Data already compressed (gain of memory storage)
- Price to pay: a reconstruction algorithm is needed

The "simple" algorithm we have presented is too slow!
(need to try exponentially many cases)

Compressed Sensing

One can reconstruct a N-dimensional sparse signal with R non-zero components from N>M>R measurements

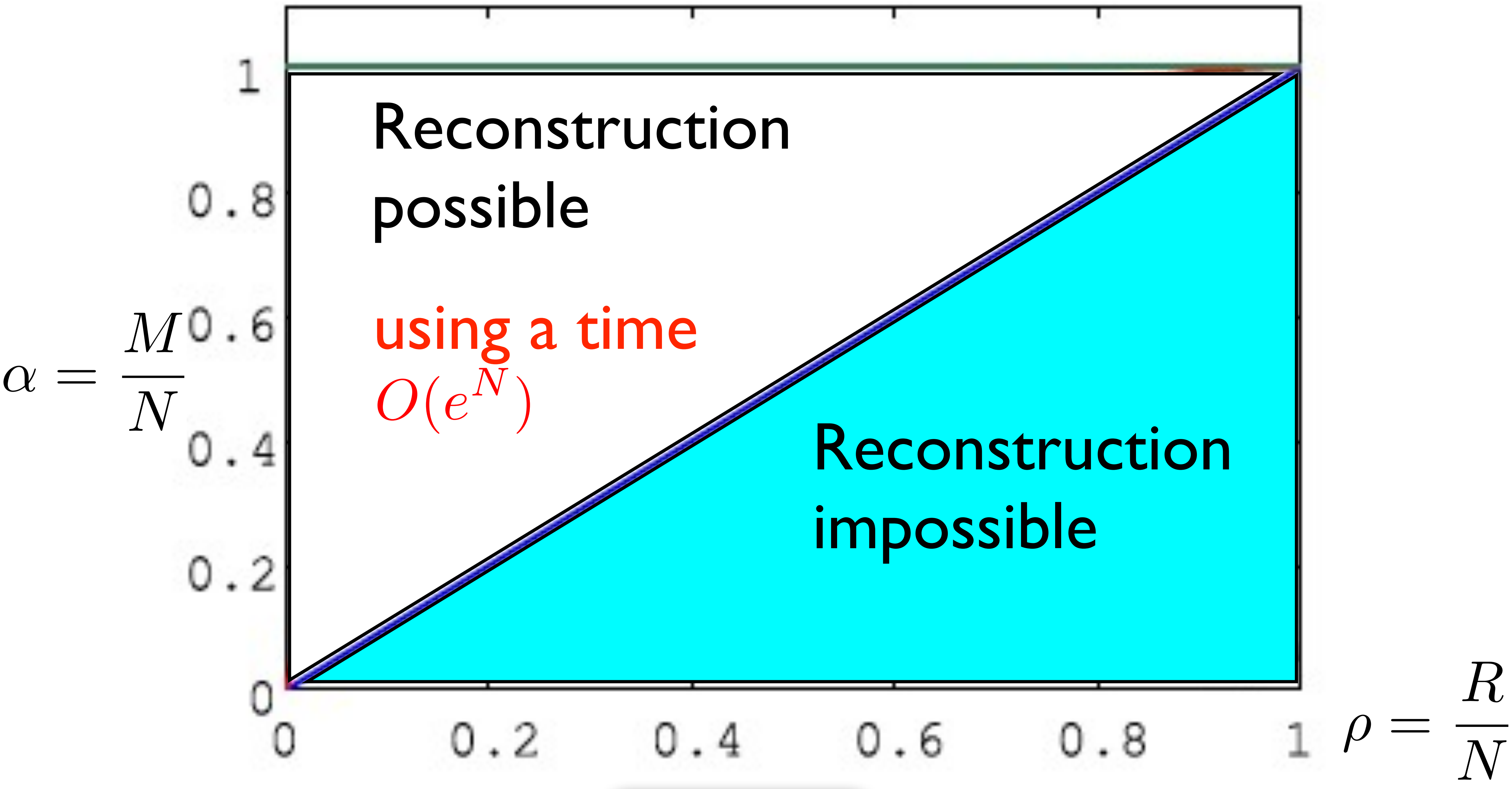


- Less measurements (gain of time and precision)
- Data already compressed (gain of memory storage)
- Price to pay: a reconstruction algorithm is needed

The goal of CS theory:

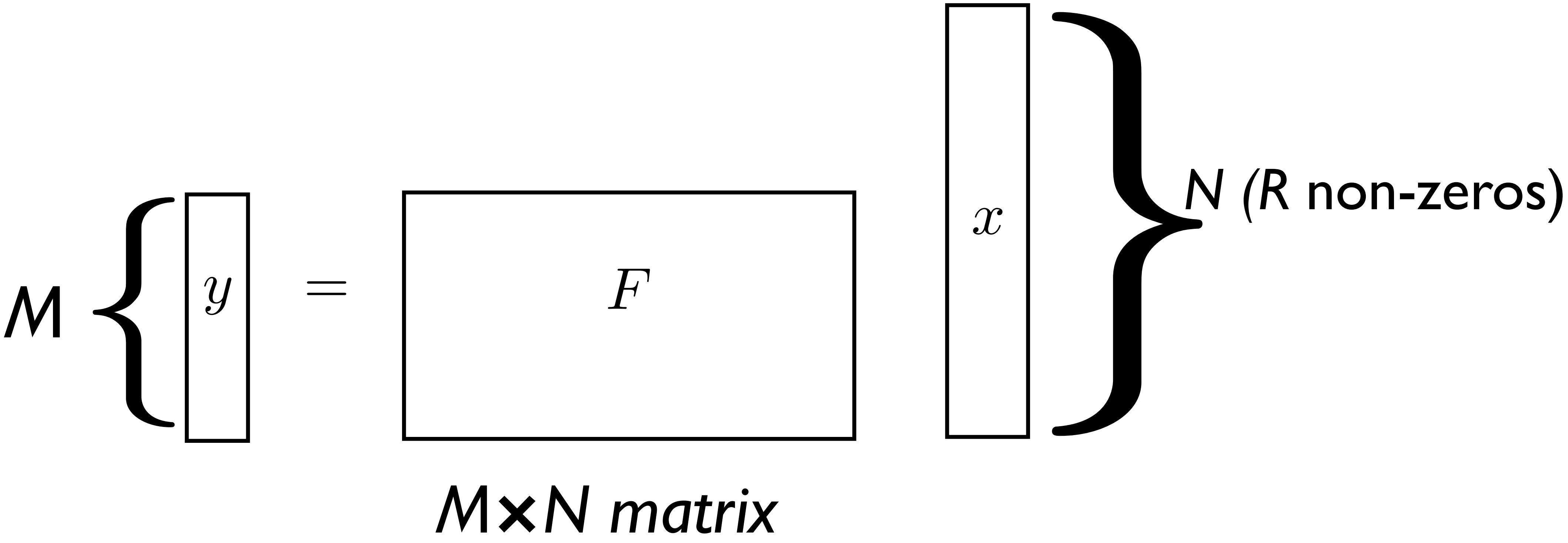
Determine a sensing matrix F and a reconstruction algorithm such that the reconstruction is possible in practice

A phase diagram



$$\boxed{\alpha_c = \rho}$$

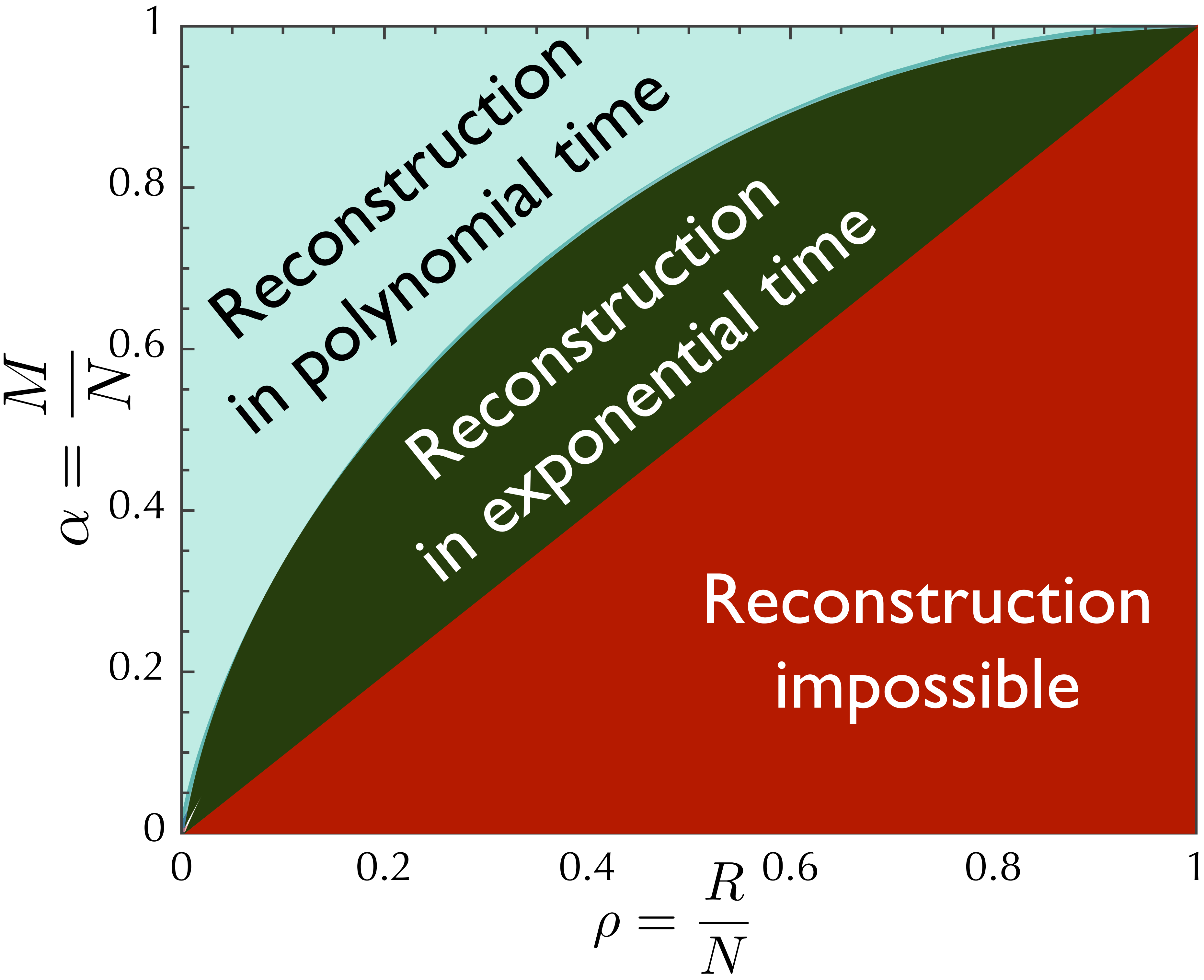
State of the art in CS



- Incoherent samplings (i.e. a random matrix F)
- Reconstruction by minimizing the L1 norm $||\vec{x}||_{L1} = \sum_i |x_i|$

Candès & Tao (2005)
Donoho and Tanner (2005)

State of the art in CS

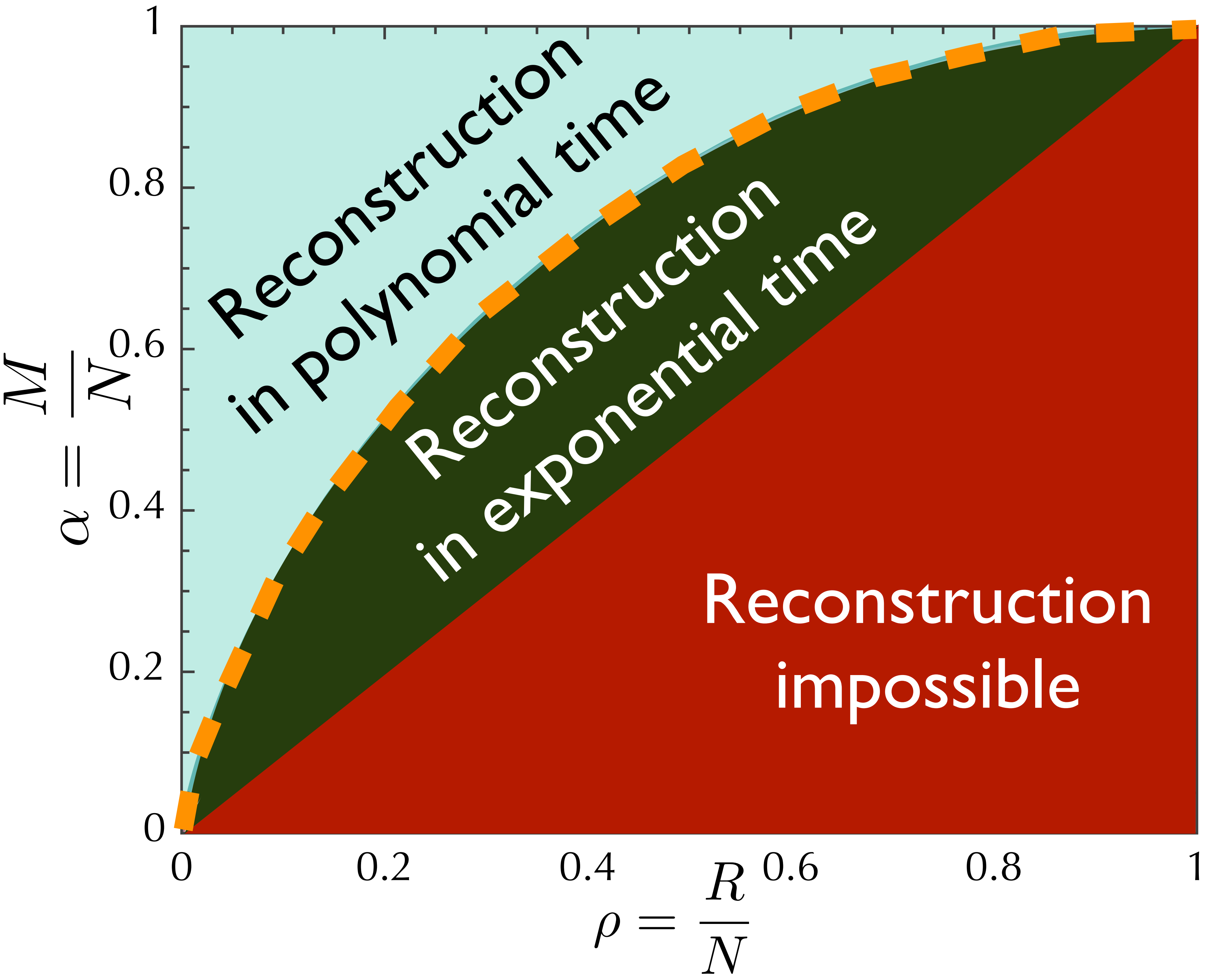


For a signal with

$$\begin{cases} (1-\rho)N \text{ zeros} \\ R=\rho N \text{ non zeros} \end{cases}$$

and a Gaussian random matrix with M =α N

State of the art in CS



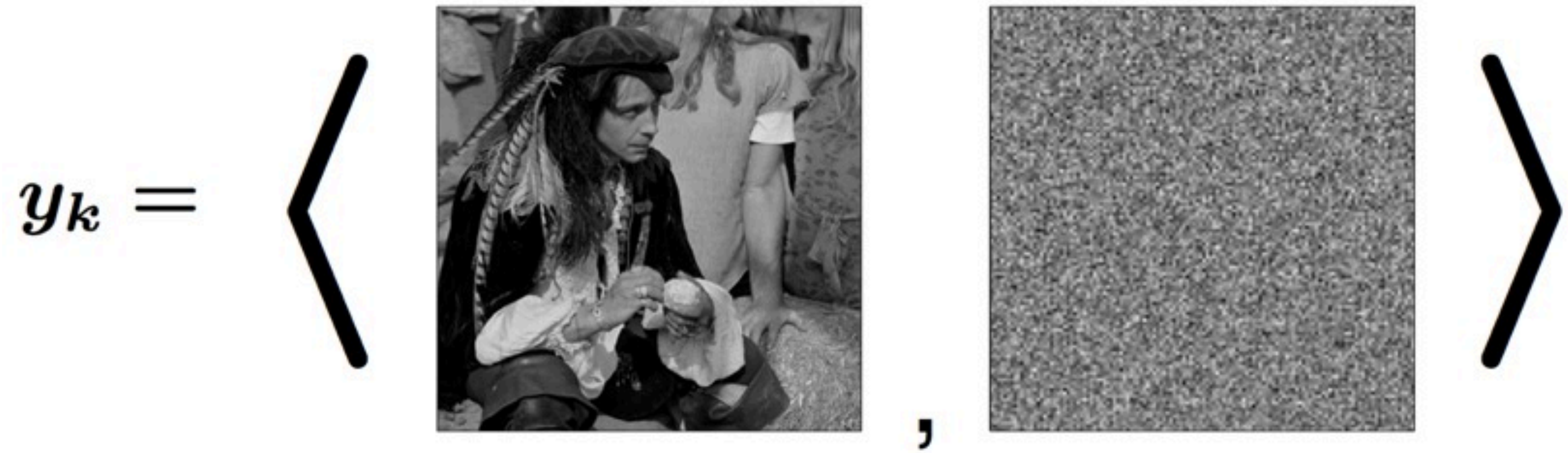
For a signal with
{ (1-ρ)N zeros
R=ρN non zeros

and a Gaussian random matrix with M =α N

Reconstruction limited by the Donoho-Tanner transition for the L_1 norm minimization

Example: measuring a picture

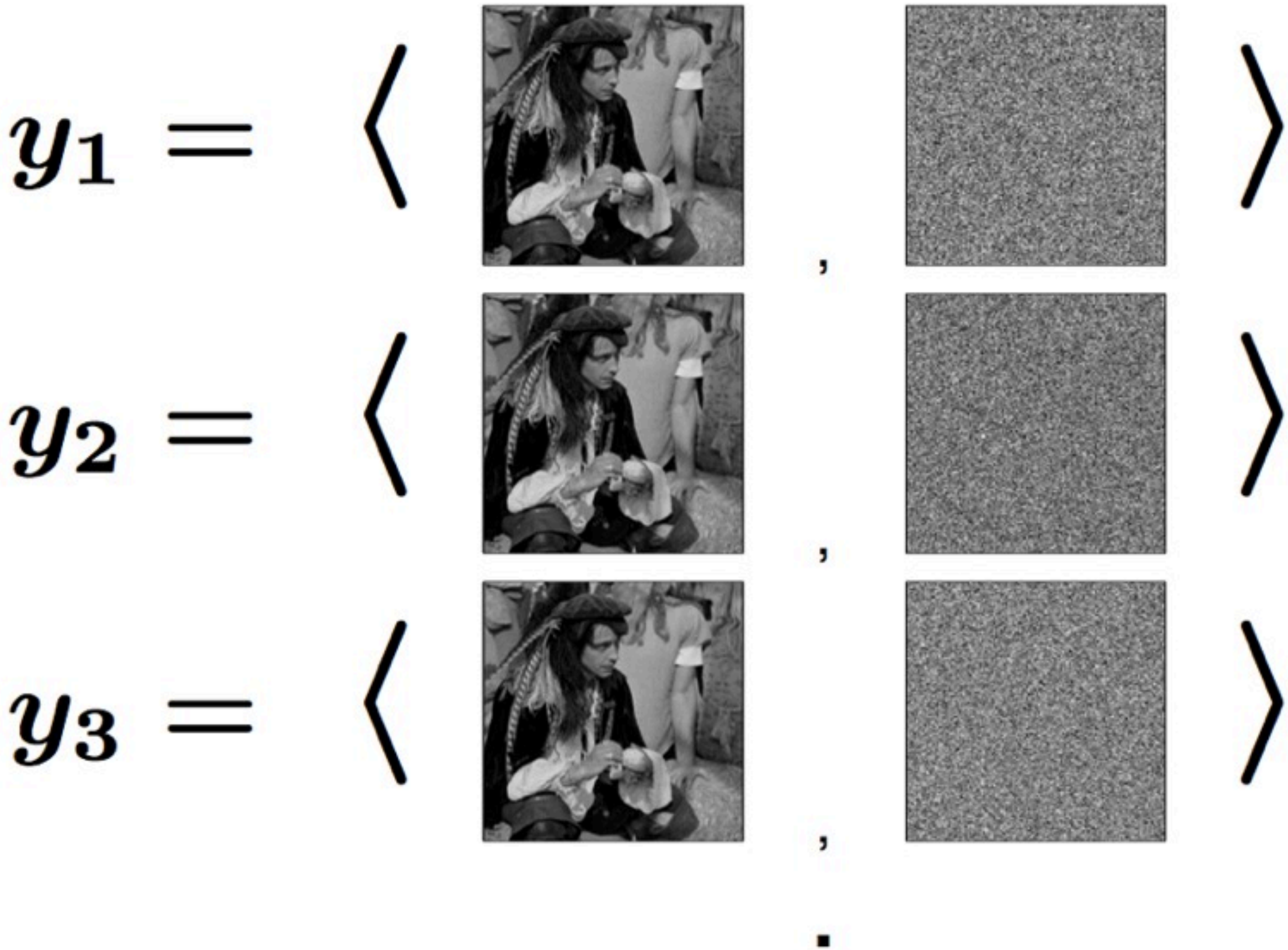
One measurement (scaling product with a random pattern)



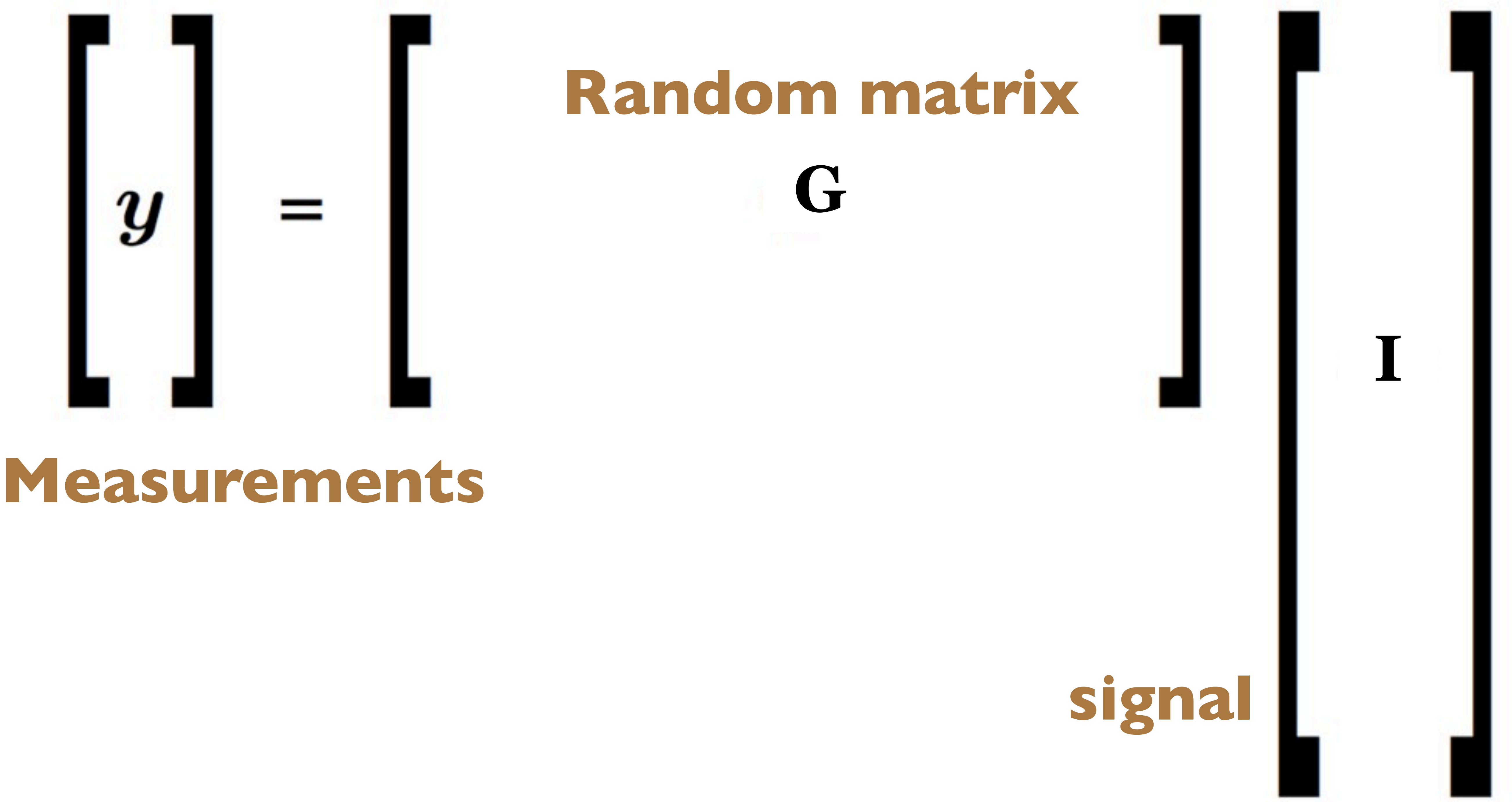
- Each measurement touches every part of the underlying signal/image

Example: measuring a picture

Many measurements (scaling product with many random patterns)



Example: measuring a picture



Example: measuring a picture

- Take $K = 96000$ incoherent measurements $y = \mathbf{GI}$

From 10^6 points, but only, 25.000 non zero

- Solve

$$\min \ \|\mathbf{x}\|_{\ell_1} \quad \text{subject to} \quad \mathbf{G}\mathbf{\Psi}\mathbf{x} = y$$

$\mathbf{\Psi}$ = wavelet transform

original (25k wavelets)

perfect recovery

Compressed Sensing:

A (short!)review of the present litterature:

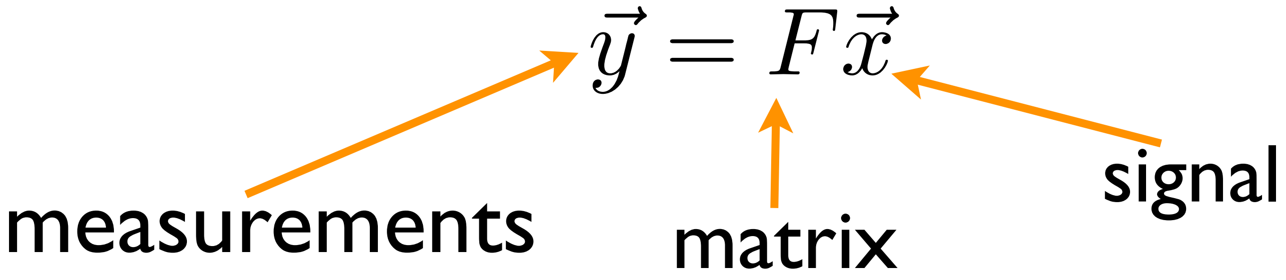
- Record data already in a compressed form
- Less measurements (faster, more precise)...
- ... but need for a reconstruction algorithm!
- State of the art: L_1-minimization and random measurements

Compressed sensing

or y=Ax revisited

- What is compressed sensing?
- **What is the link between statistical physics and compressed sensing?**
- How can one use statistical physics to improve on compressed sensing technics?

A statistical-physics approach to compressed sensing



$$\vec{y} = F\vec{x}$$

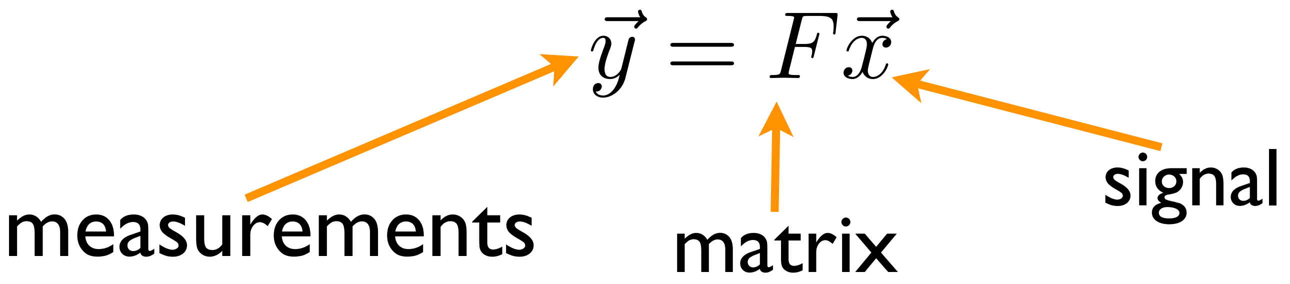
measurements

matrix

signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

A statistical-physics approach to compressed sensing



$$\vec{y} = F\vec{x}$$

measurements

matrix

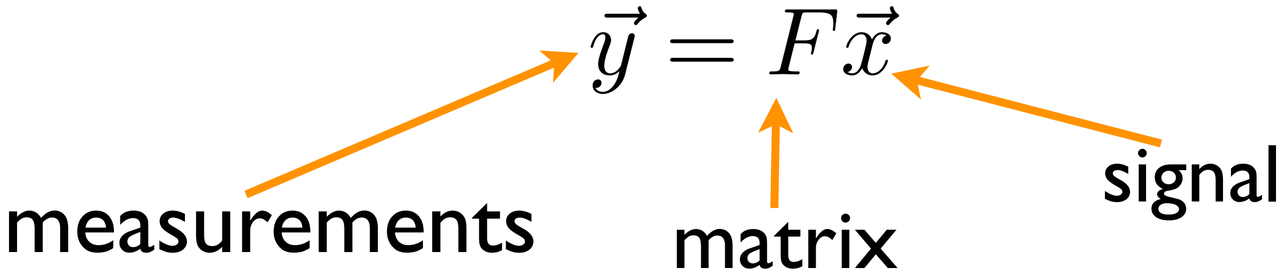
signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Inference problem:

Estimate P(x|y), and choose x accordingly

A statistical-physics approach to compressed sensing



$$\vec{y} = F\vec{x}$$

measurements

matrix

signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Inference problem:

Estimate P(x|y), and choose x accordingly

But how to estimate P(x|y)?

A statistical-physics approach to compressed sensing

$$\vec{y} = F\vec{x}$$

measurements

matrix

signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Inference problem:

Estimate P(x|y), and choose x accordingly

But how to estimate P(x|y)?

$$P(A|B)P(B) = P(B|A)P(A)$$

Rev. Thomas Bayes 1702-1762

A statistical-physics approach to compressed sensing

$$\vec{y} = F\vec{x}$$

measurements

matrix

signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem: ⇨ $$P(\vec{x}|\vec{y}) = \frac{P(\vec{x})}{P(\vec{y})} P(\vec{y}|\vec{x}) = \frac{P(\vec{x})P(\vec{y}|\vec{x})}{Z}$$

A statistical-physics approach to compressed sensing

$$\vec{y} = F\vec{x}$$

measurements matrix signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem: ➡ $P(\vec{x}|\vec{y}) = \frac{P(\vec{x})}{P(\vec{y})}P(\vec{y}|\vec{x}) = \frac{P(\vec{x})P(\vec{y}|\vec{x})}{Z}$

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

A statistical-physics approach to compressed sensing

$$\underset{\text{measurements}}{\vec{y}} = \underset{\text{matrix}}{F}\,\underset{\text{signal}}{\vec{x}}$$

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem: → $P(\vec{x}|\vec{y}) = \frac{P(\vec{x})}{P(\vec{y})} P(\vec{y}|\vec{x}) = \frac{P(\vec{x})P(\vec{y}|\vec{x})}{Z}$

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} \prod_{i=1}^{N} \left[(1-\rho)\,\delta(x_i) + \rho\phi(x_i)\right] \underbrace{\prod_{\mu=1}^{M} \delta\left(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i\right)}_{\text{Solution of the linear system}}$$

A statistical-physics approach to compressed sensing

$$\vec{y} = F\vec{x}$$

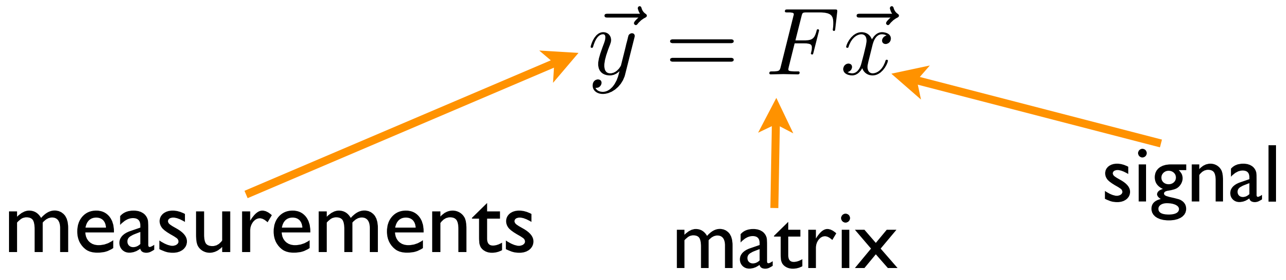
measurements matrix signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem: $\Rightarrow$ $P(\vec{x}|\vec{y}) = \frac{P(\vec{x})}{P(\vec{y})} P(\vec{y}|\vec{x}) = \frac{P(\vec{x})P(\vec{y}|\vec{x})}{Z}$

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} \underbrace{\prod_{i=1}^{N} \left[(1-\rho)\,\delta(x_i) + \rho\phi(x_i)\right]}_{\text{Sparse vector}} \underbrace{\prod_{\mu=1}^{M} \delta\left(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i\right)}_{\text{Solution of the linear system}}$$

A statistical-physics approach to compressed sensing



$$\vec{y} = F\vec{x}$$

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem:

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

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

A statistical-physics approach to compressed sensing

$$\underset{\text{measurements}}{\vec{y}} = \underset{\text{matrix}}{F}\,\underset{\text{signal}}{\vec{x}}$$

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem:

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

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

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta}\sum_{\mu=1}^{M}(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

A statistical-physics approach to compressed sensing

$$\vec{y} = F\vec{x}$$

measurements

matrix

signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

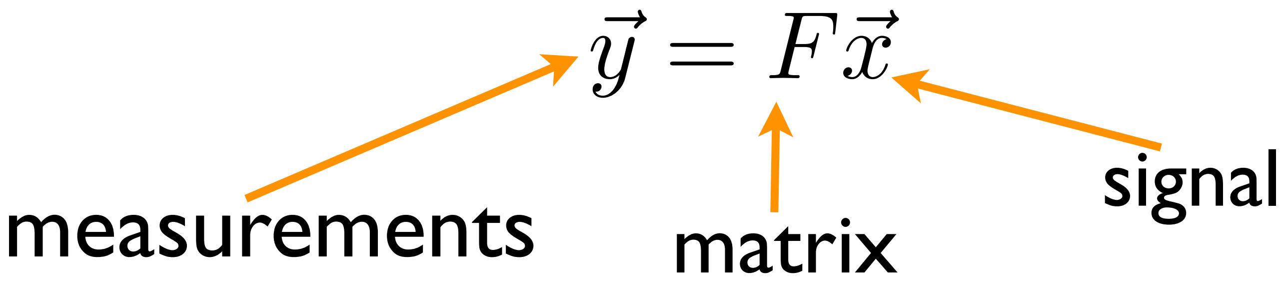
Bayes Theorem:

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

A mean-field disordered statistical physics problem

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta} \sum_{\mu=1}^{M} (y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

A statistical-physics approach to compressed sensing



$$\vec{y} = F\vec{x}$$

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem:

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

A mean-field disordered statistical physics problem

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{\overbrace{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta}\sum_{\mu=1}^{M}(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}^{\text{Hamiltonian}}}$$

A statistical-physics approach to compressed sensing

$$\underset{\text{measurements}}{\vec{y}} = \underset{\text{matrix}}{F}\,\underset{\text{signal}}{\vec{x}}$$

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem: ⇒ $$P(\vec{x}|\vec{y}) = \frac{P(\vec{x})}{P(\vec{y})} P(\vec{y}|\vec{x}) = \frac{P(\vec{x})P(\vec{y}|\vec{x})}{Z}$$

A mean-field disordered statistical physics problem

$$P(\vec{x}|\vec{y}) = \frac{1}{\underset{\text{Partition sum}}{\boxed{Z}}} e^{\overbrace{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta} \sum_{\mu=1}^{M} (y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}^{\text{Hamiltonian}}}$$

A statistical-physics approach to compressed sensing

$$\underset{\text{measurements}}{\vec{y}} = \underset{\text{matrix}}{F}\,\underset{\text{signal}}{\vec{x}}$$

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem: ⇨

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

A mean-field disordered statistical physics problem

Disordered interaction

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta}\sum_{\mu=1}^{M}(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

A statistical-physics approach to compressed sensing

$$\vec{y} = F\vec{x}$$

measurements matrix signal

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem:

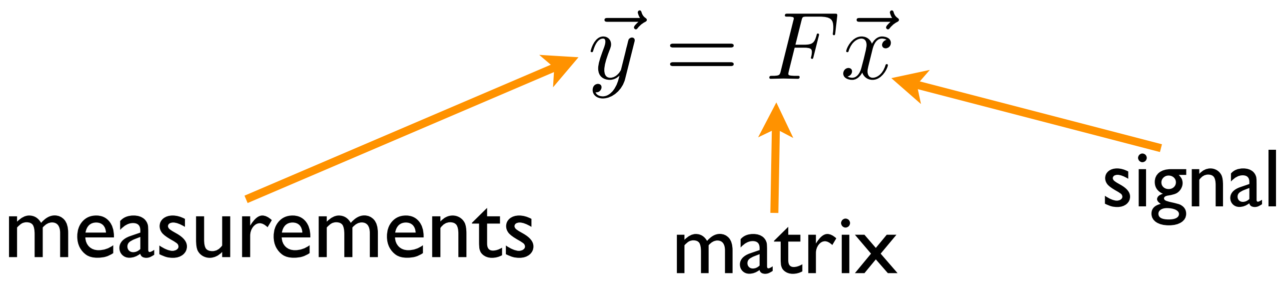
$$P(\vec{x}|\vec{y}) = \frac{P(\vec{x})}{P(\vec{y})} P(\vec{y}|\vec{x}) = \frac{P(\vec{x})P(\vec{y}|\vec{x})}{Z}$$

A mean-field disordered statistical physics problem

Mean-field
long-range interactions

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta} \sum_{\mu=1}^{M} (y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

A statistical-physics approach to compressed sensing



$$\vec{y} = F\vec{x}$$

measurements

matrix

signal

Estimating the probability of each value of x is equivalent to solving a mean-field disordered statistical physics problem

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta} \sum_{\mu=1}^{M} (y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

A statistical-physics approach to compressed sensing

$$\vec{y} = F\vec{x}$$

measurements

matrix

signal

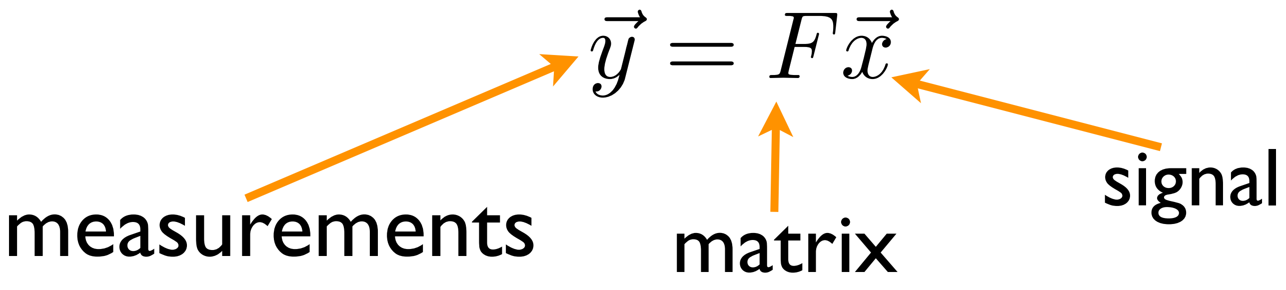
How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem:

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

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

A statistical-physics approach to compressed sensing



$$\vec{y} = F\vec{x}$$

How to reconstruct $\vec{x}$ from $F, \vec{y}$?

Bayes Theorem:

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

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} \prod_{i=1}^{N} P(x_i) \prod_{\mu=1}^{M} \delta\left(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i\right) \quad \text{with} \quad P(x_i) = (1-\rho)\delta(x_i) + \rho\phi(x_i)$$

Theorem: sampling from P(x|y) gives the correct solution as long as $\alpha>\rho_0$ if: a) $\Phi(x)>0 \ \forall x$ and b) $1>\rho>0$

The probabilistic approach is optimal, even if we do not know the correct Φ(x)! In practice, we use a Gaussian distribution

A sketch of the proof

Consider the system constrained to be at
distances larger than D with respect to the solution

$$Y(D,\epsilon) = \int \prod_{i=1}^{N} (dx_i \ [(1-\rho)\delta(x_i) + \rho\phi(x_i)]) \prod_{\mu=1}^{M} \delta_\epsilon \left(\sum_i F_{\mu i}(x_i - s_i) \right) \mathbb{I} \left(\sum_{i=1}^{N} (x_i - s_i)^2 > ND \right)$$

1) Y(0) is infinite if $\alpha > \rho_0$ (equivalently if M>R)
(just count the delta functions! N-R+M deltas versus N integrals...)

2) Y(D) is finite for any D>0
(bound by a first moment method, or "annealed" computation)

A sketch of the proof

Consider the system constrained to be at
distances larger than D with respect to the solution

$$Y(D,\epsilon) = \int \prod_{i=1}^{N} (dx_i\ [(1-\rho)\delta(x_i)+\rho\phi(x_i)]) \prod_{\mu=1}^{M} \delta_\epsilon \left(\sum_i F_{\mu i}(x_i - s_i)\right) \mathbb{I}\left(\sum_{i=1}^{N}(x_i-s_i)^2 > ND\right)$$

1) Y(0) is infinite if $\alpha>\rho_0$ (equivalently if M>R)

(just count the delta functions! N-R+M deltas versus N integrals...)

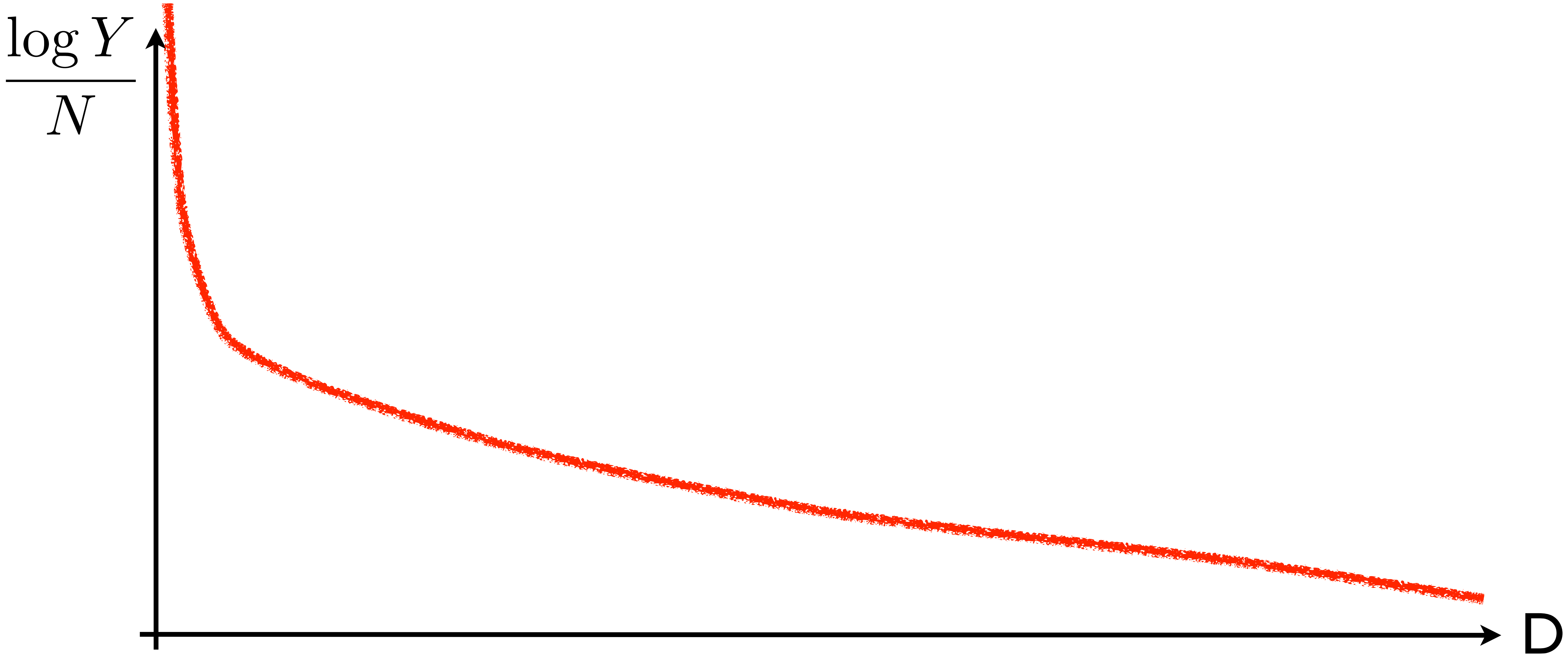
2) Y(D) is finite for any D>0

(bound by a first moment method, or "annealed" computation)

If $\alpha>\rho_0$, the measure is always dominated by the solution

A sketch of the proof

Consider the system constrained to be at distances larger than D with respect to the solution



$$Y(D,\epsilon) = \int \prod_{i=1}^{N} \left(dx_i\ \left[(1-\rho)\delta(x_i) + \rho\phi(x_i)\right]\right) \prod_{\mu=1}^{M} \delta_\epsilon \left(\sum_i F_{\mu i}(x_i - s_i) \right) \mathbb{I} \left(\sum_{i=1}^{N} (x_i - s_i)^2 > ND \right)$$

Compressed sensing

or y=Ax revisited

- What is compressed sensing?
- What is the link between statistical physics and compressed sensing?
- **How can one use statistical physics to improve on compressed sensing technics?**

A statistical physics approach to compressed sensing

One can use statistical physics tools for

I) Computing phase transitions analytically (reconstruction/non reconstruction, etc...)

Tools: Replica method from spin glass theory, etc...

II) Develop new algorithms, and design new matrices to improve on the L_1 state-of-the art.

Tools: Replica and Cavity method from spin glass theory, Mean field methods from stat-phys, Physics intuition, etc....

A statistical physics approach to compressed sensing

One can use statistical physics tools for

1) Computing phase transitions analytically (reconstruction/non reconstruction, etc...)
Tools: Replica method from spin glass theory, etc...

II) Develop new algorithms, and design new matrices to improve on the L_1 state-of-the art.
Tools: Replica and Cavity method from spin glass theory, Mean field methods from stat-phys, Physics intuition, etc....

Statistical physics of compressed sensing

Model with N infinite-range 1d interacting particles with positions x_i

What is the phase diagram of the system?

$$Z(y) = \int \prod_{i=1}^{N} dx_i P(x|y) \qquad F(\vec{y}) = -\log Z(\vec{y})$$

Statistical physics of compressed sensing

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta} \sum_{\mu=1}^{M} (y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

Model with N infinite-range 1d interacting particles with positions x_i

What is the phase diagram of the system?

$$Z(y) = \int \prod_{i=1}^{N} dx_i P(x|y) \qquad F(\vec{y}) = -\log Z(\vec{y})$$

Statistical physics of compressed sensing

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta} \sum_{\mu=1}^{M} (y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

Model with N infinite-range 1d interacting particles with positions x_i

What is the phase diagram of the system?

$$Z(y) = \int \prod_{i=1}^{N} dx_i P(x|y) \qquad F(\vec{y}) = -\log Z(\vec{y})$$

Use a random matrix F, and Gauss-Bernoulli signal

Statistical physics of compressed sensing

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta} \sum_{\mu=1}^{M} (y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

Model with N infinite-range 1d interacting particles with positions x_i

What is the phase diagram of the system?

$$Z(y) = \int \prod_{i=1}^{N} dx_i P(x|y) \qquad F(\vec{y}) = -\log Z(\vec{y})$$

Use a random matrix F, and Gauss-Bernoulli signal

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

Statistical physics of compressed sensing

$$P(\vec{x}|\vec{y}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \log P(x_i) - \frac{1}{2\Delta} \sum_{\mu=1}^{M} (y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}$$

Model with N infinite-range 1d interacting particles with positions x_i

What is the phase diagram of the system?

$$Z(y) = \int \prod_{i=1}^{N} dx_i P(x|y) \qquad F(\vec{y}) = -\log Z(\vec{y})$$

Use a random matrix F, and Gauss-Bernoulli signal

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

Replica method ➡ $\overline{\log Z} = \lim_{n \to 0} \frac{\overline{Z^n} - 1}{n}$

Analytic study: cavity equations, density evolution, replicas

$$E(Z^n) = \max_{Q,q,m,\hat{Q},\hat{q},\hat{m}} e^{Nn\Phi(Q,q,m,\hat{Q},\hat{q},\hat{m})}$$

$$\Phi(Q,q,m,\hat{Q},\hat{q},\hat{m}) = -\frac{1}{2N}\sum_\mu \frac{q-2m+\rho+\Delta_\mu}{\Delta_\mu+Q-q} - \frac{1}{2N}\sum_\mu \log\left(\Delta_\mu+Q-q\right) + \frac{Q\hat{Q}}{2} - m\hat{m} + \frac{q\hat{q}}{2}$$

$$+\int \mathcal{D}z \int \mathrm{d}x_0 \left[(1-\rho_0)\delta(x_0)+\rho_0\phi_0(x_0)\right] \log\left\{\int \mathrm{d}x\, e^{-\frac{\hat{Q}+\hat{q}}{2}x^2+\hat{m}xx_0+z\sqrt{\hat{q}}x}\left[(1-\rho)\delta(x)+\rho\phi(x)\right]\right\}$$

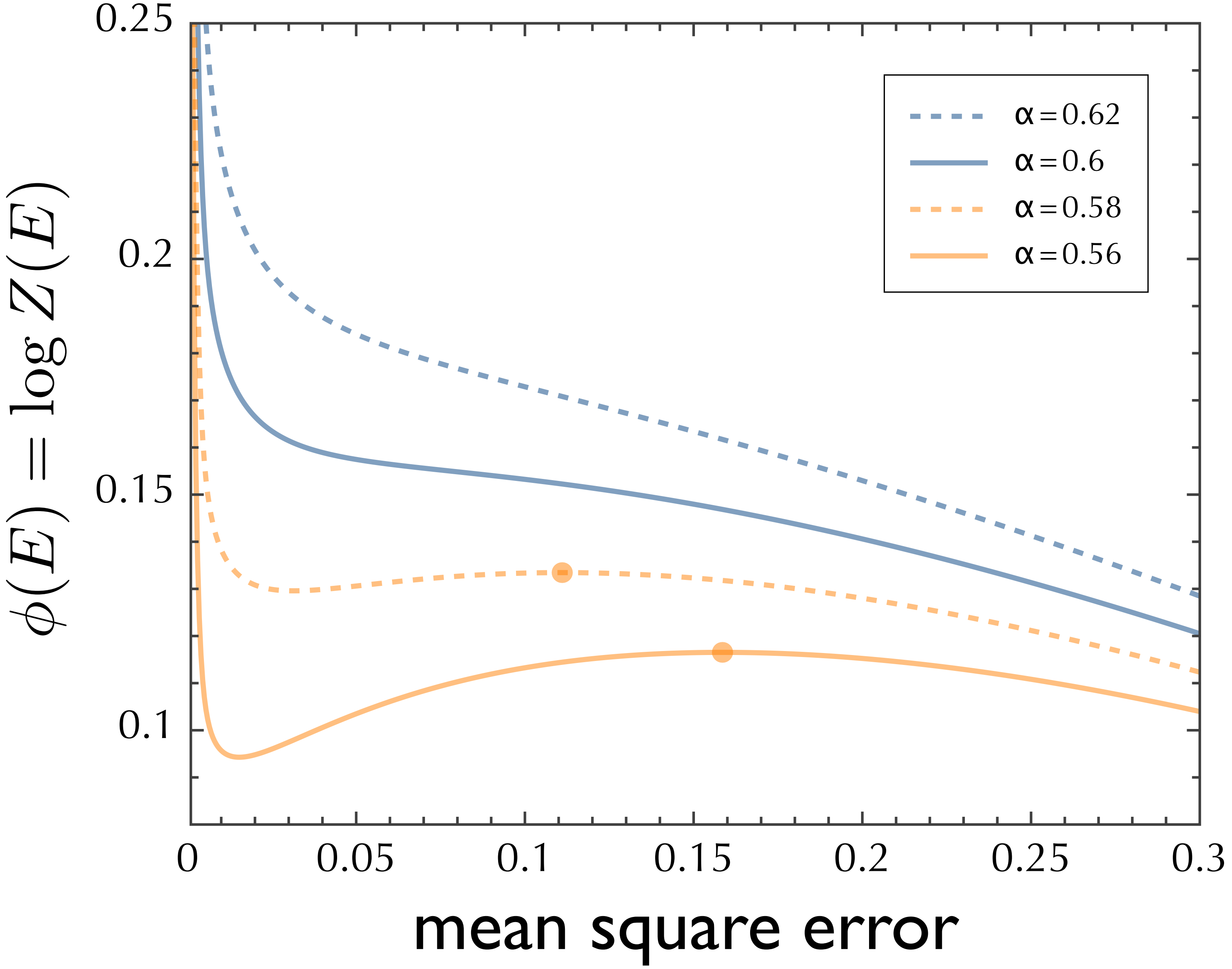
Order parameters:

$$Q = \frac{1}{N}\sum_i \langle x_i^2\rangle \qquad q = \frac{1}{N}\sum_i \langle x_i\rangle^2 \qquad m = \frac{1}{N}\sum_i x_i^0\langle x_i\rangle$$

Mean square error: $$E = \frac{1}{N}\sum_i \left(\langle x_i\rangle - x_i^0\right)^2 = q - 2m + \langle (x_i^0)^2\rangle_0$$

Computing the free entropy

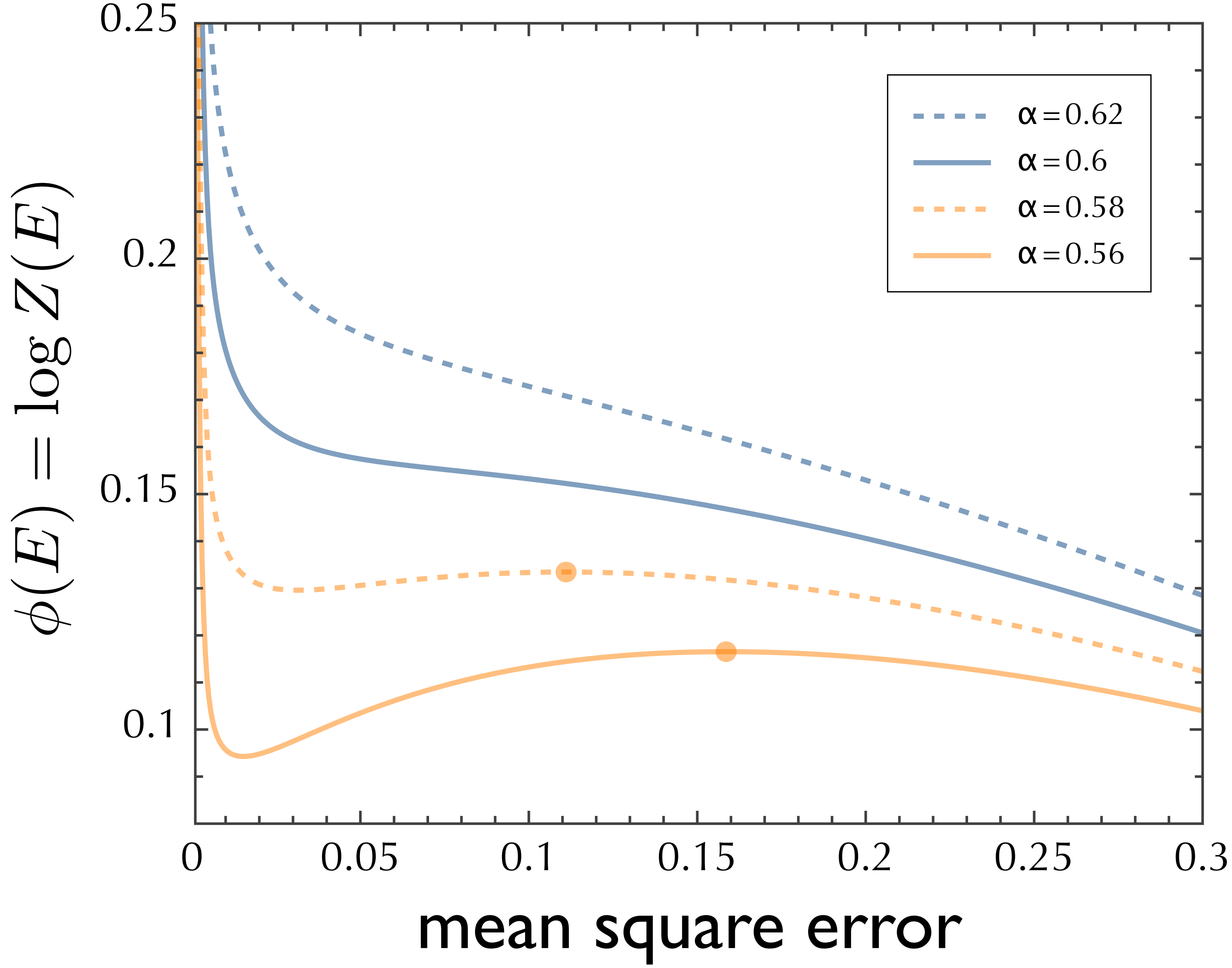
Example with ρ_0=0.4, and Φ_0 a Gaussian distribution with zero mean and unit variance



$$E = \frac{1}{N} \sum_i \left(\langle x_i \rangle - x_i^0 \right)^2$$

Computing the free entropy

Example with ρ$_0$=0.4, and Φ$_0$ a Gaussian distribution with zero mean and unit variance

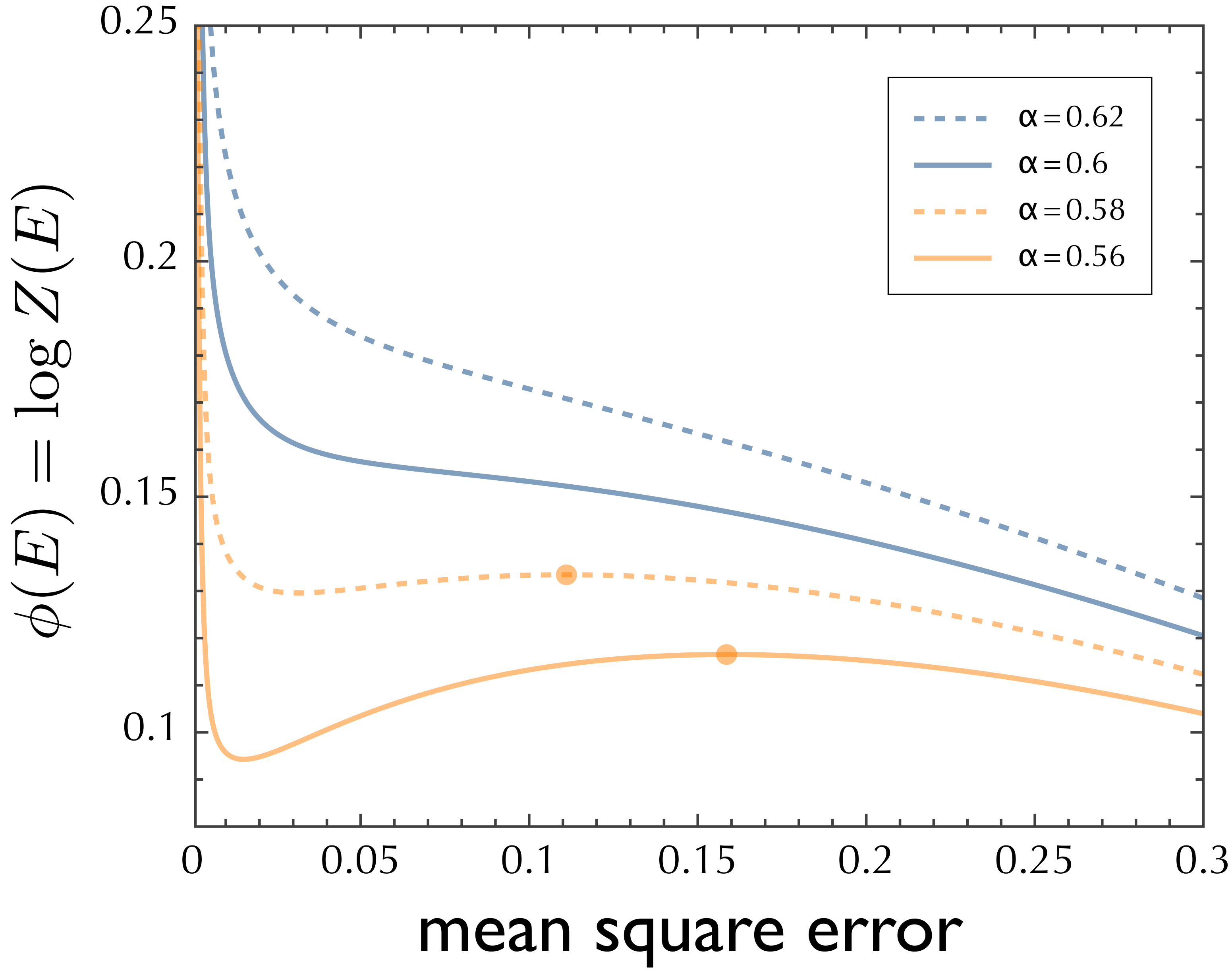


$$E = \frac{1}{N}\sum_i \left(\langle x_i \rangle - x_i^0\right)^2$$

- Maximum is at E=0 (as long as α>ρ0): Equilibrium behavior dominated by the original signal

Computing the free entropy

Example with ρ₀=0.4, and Φ₀ a Gaussian distribution with zero mean and unit variance

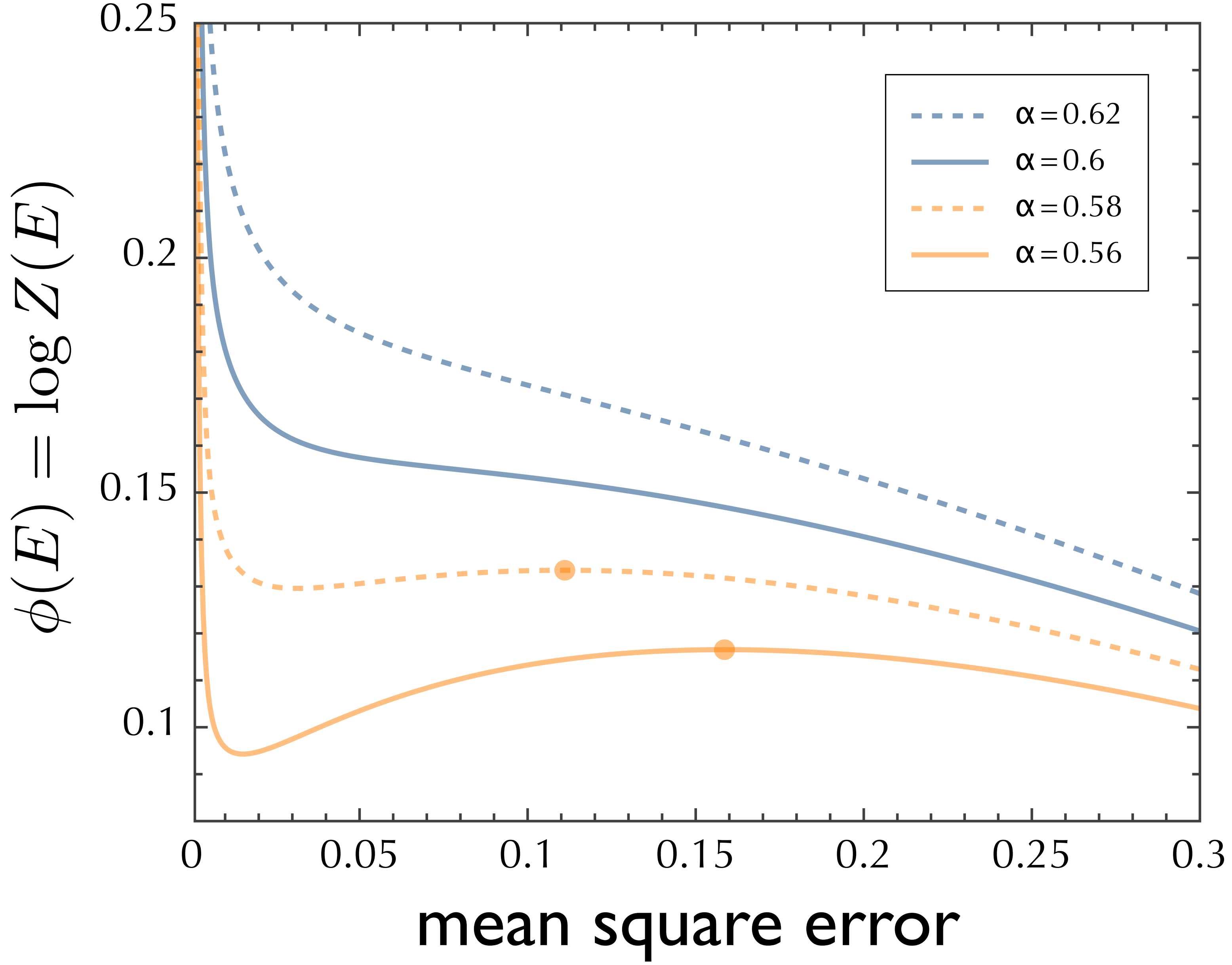


$$E = \frac{1}{N}\sum_i \left(\langle x_i\rangle - x_i^0\right)^2$$

- Maximum is at E=0 (as long as α>ρ0): Equilibrium behavior dominated by the original signal
- For α<0.58, a secondary maximum appears (meta-stable state): spinodal point

Computing the free entropy

Example with ρ_0=0.4, and Φ_0 a Gaussian distribution with zero mean and unit variance

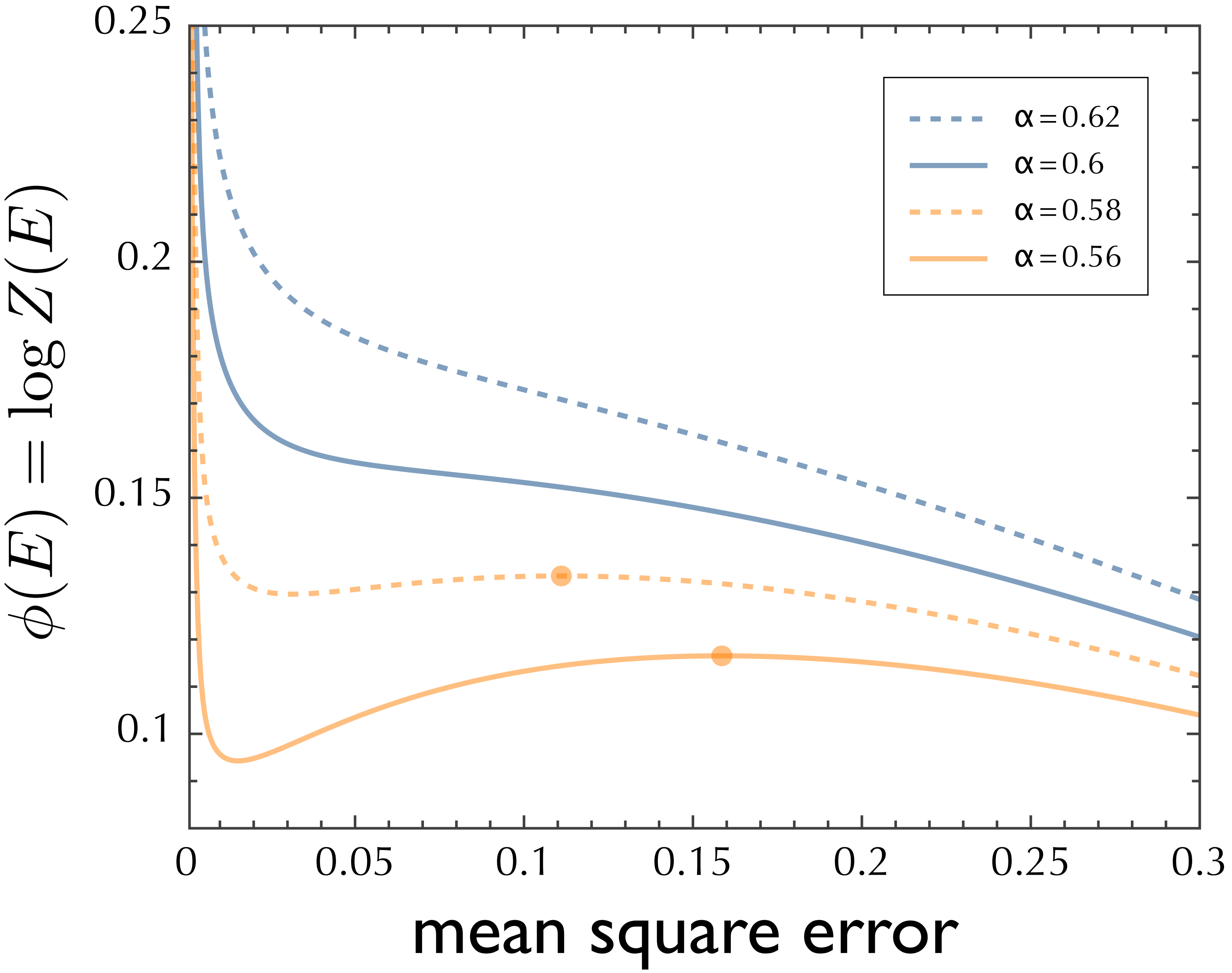


$$E = \frac{1}{N} \sum_i \left(\langle x_i \rangle - x_i^0 \right)^2$$

- Maximum is at E=0 (as long as α>ρ0): Equilibrium behavior dominated by the original signal
- For α<0.58, a secondary maximum appears (meta-stable state): spinodal point
- A steepest ascent dynamics starting from large E would reach the signal for α>0.58, but would stay block in the meta-stable state for α<0.58, even if the true equilibrium is at E=0.

Computing the free entropy

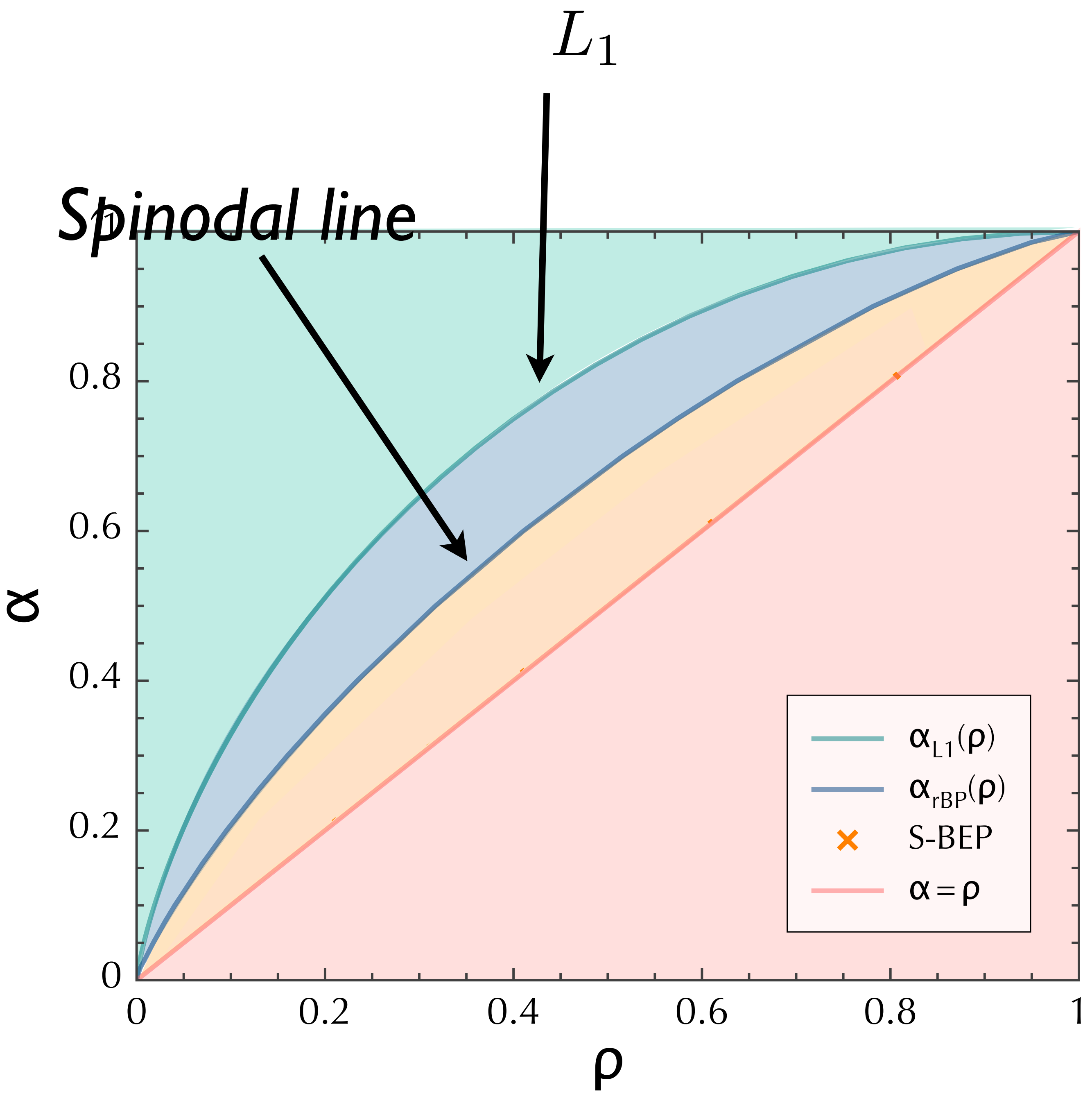
Example with ρ0=0.4, and Φ0 a Gaussian distribution with zero mean and unit variance

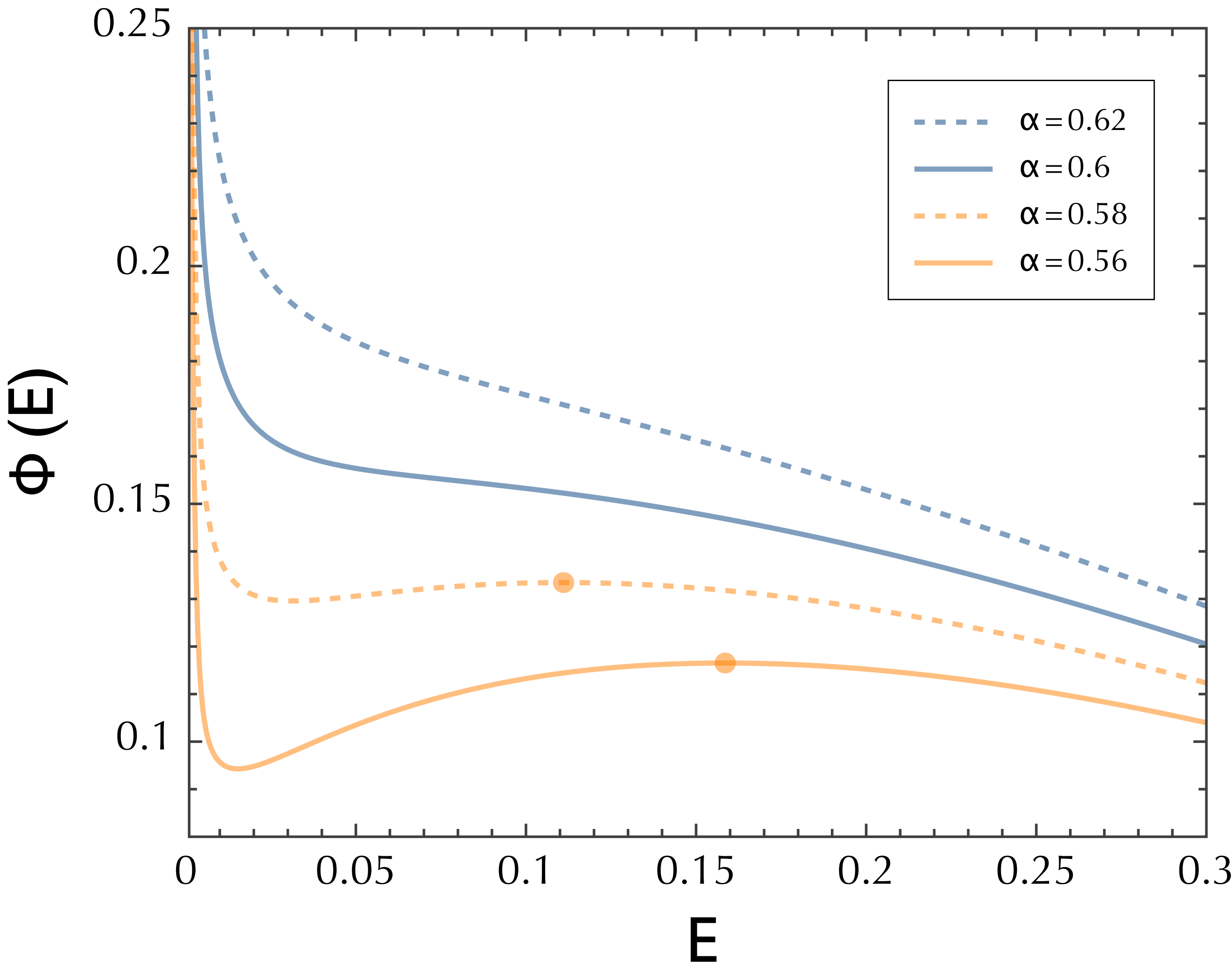


$$E = \frac{1}{N}\sum_i \left(\langle x_i \rangle - x_i^0\right)^2$$

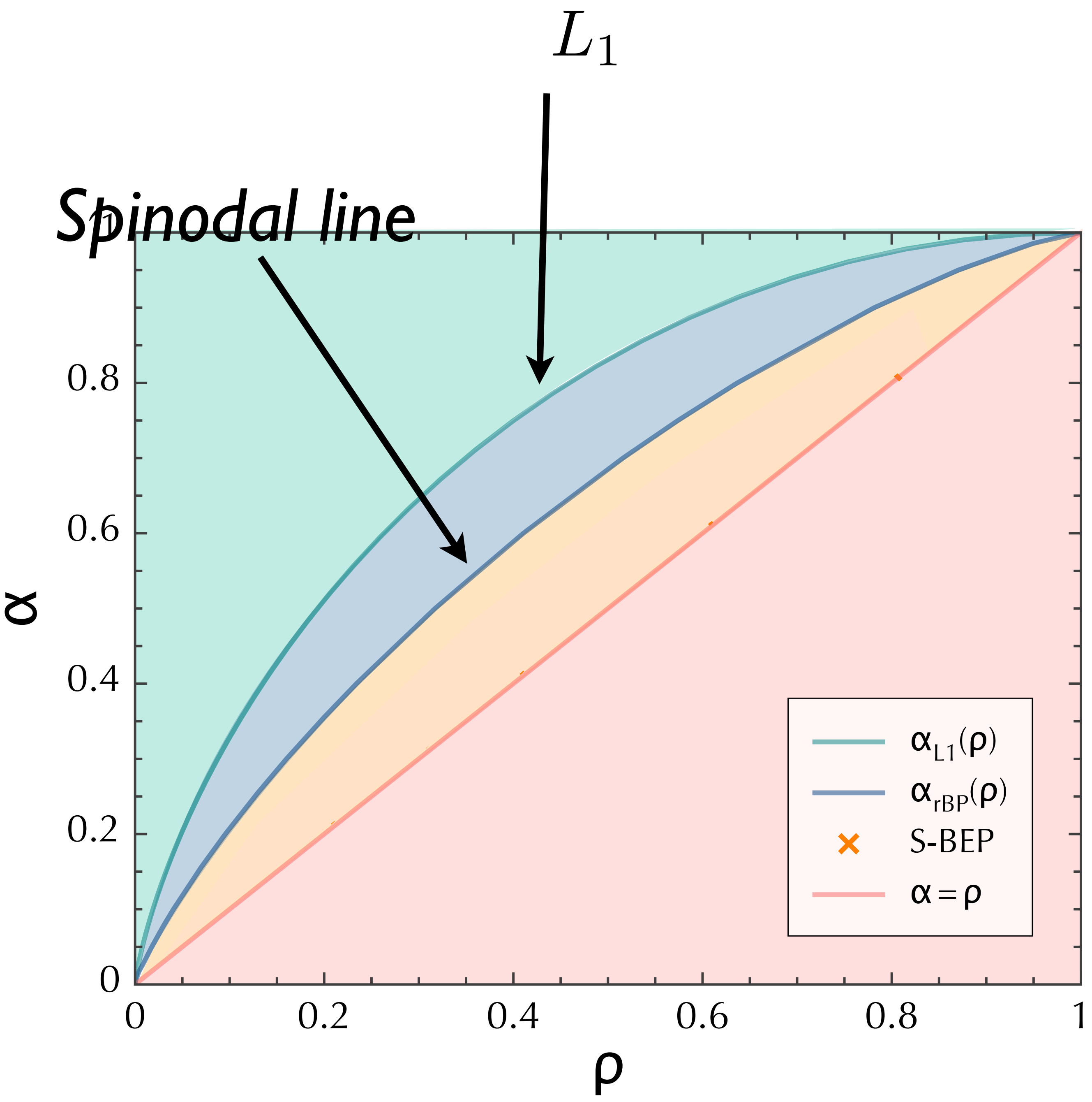
- Maximum is at E=0 (as long as α>ρ0): Equilibrium behavior dominated by the original signal
- For α<0.58, a secondary maximum appears (meta-stable state): spinodal point
- A steepest ascent dynamics starting from large E would reach the signal for α>0.58, but would stay block in the meta-stable state for α<0.58, even if the true equilibrium is at E=0.
- Similarity with metastable phase in first-order transition (supercooled liquids)

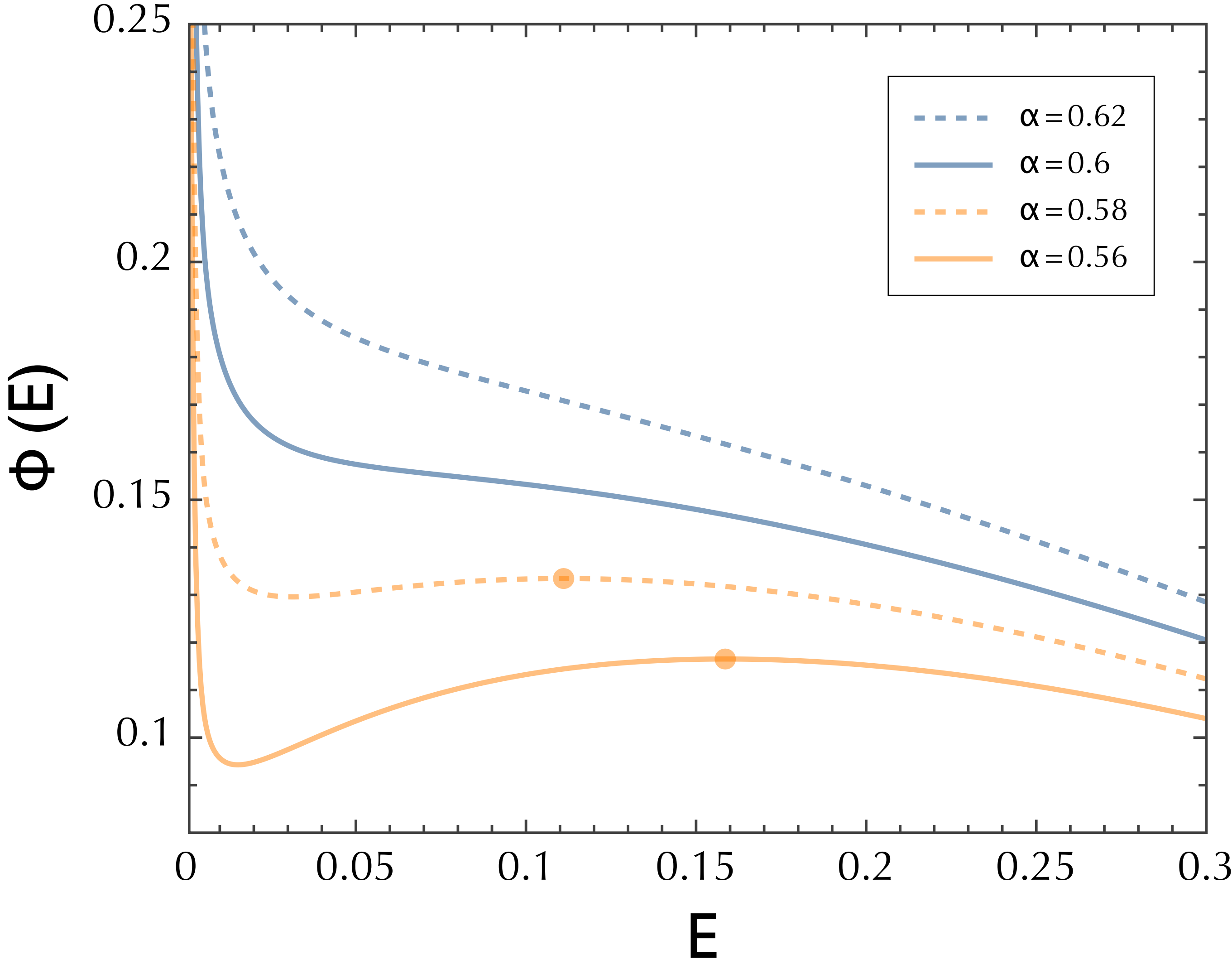
Computing the Phase Diagram





Computing the Phase Diagram





A steepest ascent of the free entropy should perform a perfect reconstruction until the spinodal line: This should be more efficient than L_1-minimization

A statistical physics approach to compressed sensing

One can use statistical physics tools for

1) Computing phase transitions analytically (reconstruction/non reconstruction, etc...)

Tools: Replica method from spin glass theory, etc...

II) Develop new algorithms, and design new matrices to improve on the L_1 state-of-the art.

Tools: Bethe-Peirls method/Belief propagation,
Mean field methods from stat-phys, Physics intuition, etc....

The Belief-Propagation algorithm (a sketchy description)

- NO averaging: work on a given problem
- Compute $f\left(\{\mathcal{P}_i(x_i)\}\right) = \log Z\left(\{\mathcal{P}_i(x_i)\},\right)$ the potential with constrained local probabilities (marginals) for each variable.
- Derive the recursion equation for by steepest ascent/descent:

$$\mathcal{P}_i^{t+1} = \nabla_{\mathcal{P}_i^t} f\left(\{\mathcal{P}_i^t\}\right)$$

The Belief-Propagation algorithm (a sketchy description)

- NO averaging: work on a given problem

- Compute $f\left(\{\mathcal{P}_i(x_i)\}\right) = \log Z\left(\{\mathcal{P}_i(x_i)\}\right)$ the potential with constrained local probabilities (marginals) for each variable.

- Derive the recursion equation for by steepest ascent/descent:

$$\mathcal{P}_i^{t+1} = \nabla_{\mathcal{P}_i^t} f\left(\{\mathcal{P}_i^t\}\right)$$

- This approach has been used :
  - Mean-field, Curie-Weiss, TAP (Thouless-Anderson-Palmer), or Cavity Method in Physics, and can be traced to Bethe-Peierls and Onsager ('35).
  - Belief Propagation in Artificial Intelligence (Pearl, '82)
  - Sum-Product in Error-Correcting-Codes (Gallager, '60)

How does BP works?

Gibbs free energy approach: $\log Z = \max_{\{\mathcal{P}(\vec{x})\}} f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right)$

With $f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right) = -\langle \log P(\vec{x}|\vec{y}) \rangle_{\mathcal{P}(\vec{x})} - \int d\vec{x} \mathcal{P}(\vec{x}) \log \mathcal{P}(\vec{x})$

How does BP works?

Gibbs free energy approach: $\log Z = \max_{\{\mathcal{P}(\vec{x})\}} f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right)$

With $f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right) = -\langle \log P(\vec{x}|\vec{y}) \rangle_{\mathcal{P}(\vec{x})} - \int d\vec{x} \mathcal{P}(\vec{x}) \log \mathcal{P}(\vec{x})$

Mean-Field $\Rightarrow$ $\mathcal{P}(\vec{x}) = \prod_i \mathcal{P}_i(\vec{x}_i)$

How does BP works?

Gibbs free energy approach: $\log Z = \max_{\{\mathcal{P}(\vec{x})\}} f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right)$

With $f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right) = -\langle \log P(\vec{x}|\vec{y}) \rangle_{\mathcal{P}(\vec{x})} - \int d\vec{x} \mathcal{P}(\vec{x}) \log \mathcal{P}(\vec{x})$

Mean-Field $\Rightarrow$ $\mathcal{P}(\vec{x}) = \prod_i \mathcal{P}_i(\vec{x}_i)$

Not correct
+Convergence problems

How does BP works?

Gibbs free energy approach: $\log Z = \max_{\{\mathcal{P}(\vec{x})\}} f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right)$

With $f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right) = -\langle \log P(\vec{x}|\vec{y})\rangle_{\mathcal{P}(\vec{x})} - \int d\vec{x}\mathcal{P}(\vec{x}) \log \mathcal{P}(\vec{x})$

Mean-Field $\Rightarrow$ $\mathcal{P}(\vec{x}) = \prod_i \mathcal{P}_i(\vec{x}_i)$

Not correct
+Convergence problems

Belief-Propagation$\Rightarrow$ $\mathcal{P}(\vec{x}) = \frac{\prod_{ij} \mathcal{P}_{ij}(\vec{x}_i, \vec{x}_j)}{\prod_i \mathcal{P}_i(\vec{x}_i)^{M-1}}$

How does BP works?

Gibbs free energy approach: $\log Z = \max_{\{\mathcal{P}(\vec{x})\}} f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right)$

With $f_{Gibbs}\left(\{\mathcal{P}(\vec{x})\}\right) = -\langle \log P(\vec{x}|\vec{y})\rangle_{\mathcal{P}(\vec{x})} - \int d\vec{x}\mathcal{P}(\vec{x}) \log \mathcal{P}(\vec{x})$

Mean-Field $\Rightarrow$ $\mathcal{P}(\vec{x}) = \prod_i \mathcal{P}_i(\vec{x}_i)$

Not correct
+Convergence problems

Belief-Propagation$\Rightarrow$ $\mathcal{P}(\vec{x}) = \frac{\prod_{ij} \mathcal{P}_{ij}(\vec{x}_i, \vec{x}_j)}{\prod_i \mathcal{P}_i(\vec{x}_i)^{M-1}}$

(asymptotically)
exact in CS
with random matrices

How does BP works?

Simplification thanks to the dense matrix limit:
Projection on first two moments is enough :

$$f\left(\{\mathcal{P}_i(x_i), \mathcal{P}_{ij}(x_i, x_j)\}\right) \quad \Longrightarrow \quad f\left(\{\langle x_i\rangle, \langle x_i^2\rangle\}\right)$$

Belief-Propagation equations ⇒

$$\begin{cases} \langle x_i\rangle^{t+1} = \langle x_i\rangle^{t} + \dfrac{\partial f}{\partial\langle x_i\rangle} \\ \langle x_i^2\rangle^{t+1} = \langle x_i^2\rangle^{t} + \dfrac{\partial f}{\partial\langle x_i^2\rangle} \end{cases}$$

The Belief-Propagation algorithm

Iterate these variables

$$
\begin{aligned}
U_i^{(t+1)} &= \frac{\alpha}{M}\sum_\mu \frac{1}{\Delta_\mu + \gamma^{(t)}} \\
V_i^{(t+1)} &= \sum_\mu F_{\mu i}\frac{(y_\mu - \alpha_\mu^{(t)})}{\Delta_\mu + \gamma_\mu^{(t)}} + f_a\left(U_i^{(t)}, V_i^{(t)}\right)\frac{\alpha}{M}\sum_\mu \frac{1}{\Delta_\mu + \gamma^{(t)}} \\
\alpha_\mu^{(t+1)} &= \sum_i F_{\mu i} f_a(U_i^{(t+1)}, V_i^{(t+1)}) - \frac{(y_\mu - \alpha_\mu^{(t)})}{\Delta_\mu + \gamma^{(t)}}\frac{1}{N}\sum_i \frac{\partial f_a}{\partial Y}\left(U_i^{(t+1)}, V_i^{(t+1)}\right) \\
\gamma^{(t+1)} &= \frac{1}{N}\sum_i f_c(U_i^{(t+1)}, V_i^{(t+1)})
\end{aligned}
$$

Using these functions:

$$
\begin{aligned}
f_a(X,Y) &= \left[\frac{\rho Y}{(1+X)^{3/2}} e^{Y^2/(2(1+X))}\right]\left[1-\rho+\frac{\rho}{(1+X)^{1/2}} e^{Y^2/(2(1+X))}\right]^{-1} \\
f_c(X,Y) &= \left[\frac{\rho}{(1+X)^{3/2}} e^{Y^2/(2(1+X))}\left(1+\frac{Y^2}{1+X}\right)\right]\left[1-\rho+\frac{\rho}{(1+X)^{1/2}} e^{Y^2/(2(1+X))}\right]^{-1} - f_a(X,Y)^2
\end{aligned}
$$

And finally at the end:

$$
\langle x_i \rangle = f_a\left(U_i, V_i\right)
$$

The Belief-Propagation algorithm

Iterate these variables

$$\begin{aligned}
U_i^{(t+1)} &= \frac{\alpha}{M} \sum_\mu \frac{1}{\Delta_\mu + \gamma^{(t)}} \\
V_i^{(t+1)} &= \sum_\mu F_{\mu i} \frac{(y_\mu - \alpha_\mu^{(t)})}{\Delta_\mu + \gamma_\mu^{(t)}} + f_a\left(U_i^{(t)}, V_i^{(t)}\right) \frac{\alpha}{M} \sum_\mu \frac{1}{\Delta_\mu + \gamma^{(t)}} \\
\alpha_\mu^{(t+1)} &= \sum_i F_{\mu i} f_a(U_i^{(t+1)}, V_i^{(t+1)}) - \frac{(y_\mu - \alpha_\mu^{(t)})}{\Delta_\mu + \gamma^{(t)}} \frac{1}{N} \sum_i \frac{\partial f_a}{\partial Y}\left(U_i^{(t+1)}, V_i^{(t+1)}\right) \\
\gamma^{(t+1)} &= \frac{1}{N} \sum_i f_c(U_i^{(t+1)}, V_i^{(t+1)})
\end{aligned}$$

Simple algebraic operations!

Using these functions:

$$\begin{aligned}
f_a(X,Y) &= \left[\frac{\rho Y}{(1+X)^{3/2}} e^{Y^2/(2(1+X))}\right] \left[1 - \rho + \frac{\rho}{(1+X)^{1/2}} e^{Y^2/(2(1+X))}\right]^{-1} \\
f_c(X,Y) &= \left[\frac{\rho}{(1+X)^{3/2}} e^{Y^2/(2(1+X))} \left(1 + \frac{Y^2}{1+X}\right)\right] \left[1 - \rho + \frac{\rho}{(1+X)^{1/2}} e^{Y^2/(2(1+X))}\right]^{-1} - f_a(X,Y)^2
\end{aligned}$$

And finally at the end:

$$\langle x_i \rangle = f_a\left(U_i, V_i\right)$$

The Belief-Propagation algorithm

Iterate these variables

$$
\begin{aligned}
U_i^{(t+1)} &= \frac{\alpha}{M}\sum_\mu \frac{1}{\Delta_\mu+\gamma^{(t)}} \\
V_i^{(t+1)} &= \sum_\mu F_{\mu i}\frac{(y_\mu-\alpha_\mu^{(t)})}{\Delta_\mu+\gamma_\mu^{(t)}} + f_a\left(U_i^{(t)},V_i^{(t)}\right)\frac{\alpha}{M}\sum_\mu\frac{1}{\Delta_\mu+\gamma^{(t)}} \\
\alpha_\mu^{(t+1)} &= \sum_i F_{\mu i} f_a(U_i^{(t+1)},V_i^{(t+1)}) - \frac{(y_\mu-\alpha_\mu^{(t)})}{\Delta_\mu+\gamma^{(t)}}\frac{1}{N}\sum_i\frac{\partial f_a}{\partial Y}\left(U_i^{(t+1)},V_i^{(t+1)}\right) \\
\gamma^{(t+1)} &= \frac{1}{N}\sum_i f_c(U_i^{(t+1)},V_i^{(t+1)})
\end{aligned}
$$

Simple algebraic operations!

Using these functions:

$$
\begin{aligned}
f_a(X,Y) &= \left[\frac{\rho Y}{(1+X)^{3/2}}e^{Y^2/(2(1+X))}\right]\left[1-\rho+\frac{\rho}{(1+X)^{1/2}}e^{Y^2/(2(1+X))}\right]^{-1} \\
f_c(X,Y) &= \left[\frac{\rho}{(1+X)^{3/2}}e^{Y^2/(2(1+X))}\left(1+\frac{Y^2}{1+X}\right)\right]\left[1-\rho+\frac{\rho}{(1+X)^{1/2}}e^{Y^2/(2(1+X))}\right]^{-1} - f_a(X,Y)^2
\end{aligned}
$$

And finally at the end:

Complexity is O(N^2×convergence time)

$$\langle x_i\rangle = f_a\left(U_i,V_i\right)$$

The Belief-Propagation algorithm

Iterate these variables

$$
\begin{aligned}
U_i^{(t+1)} &= \frac{\alpha}{M}\sum_\mu \frac{1}{\Delta_\mu + \gamma^{(t)}} \\
V_i^{(t+1)} &= \sum_\mu F_{\mu i} \frac{(y_\mu - \alpha_\mu^{(t)})}{\Delta_\mu + \gamma_\mu^{(t)}} + f_a\left(U_i^{(t)}, V_i^{(t)}\right) \frac{\alpha}{M}\sum_\mu \frac{1}{\Delta_\mu + \gamma^{(t)}} \\
\alpha_\mu^{(t+1)} &= \sum_i F_{\mu i} f_a(U_i^{(t+1)}, V_i^{(t+1)}) - \frac{(y_\mu - \alpha_\mu^{(t)})}{\Delta_\mu + \gamma^{(t)}} \frac{1}{N}\sum_i \frac{\partial f_a}{\partial Y}\left(U_i^{(t+1)}, V_i^{(t+1)}\right) \\
\gamma^{(t+1)} &= \frac{1}{N}\sum_i f_c(U_i^{(t+1)}, V_i^{(t+1)})
\end{aligned}
$$

Simple algebraic operations!

Using these functions:

$$
\begin{aligned}
f_a(X,Y) &= \left[\frac{\rho Y}{(1+X)^{3/2}} e^{Y^2/(2(1+X))}\right]\left[1-\rho+\frac{\rho}{(1+X)^{1/2}}e^{Y^2/(2(1+X))}\right]^{-1} \\
f_c(X,Y) &= \left[\frac{\rho}{(1+X)^{3/2}} e^{Y^2/(2(1+X))}\left(1+\frac{Y^2}{1+X}\right)\right]\left[1-\rho+\frac{\rho}{(1+X)^{1/2}}e^{Y^2/(2(1+X))}\right]^{-1} - f_a(X,Y)^2
\end{aligned}
$$

And finally at the end:

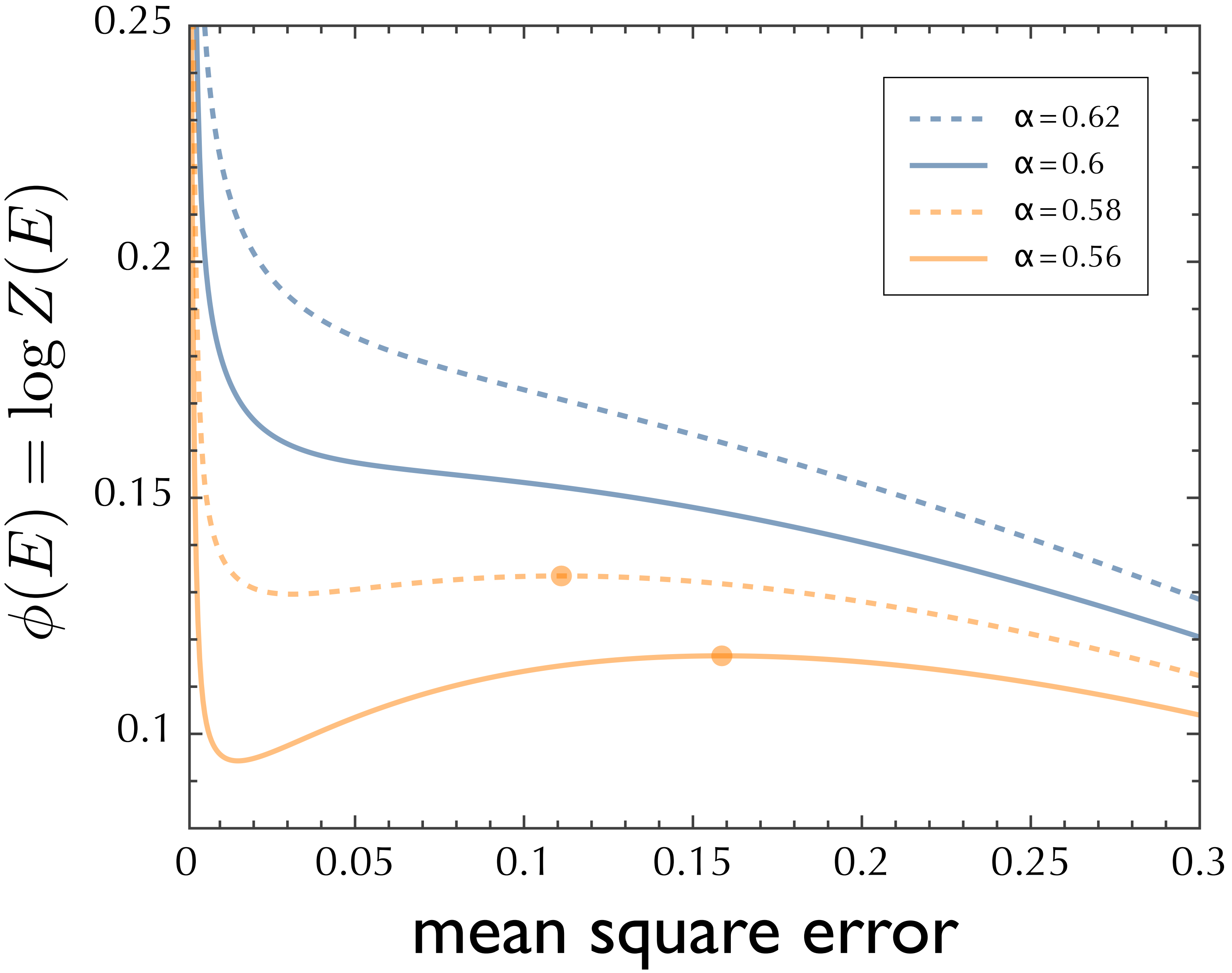
$$\langle x_i \rangle = f_a\left(U_i, V_i\right)$$

Complexity is O(N^2×convergence time)

http://aspics.krzakala.org

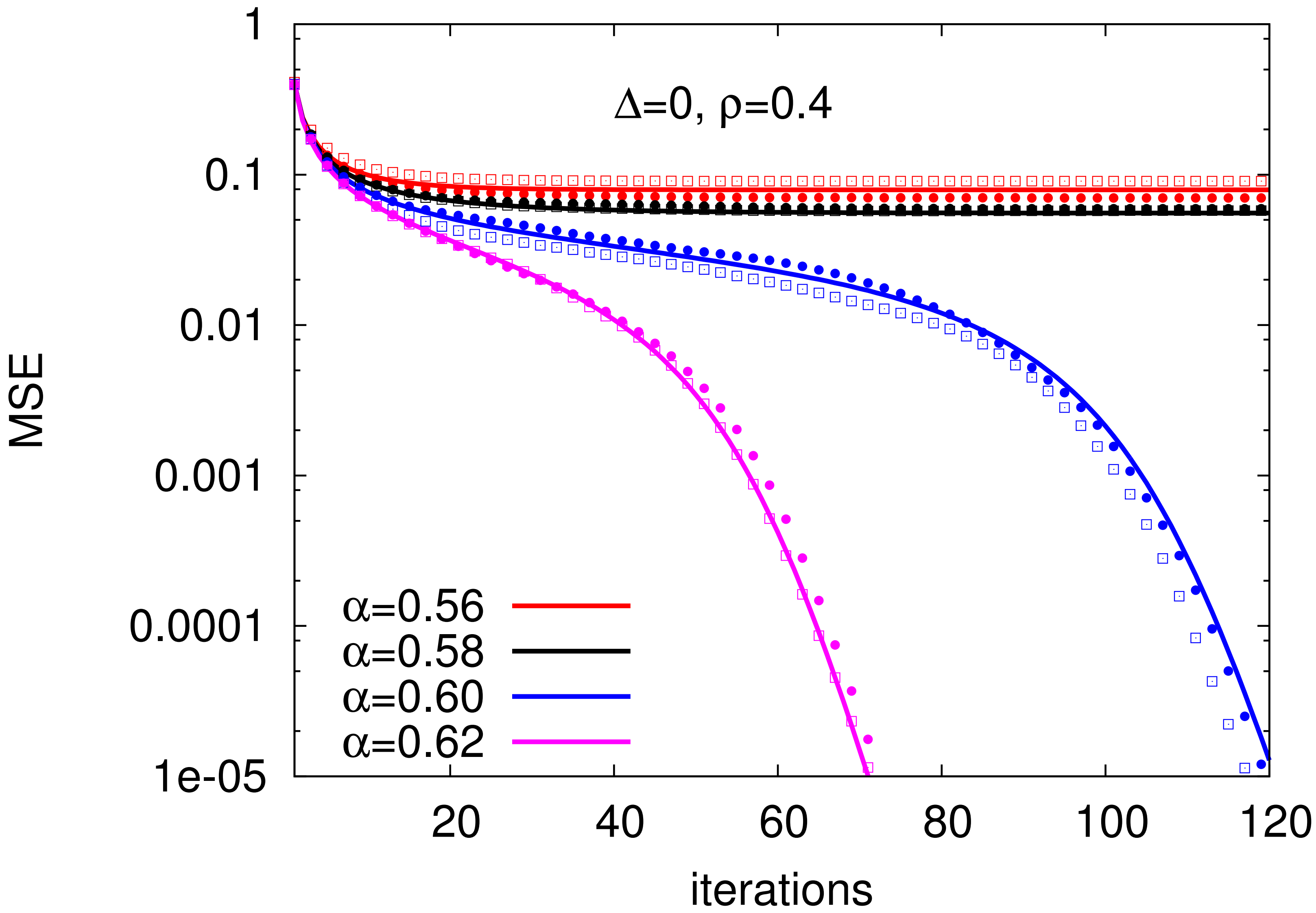
http://k11p.sourceforge.net/home.html

Steepest ascent of the free entropy



$$E = \frac{1}{N} \sum_i \left(\langle x_i \rangle - x_i^0 \right)^2$$

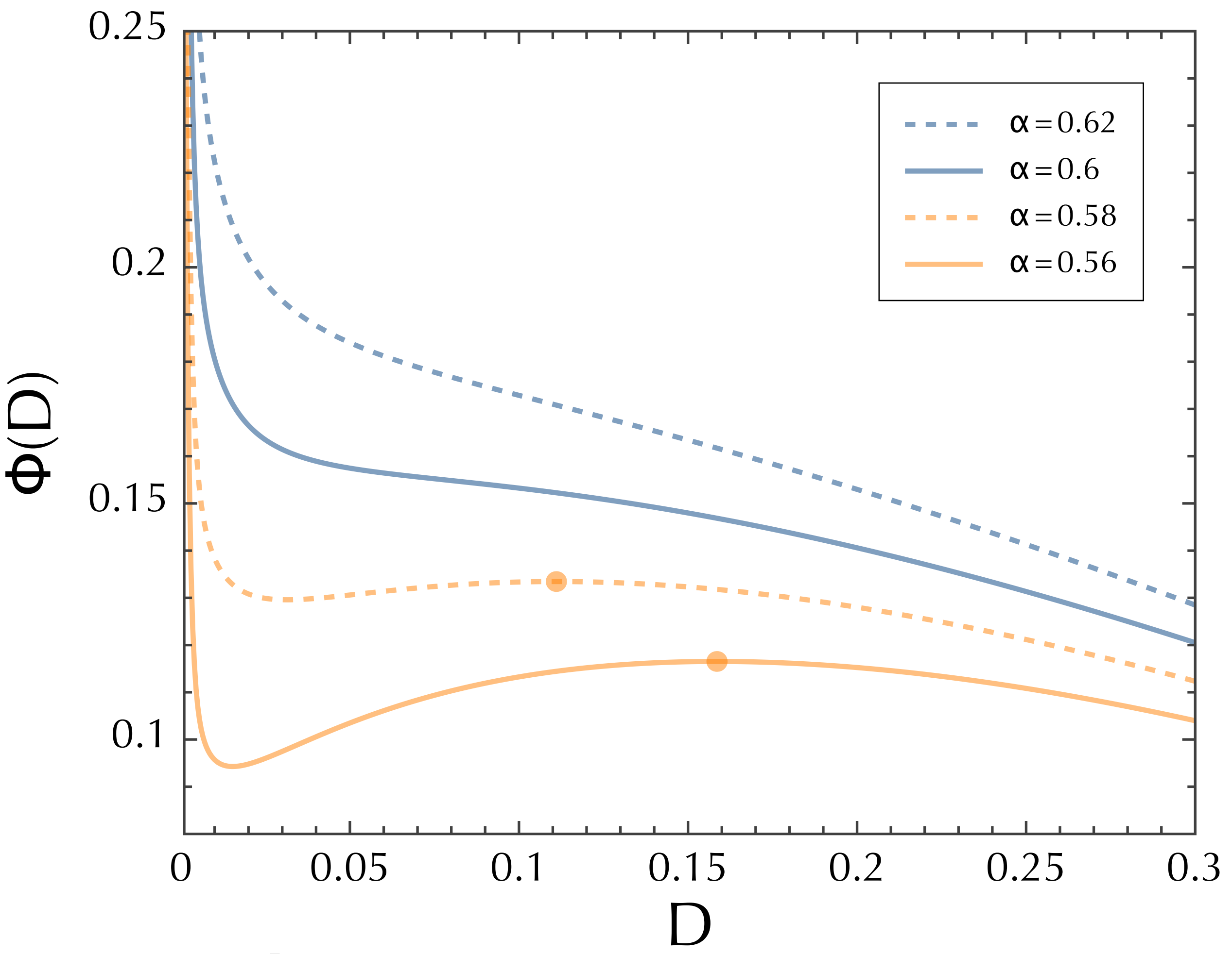
Steepest ascent of the free entropy



$$E = \frac{1}{N} \sum_i \left(\langle x_i \rangle - x_i^0 \right)^2$$

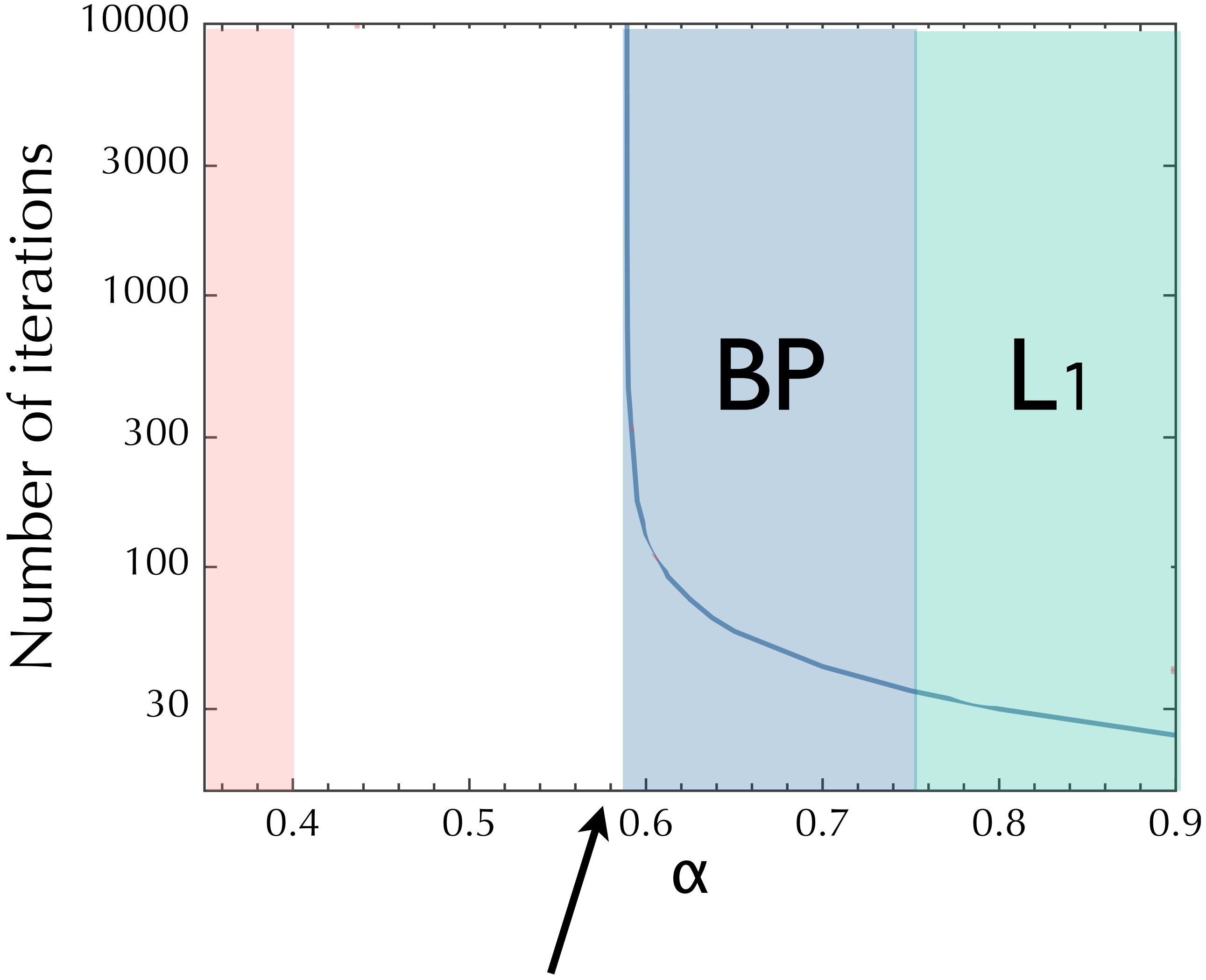
Replica (lines) vs Algo (points)

Thermodynamic potential



distance to native state

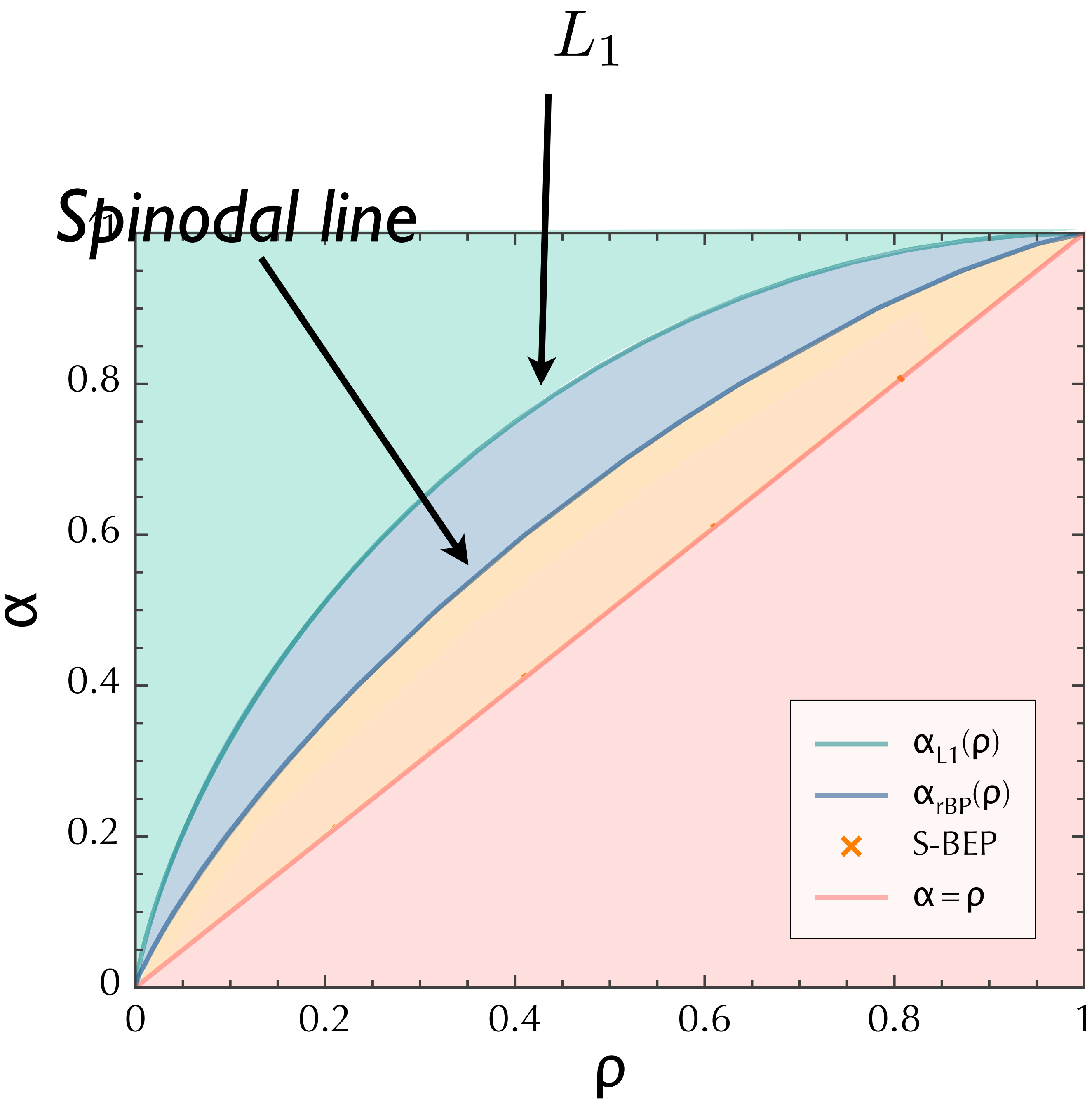
BP convergence time



Spinodal transition

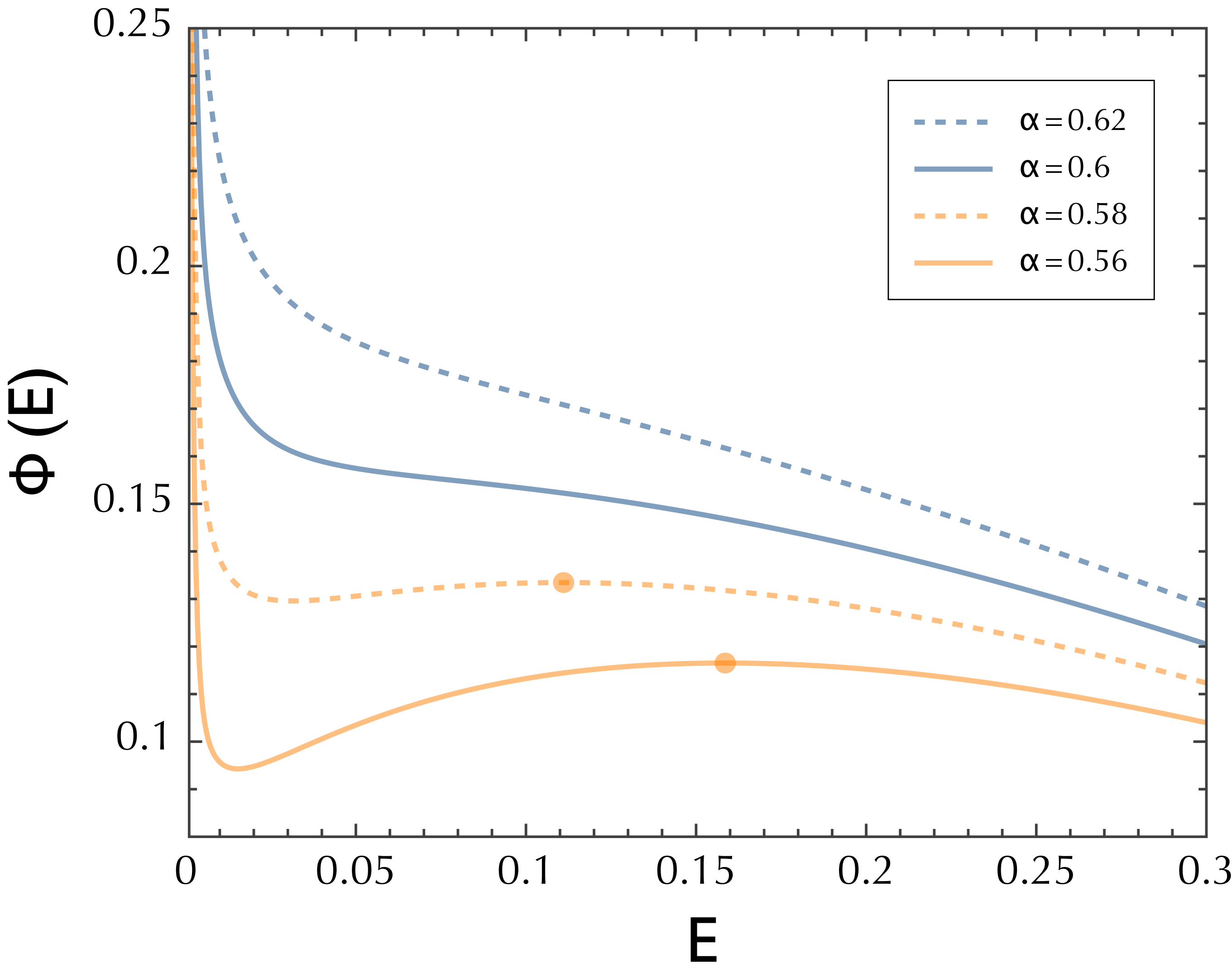
- Maximum is at E=0 (as long as α>ρ0): Equilibrium behavior dominated by the original signal
- For α<0.58, a secondary maximum appears (meta-stable state): spinodal point
- A steepest ascent dynamics starting from large E reaches the signal for α>0.58, but stay blocked in the meta-stable state for α<0.58, even if the true maximum is at E=0.
- Similarity with the physics of supercooled liquids

Computing the Phase Diagram

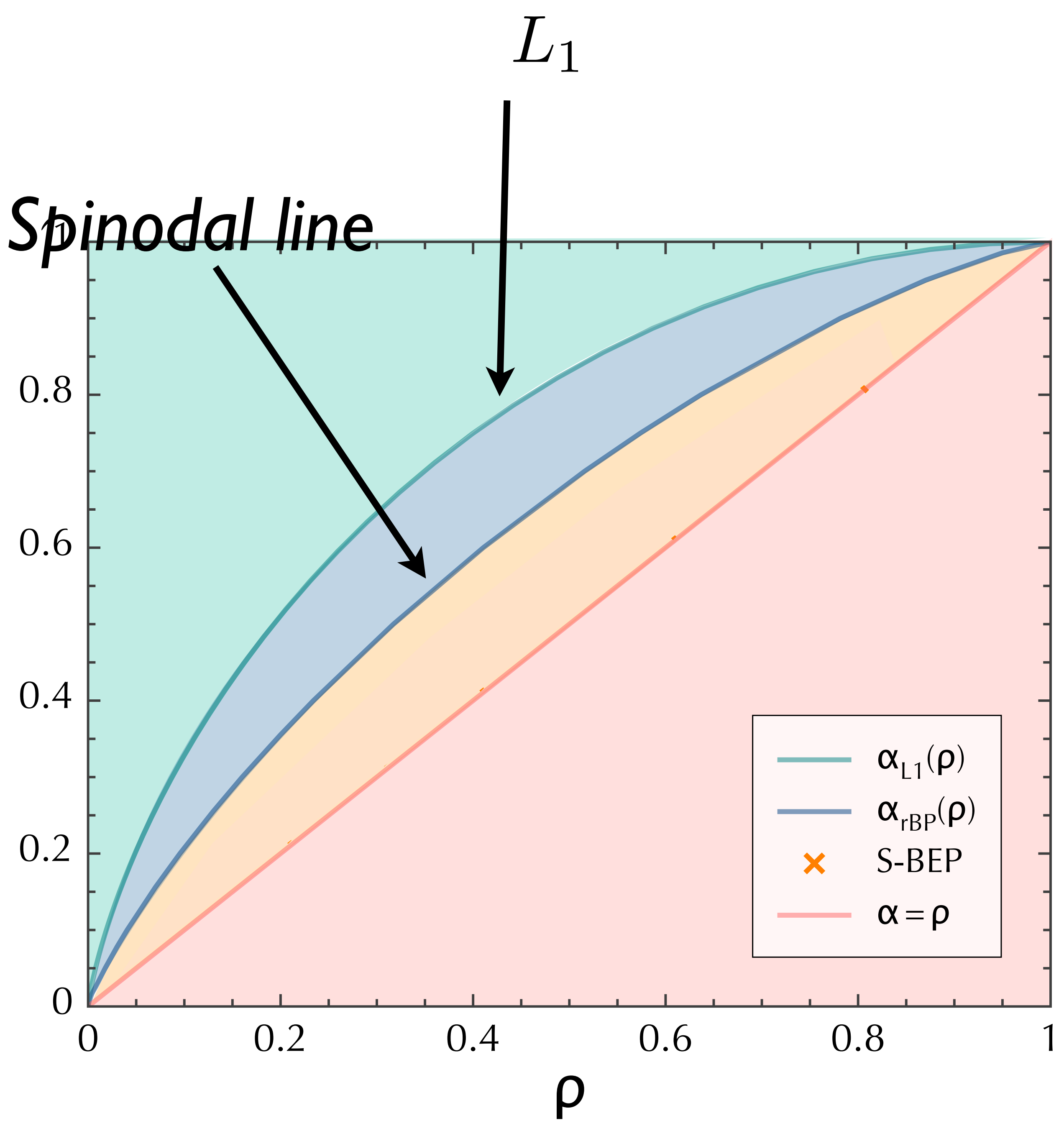


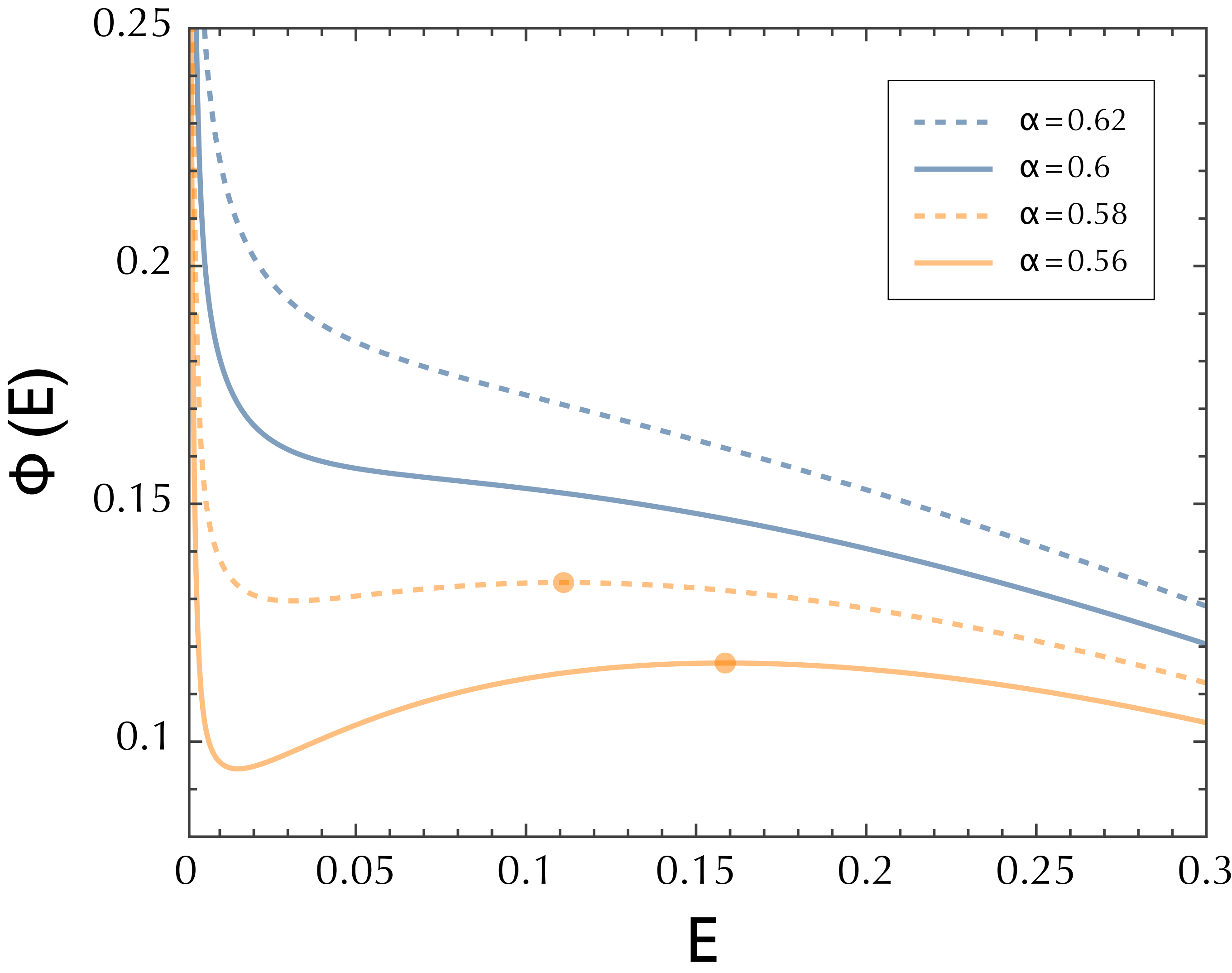
L_1

Spinodal line



Computing the Phase Diagram





A steepest ascent of the free entropy allows
a perfect reconstruction until the spinodal line.
This is more efficient than L_1-minimization

BP + probabilistic approach

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

BP + probabilistic approach

- Efficient and fast
- Robust to many type of noises (measurement, matrix coefficients, etc..)
- Very flexible (more information can be put in the prior, correlated variables, etc...)

GOOD!

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

BP + probabilistic approach

- Efficient and fast

GOOD!

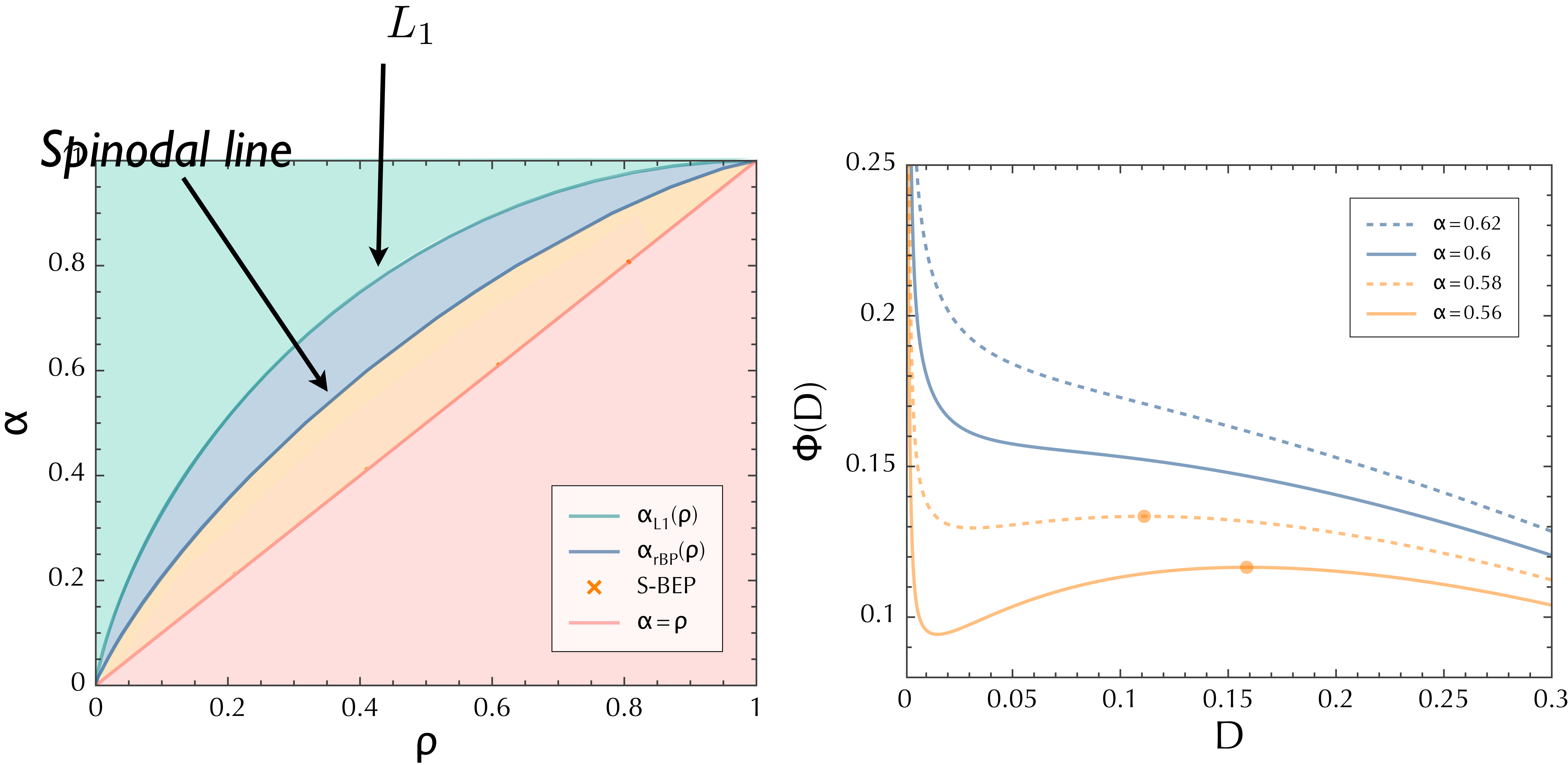
- Robust to many type of noises (measurement, matrix coefficients, etc..)
- Very flexible (more information can be put in the prior, correlated variables, etc...)

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

- Still not optimal

BAD!

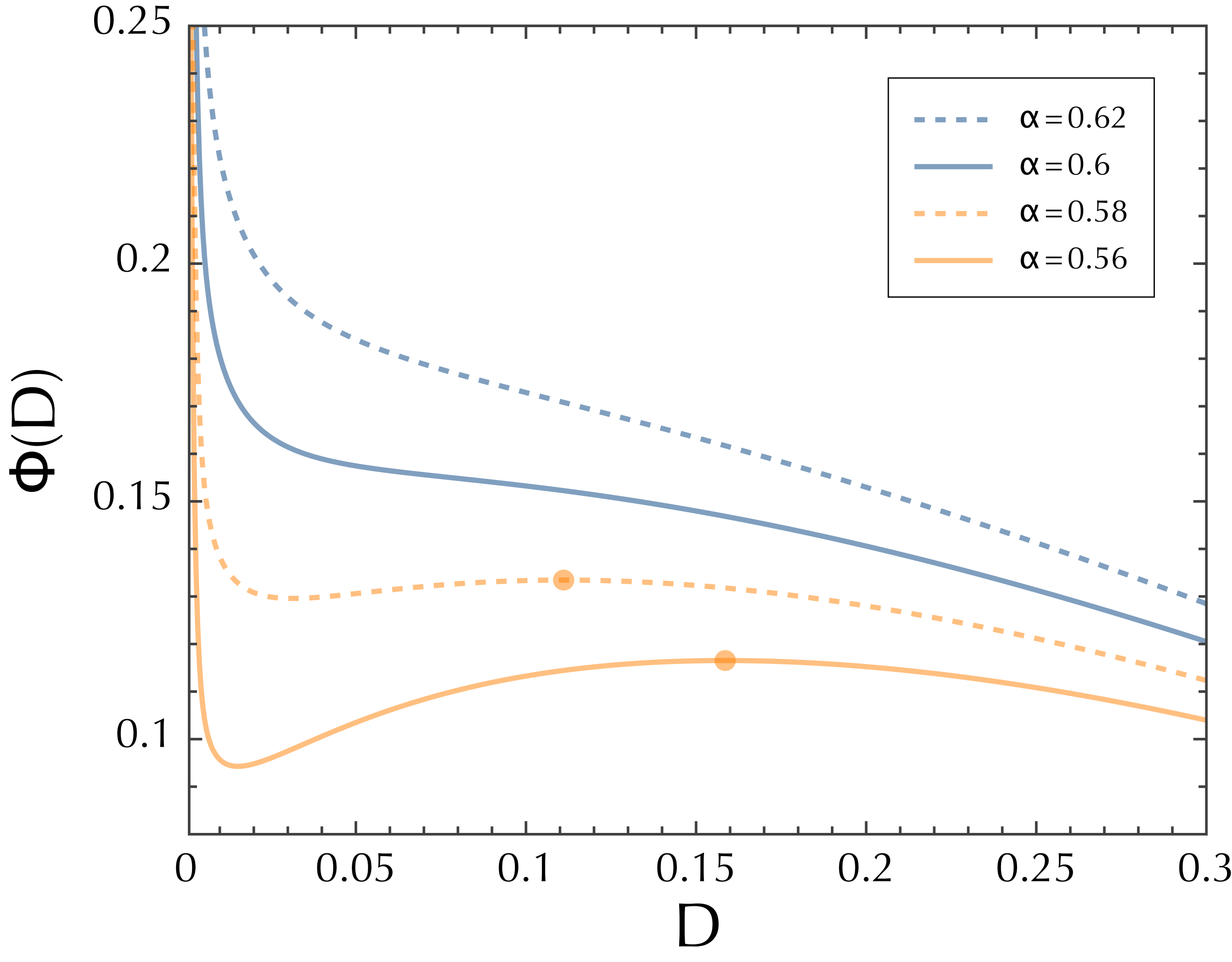
This is good, but not good enough



The dynamics is stuck in a metastable state, just as a liquid cooled too fast remains in a supercooled liquid state instead of crystalizing

This is good, but not good enough

How to pass the spinodal point?



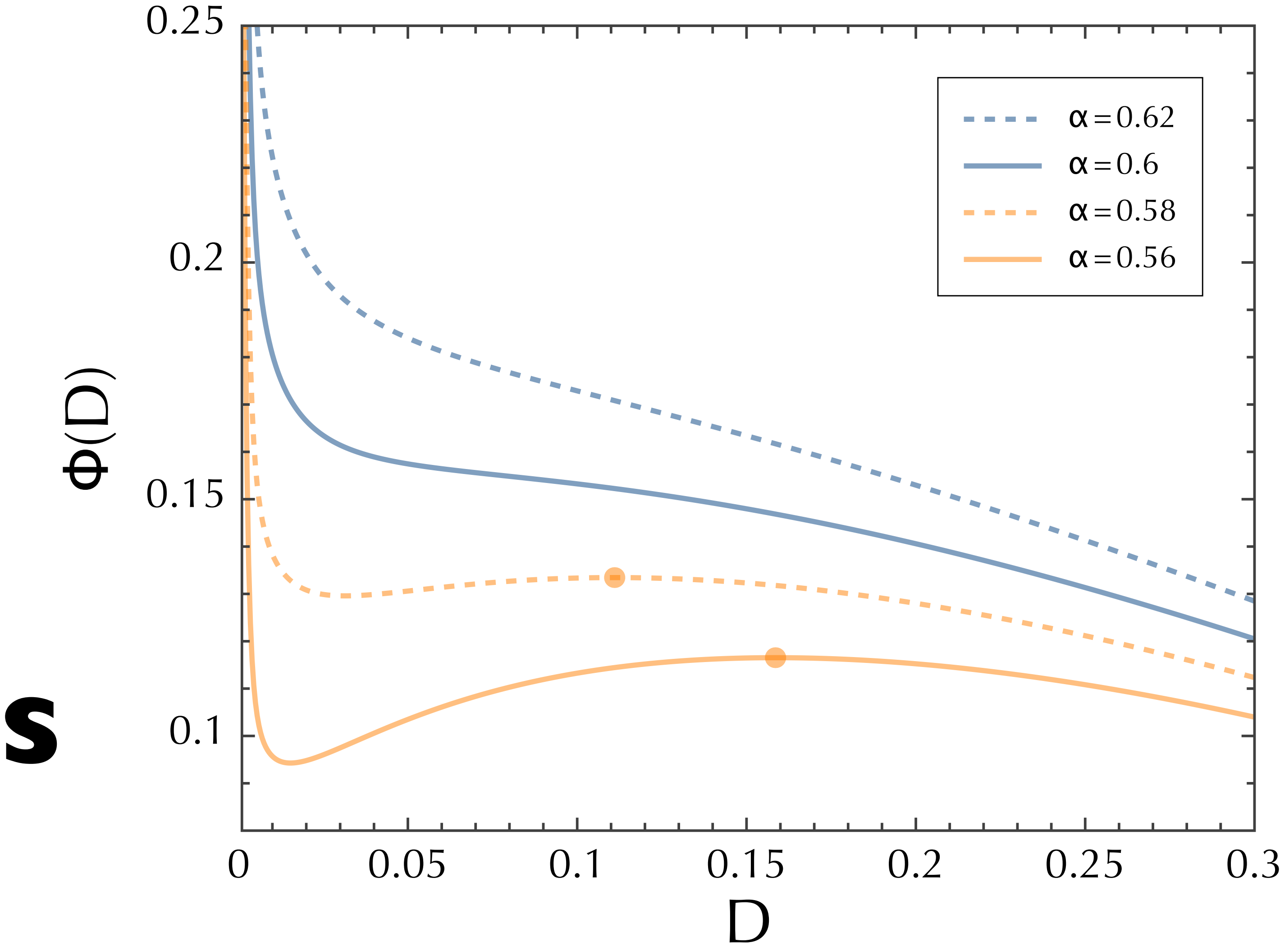
The dynamics is stuck in a metastable state, just as a liquid cooled too fast remains in a supercooled liquid state instead of crystalizing

This is good, but not good enough

How to pass the spinodal point?

By nucleation!

Special design of “seeded” matrices



The dynamics is stuck in a metastable state, just as a liquid cooled too fast remains in a supercooled liquid state instead of crystalizing

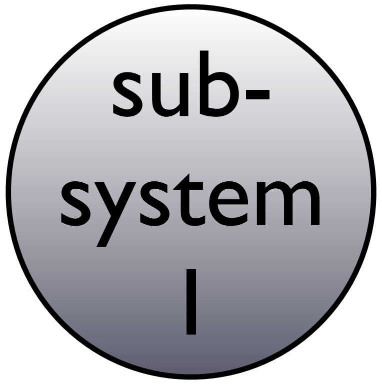
A coupled one-dimensional system:

A coupled one-dimensional system:

1) Create many sub-systems

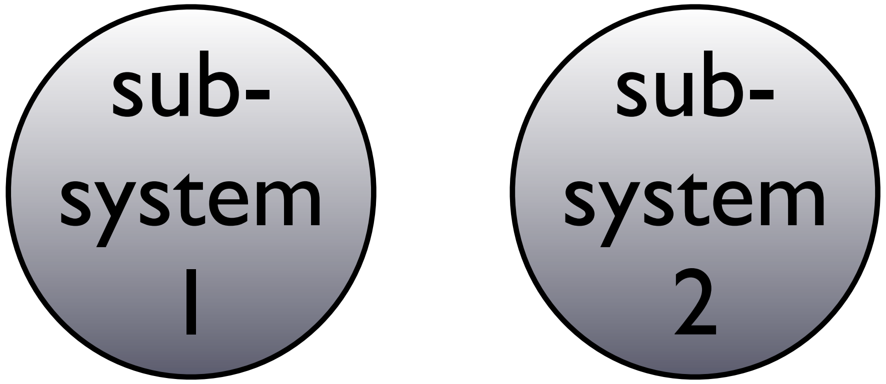
A coupled one-dimensional system:

1) Create many sub-systems



A coupled one-dimensional system:

1) Create many sub-systems

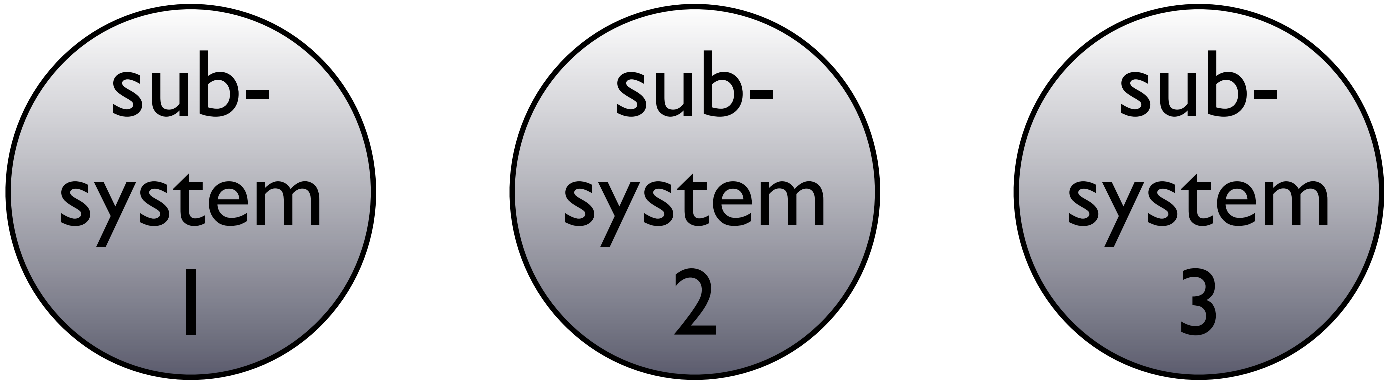


sub-system 1

sub-system 2

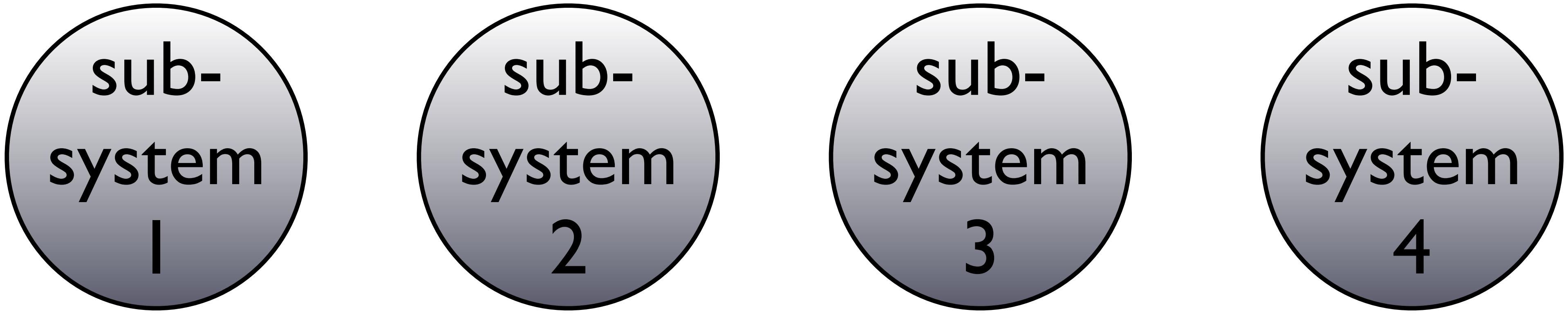
A coupled one-dimensional system:

1) Create many sub-systems



A coupled one-dimensional system:

1) Create many sub-systems



sub-system 1

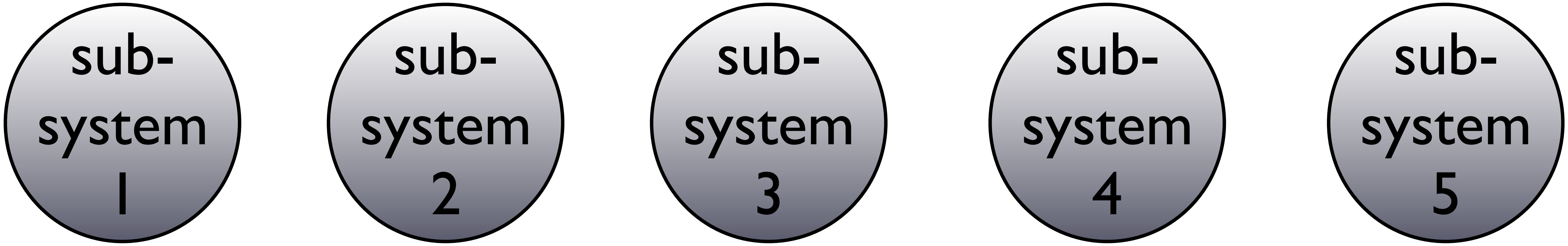
sub-system 2

sub-system 3

sub-system 4

A coupled one-dimensional system:

1) Create many sub-systems



sub-system 1

sub-system 2

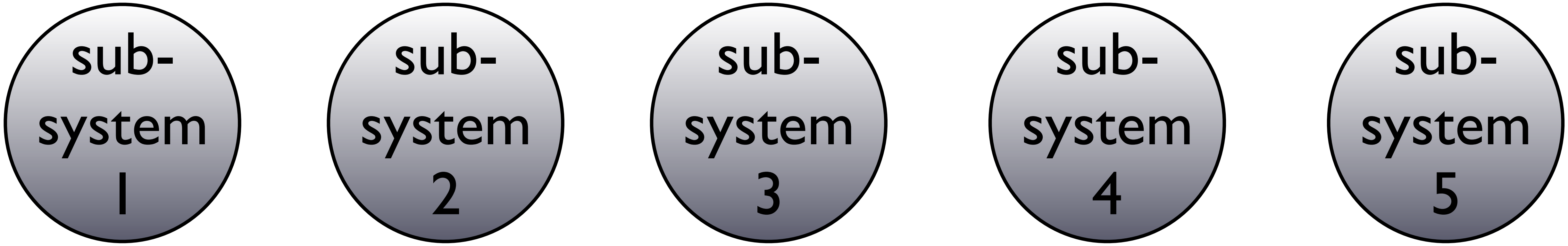
sub-system 3

sub-system 4

sub-system 5

A coupled one-dimensional system:

2) Add a first neighbor coupling



sub-system 1

sub-system 2

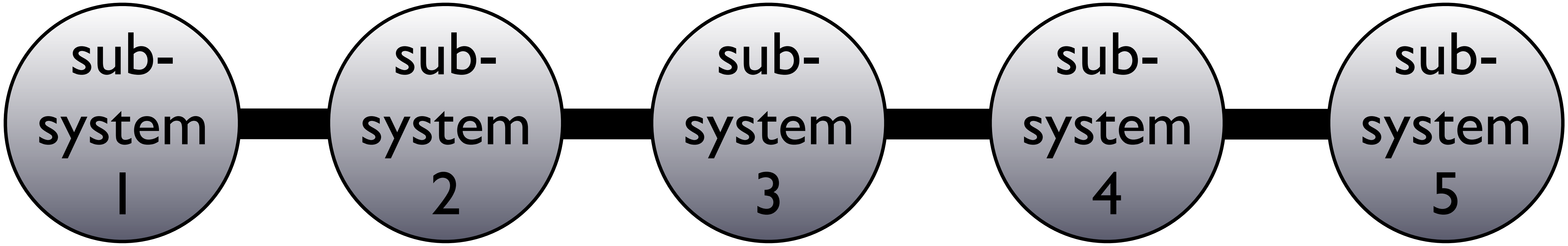
sub-system 3

sub-system 4

sub-system 5

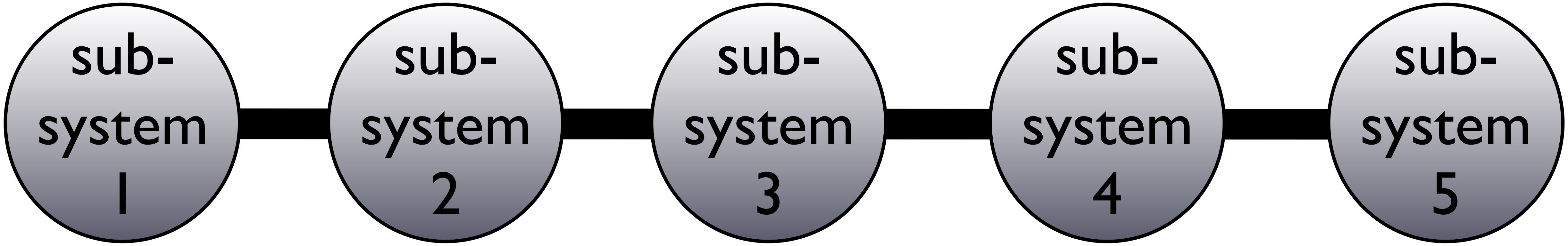
A coupled one-dimensional system:

2) Add a first neighbor coupling



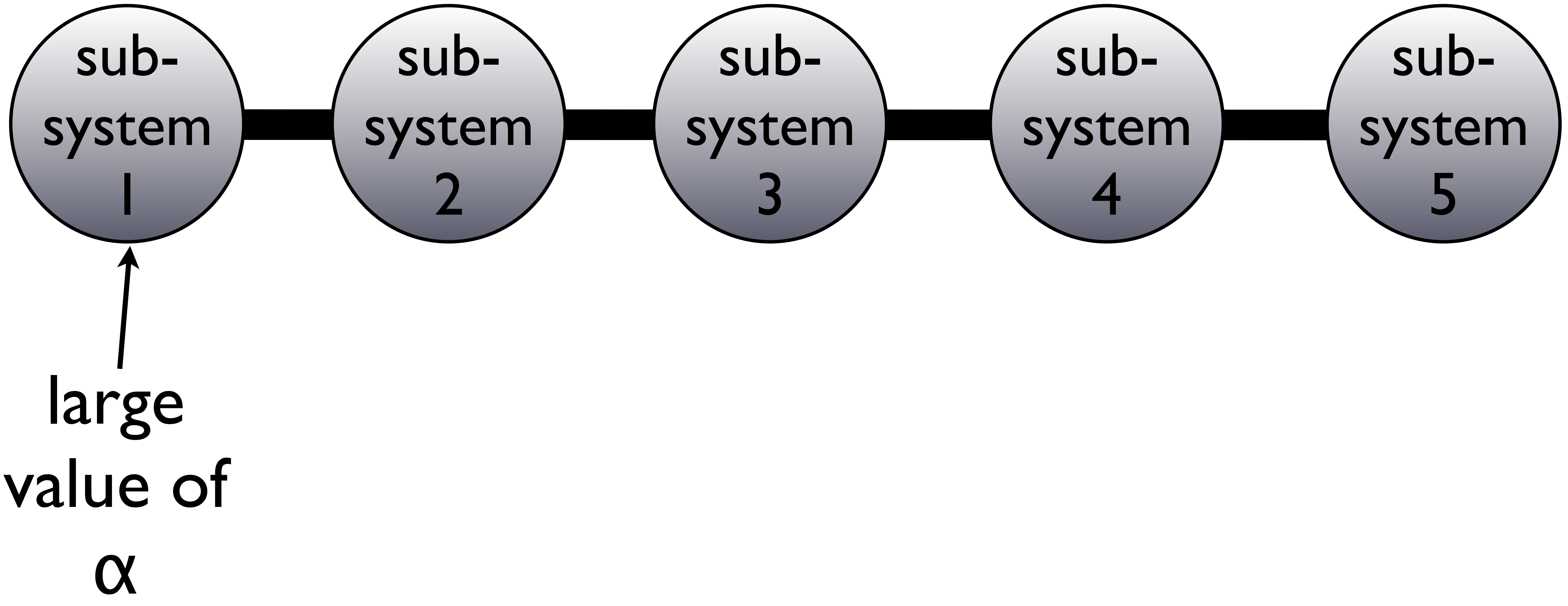
A coupled one-dimensional system:

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



A coupled one-dimensional system:

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



sub-system 1

sub-system 2

sub-system 3

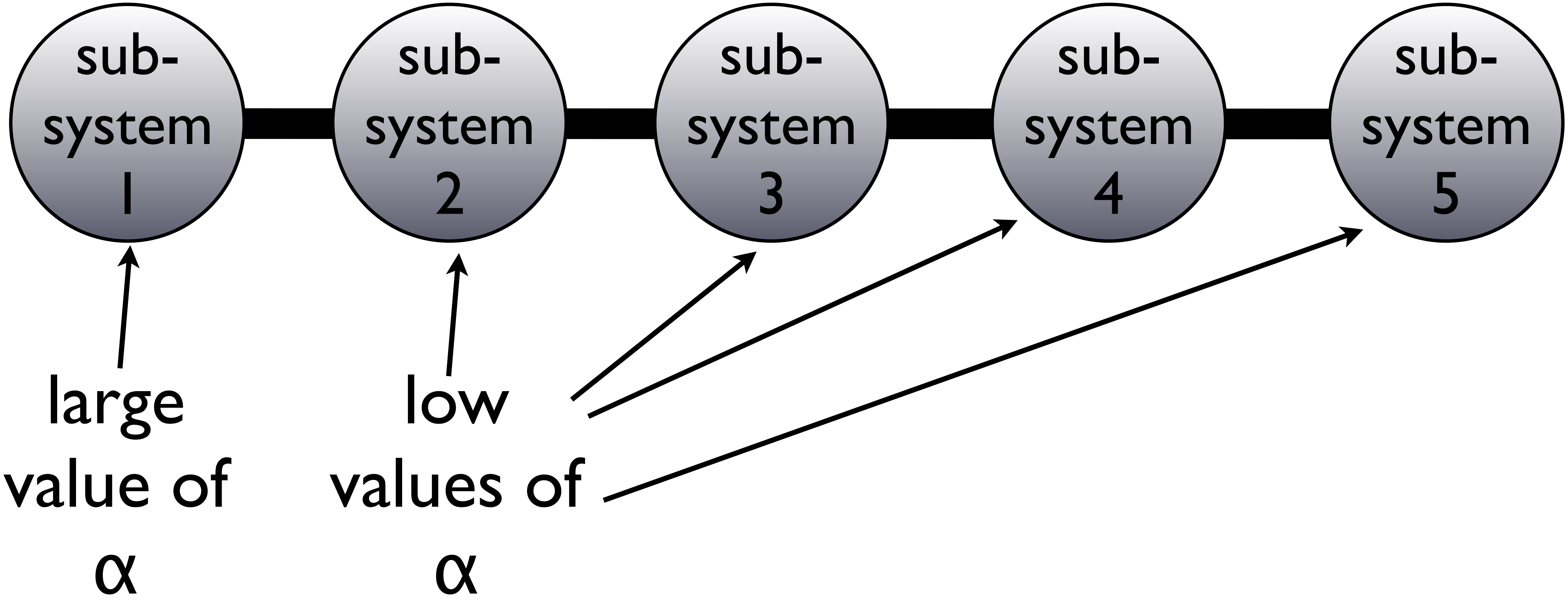
sub-system 4

sub-system 5

large value of α

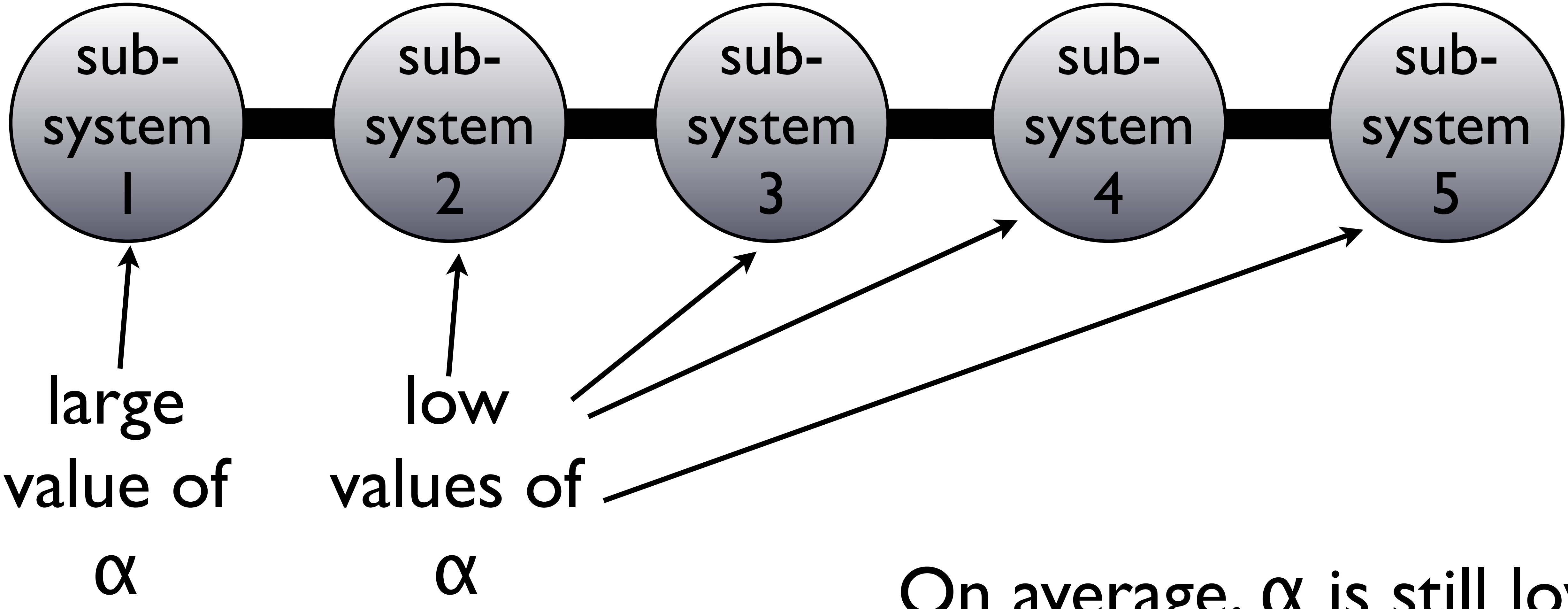
A coupled one-dimensional system:

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



A coupled one-dimensional system:

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



sub-system 1

sub-system 2

sub-system 3

sub-system 4

sub-system 5

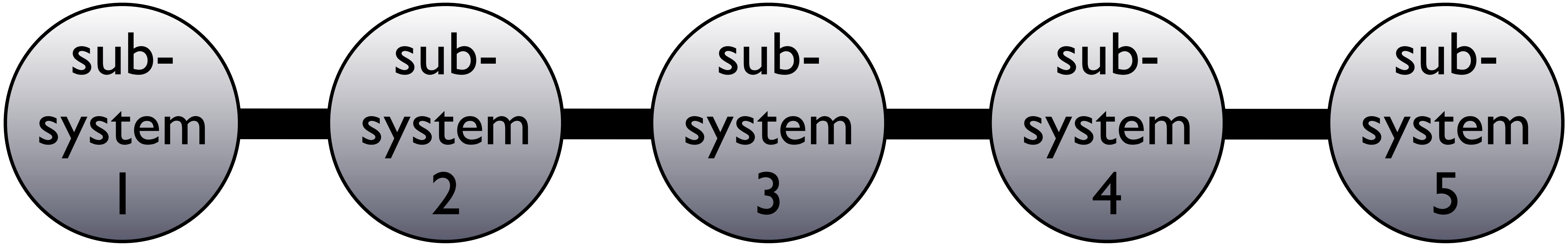
large value of α

low values of α

On average, α is still low !

A coupled one-dimensional system:

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



sub-system 1

sub-system 2

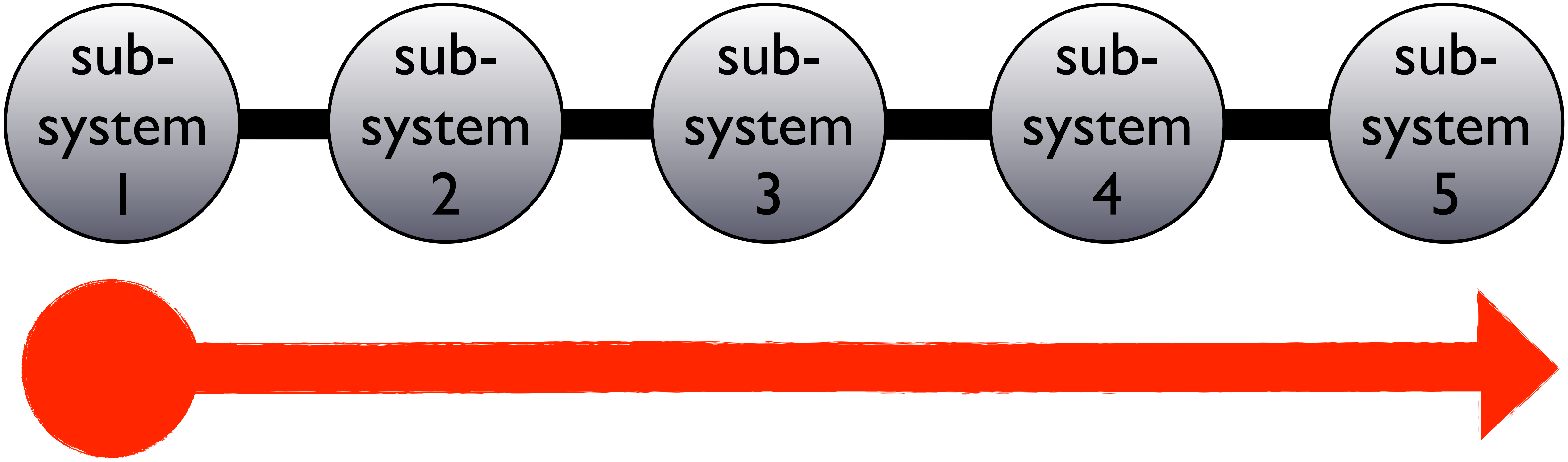
sub-system 3

sub-system 4

sub-system 5

A coupled one-dimensional system:

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



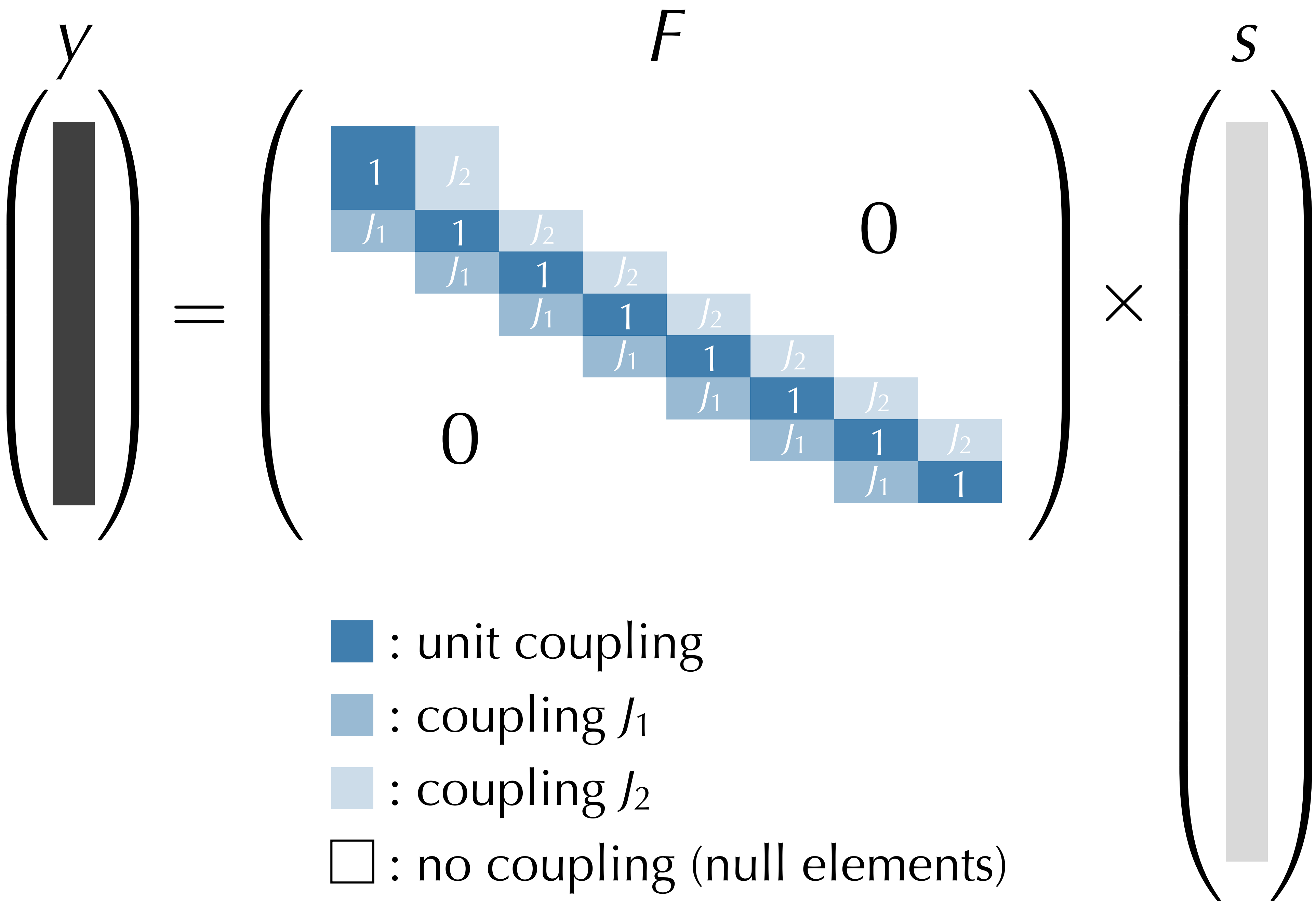
sub-system 1

sub-system 2

sub-system 3

sub-system 4

sub-system 5



$$L = 8$$

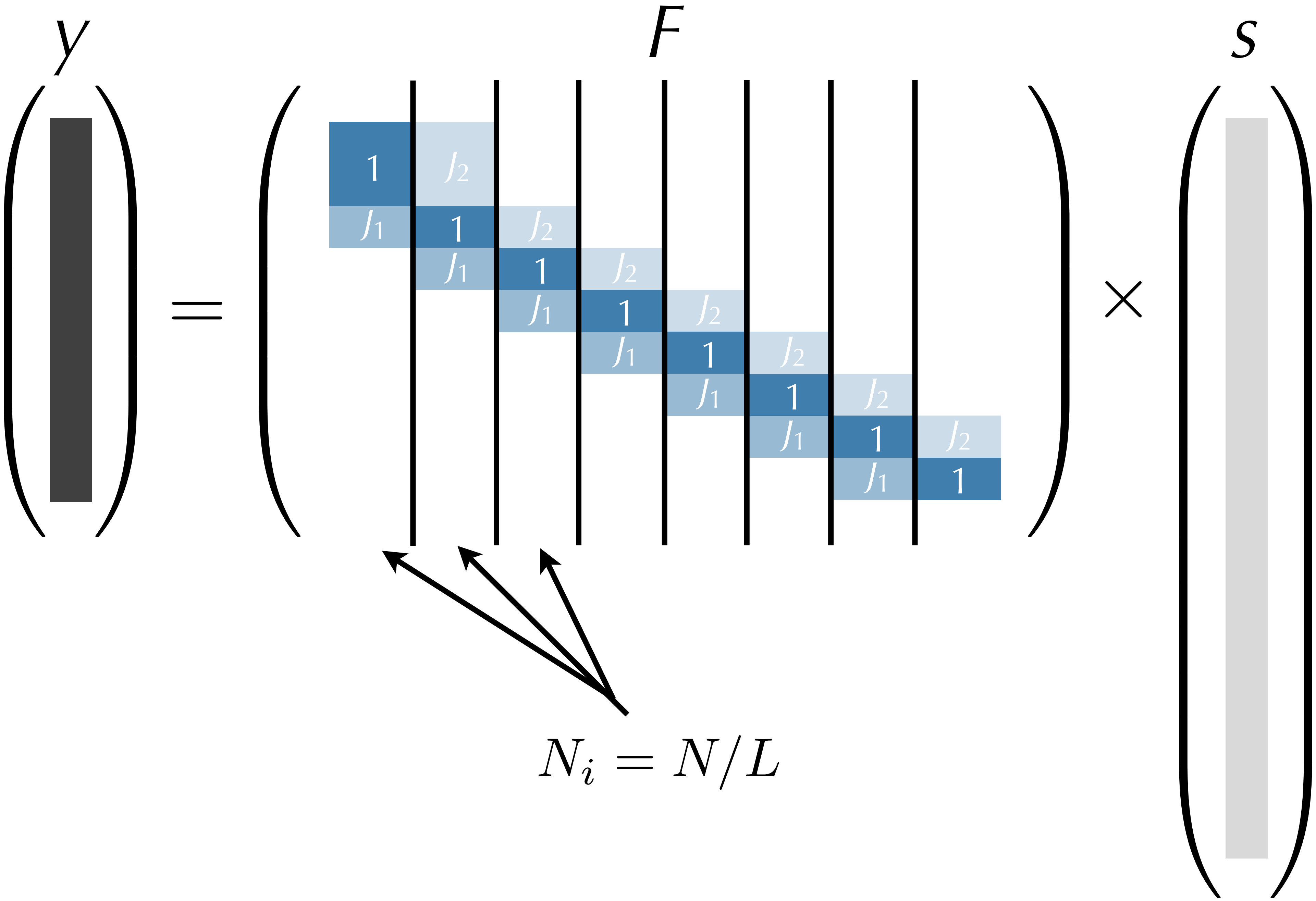
$$N_i = N/L$$

$$M_i = \alpha_i N/L$$

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

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

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



$$L = 8$$

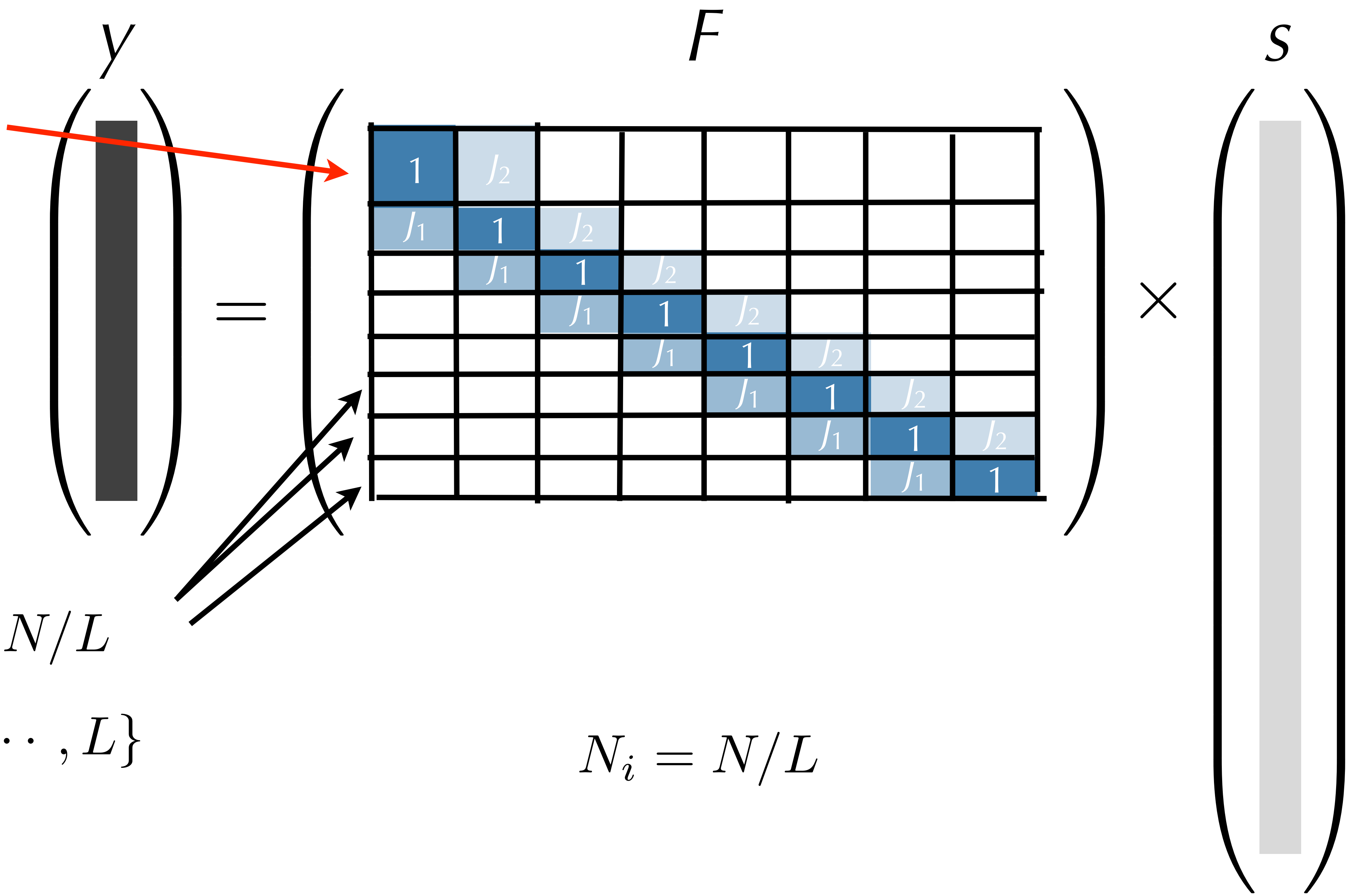
$$N_i = N/L$$

$$M_i = \alpha_i N/L$$

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

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

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



$M_1 = \alpha_1 N/L$

$M_i = \alpha' N/L$

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

$N_i = N/L$

$L = 8$

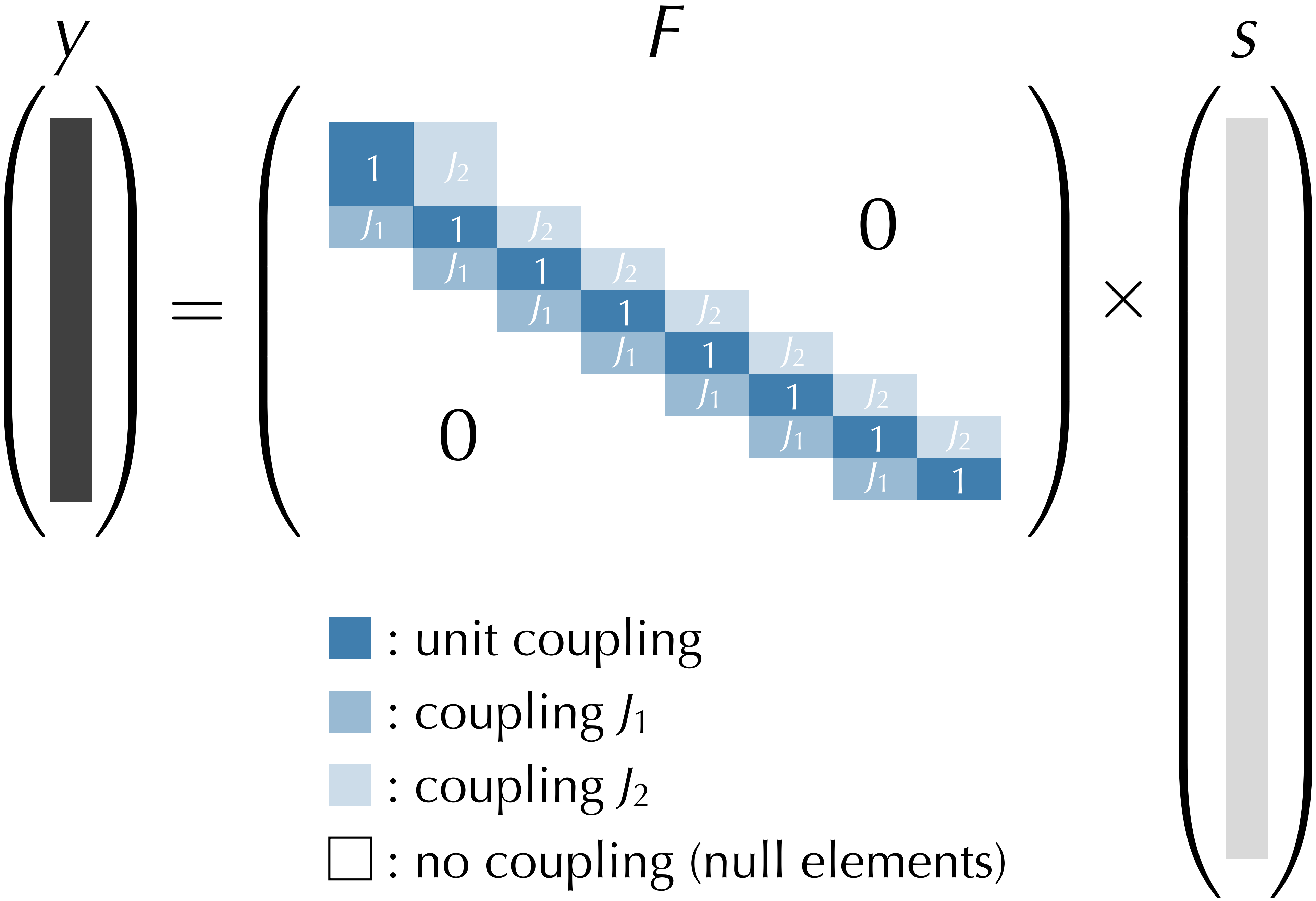
$N_i = N/L$

$M_i = \alpha_i N/L$

$\alpha_1 > \alpha_{BP}$

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

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



$$L = 8$$

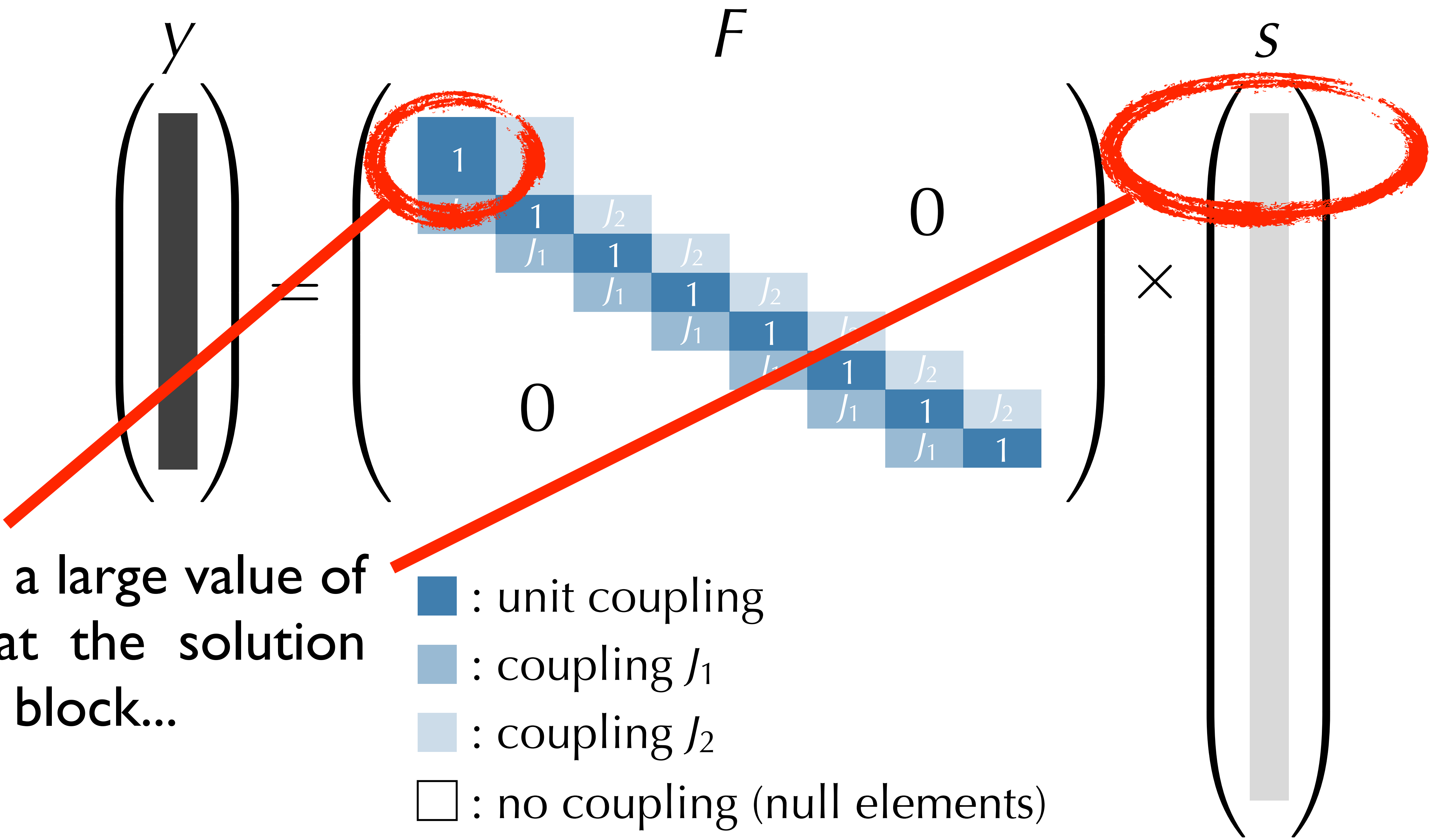
$$N_i = N/L$$

$$M_i = \alpha_i N/L$$

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

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

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



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

■ : unit coupling

▩ : coupling J_1

░ : coupling J_2

□ : no coupling (null elements)

$$L = 8$$

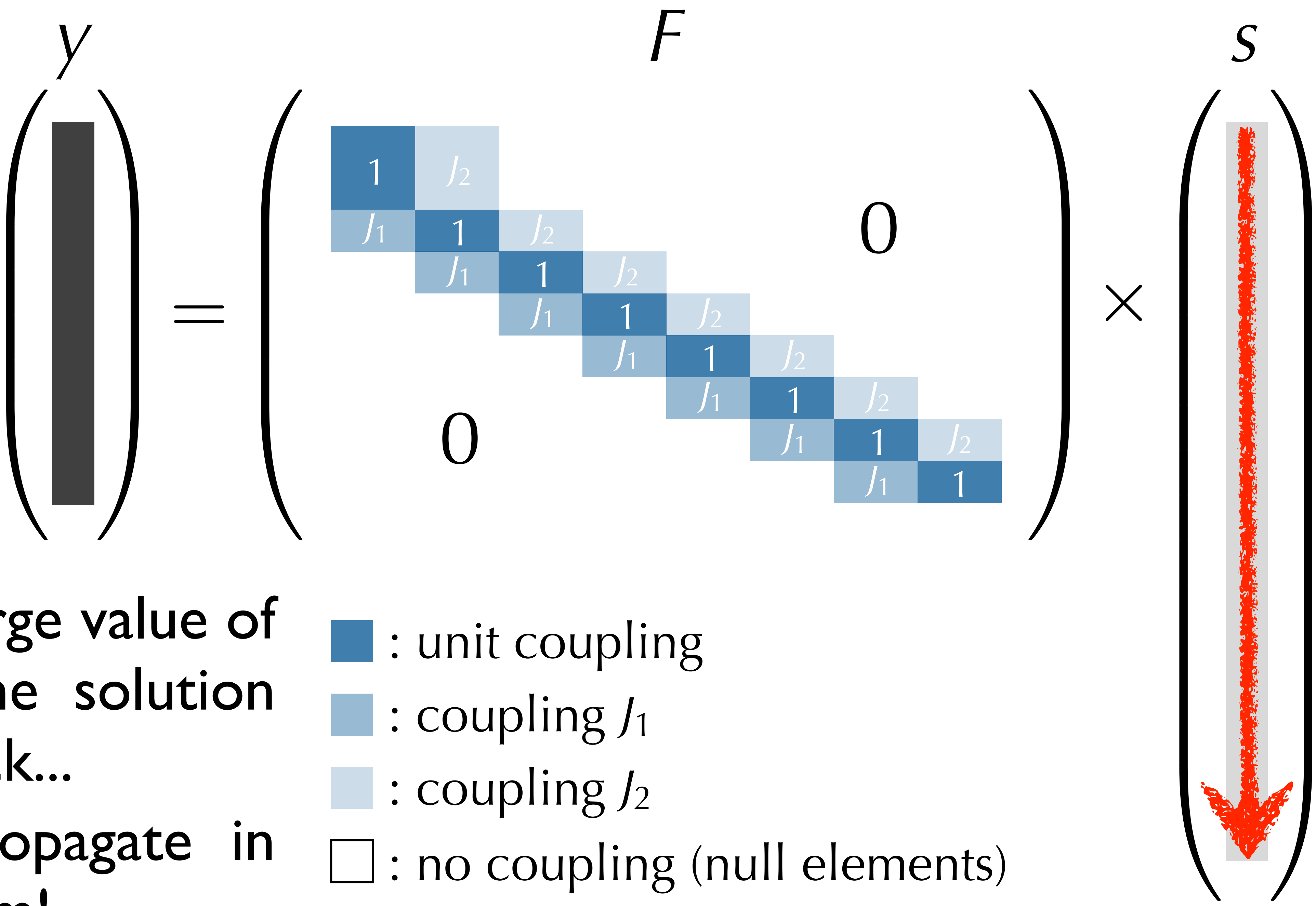
$$N_i = N/L$$

$$M_i = \alpha_i N/L$$

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

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

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



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

... and then propagate in the whole system!

- ■ : unit coupling
- ▨ : coupling J_1
- ░ : coupling J_2
- ☐ : no coupling (null elements)

$$L = 8$$

$$N_i = N/L$$

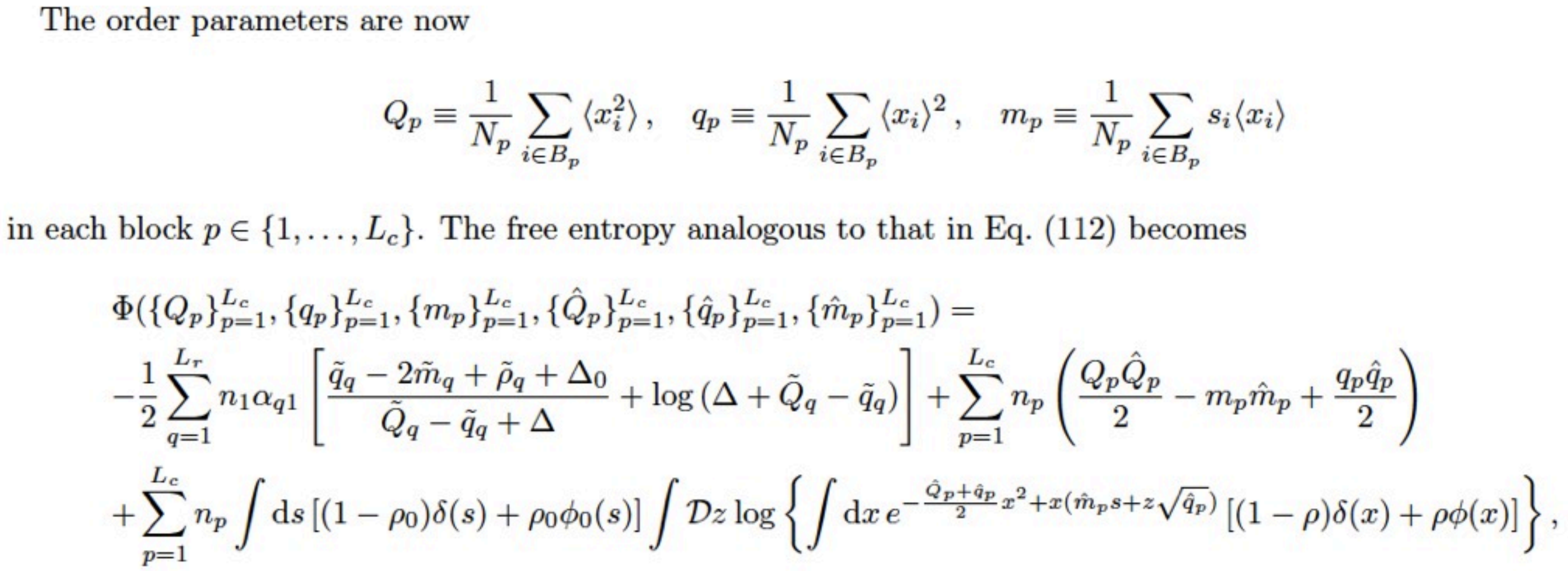
$$M_i = \alpha_i N/L$$

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

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

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

Replica solution for coupled seeded matrix



The order parameters are now

$$Q_p \equiv \frac{1}{N_p} \sum_{i \in B_p} \langle x_i^2 \rangle \,, \quad q_p \equiv \frac{1}{N_p} \sum_{i \in B_p} \langle x_i \rangle^2 \,, \quad m_p \equiv \frac{1}{N_p} \sum_{i \in B_p} s_i \langle x_i \rangle$$

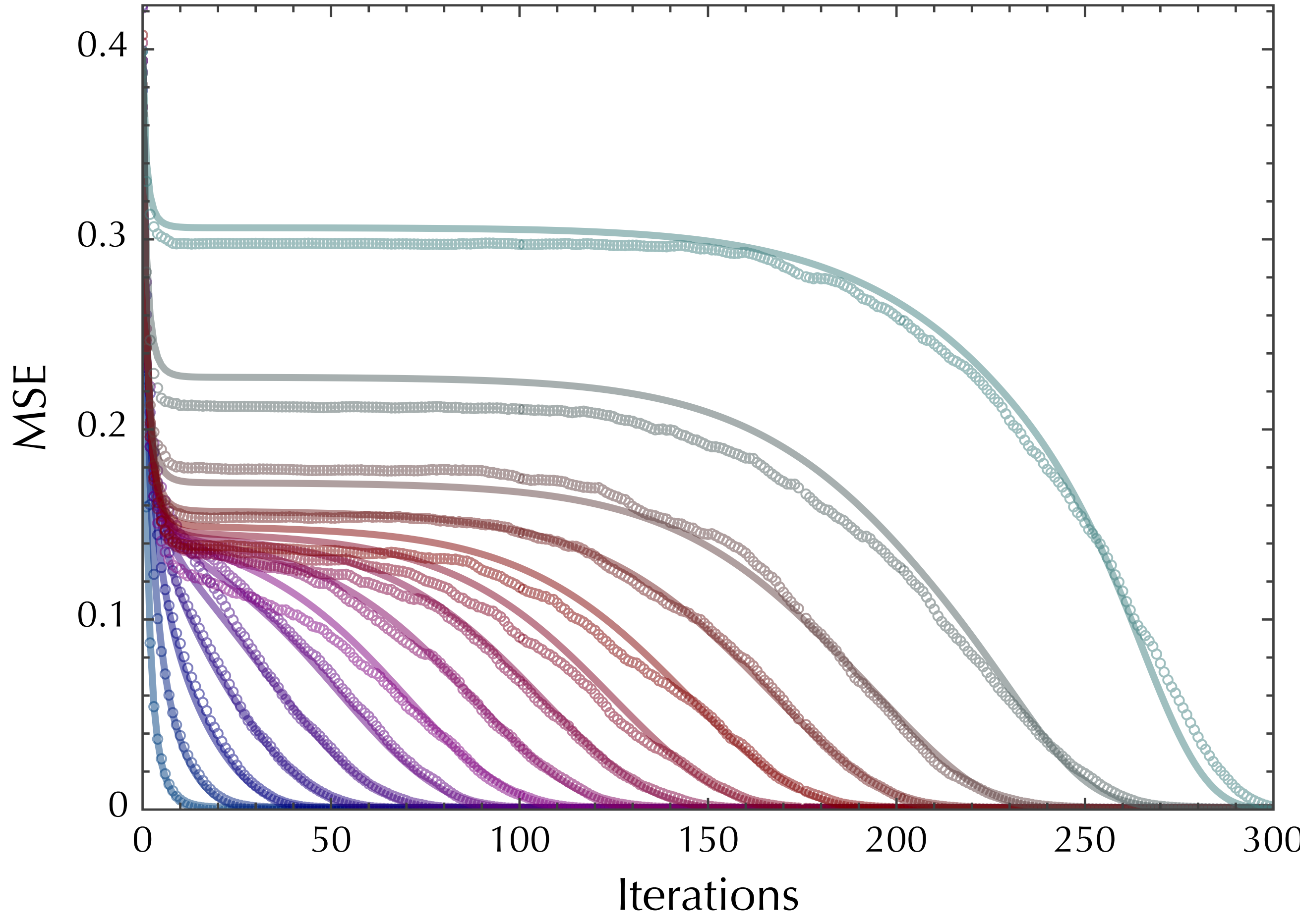
in each block $p \in \{1, \dots, L_c\}$. The free entropy analogous to that in Eq. (112) becomes

$$\begin{aligned}
&\Phi(\{Q_p\}_{p=1}^{L_c}, \{q_p\}_{p=1}^{L_c}, \{m_p\}_{p=1}^{L_c}, \{\hat{Q}_p\}_{p=1}^{L_c}, \{\hat{q}_p\}_{p=1}^{L_c}, \{\hat{m}_p\}_{p=1}^{L_c}) = \\
&-\frac{1}{2} \sum_{q=1}^{L_r} n_1 \alpha_{q1} \left[\frac{\tilde{q}_q - 2\tilde{m}_q + \tilde{\rho}_q + \Delta_0}{\tilde{Q}_q - \tilde{q}_q + \Delta} + \log\left(\Delta + \tilde{Q}_q - \tilde{q}_q\right) \right] + \sum_{p=1}^{L_c} n_p \left(\frac{Q_p \hat{Q}_p}{2} - m_p \hat{m}_p + \frac{q_p \hat{q}_p}{2} \right) \\
&+ \sum_{p=1}^{L_c} n_p \int \mathrm{d}s \, [(1-\rho_0)\delta(s) + \rho_0 \phi_0(s)] \int \mathcal{D}z \log \left\{ \int \mathrm{d}x \, e^{-\frac{\hat{Q}_p + \hat{q}_p}{2} x^2 + x(\hat{m}_p s + z\sqrt{\hat{q}_p})} [(1-\rho)\delta(x) + \rho\phi(x)] \right\} ,
\end{aligned}$$

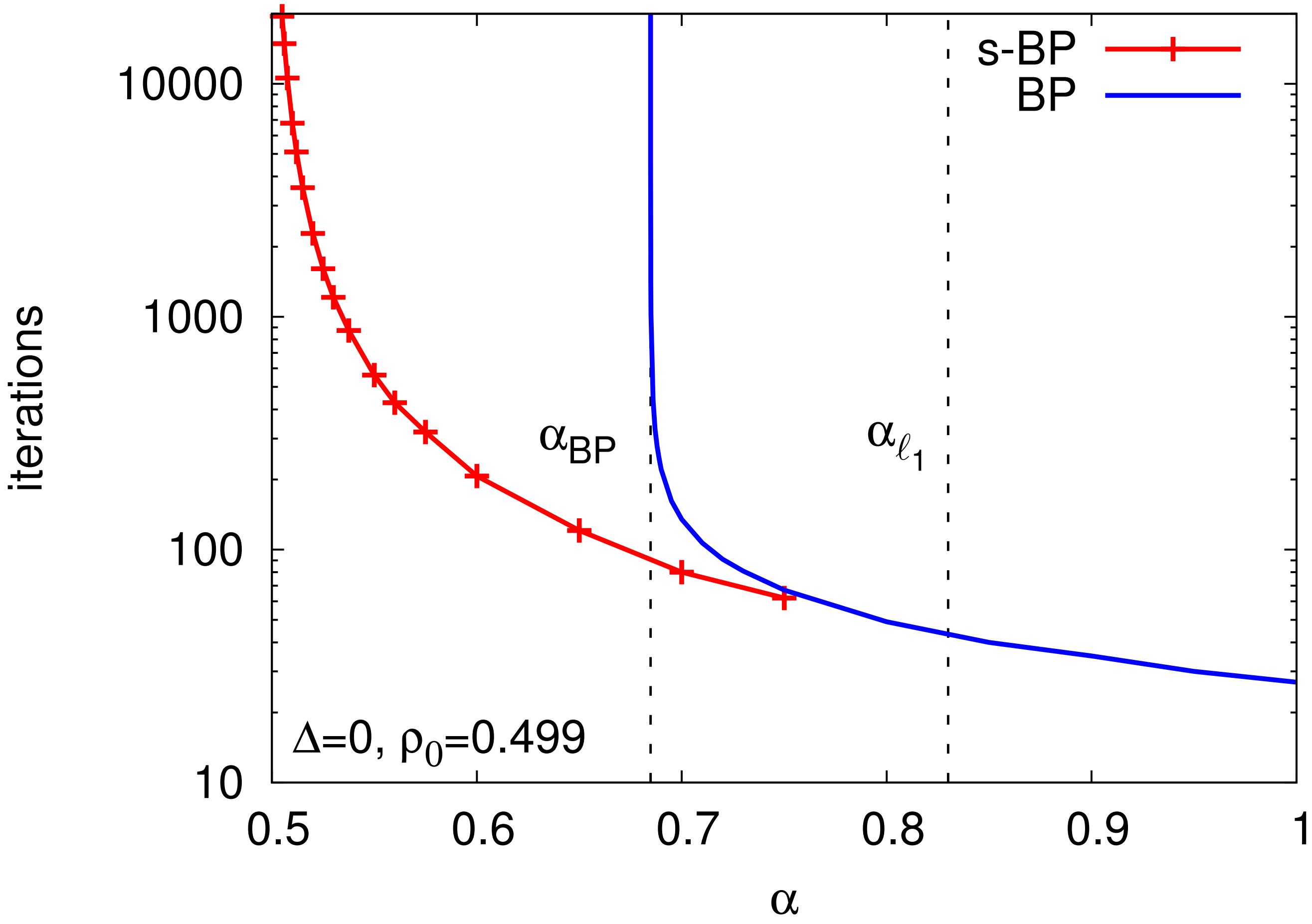
(after a bit of work...)

Comparing the algorithm and replica theory

BP analyzed by density evolution versus
an actual test with N=40000
(MSE in the different block versus time)



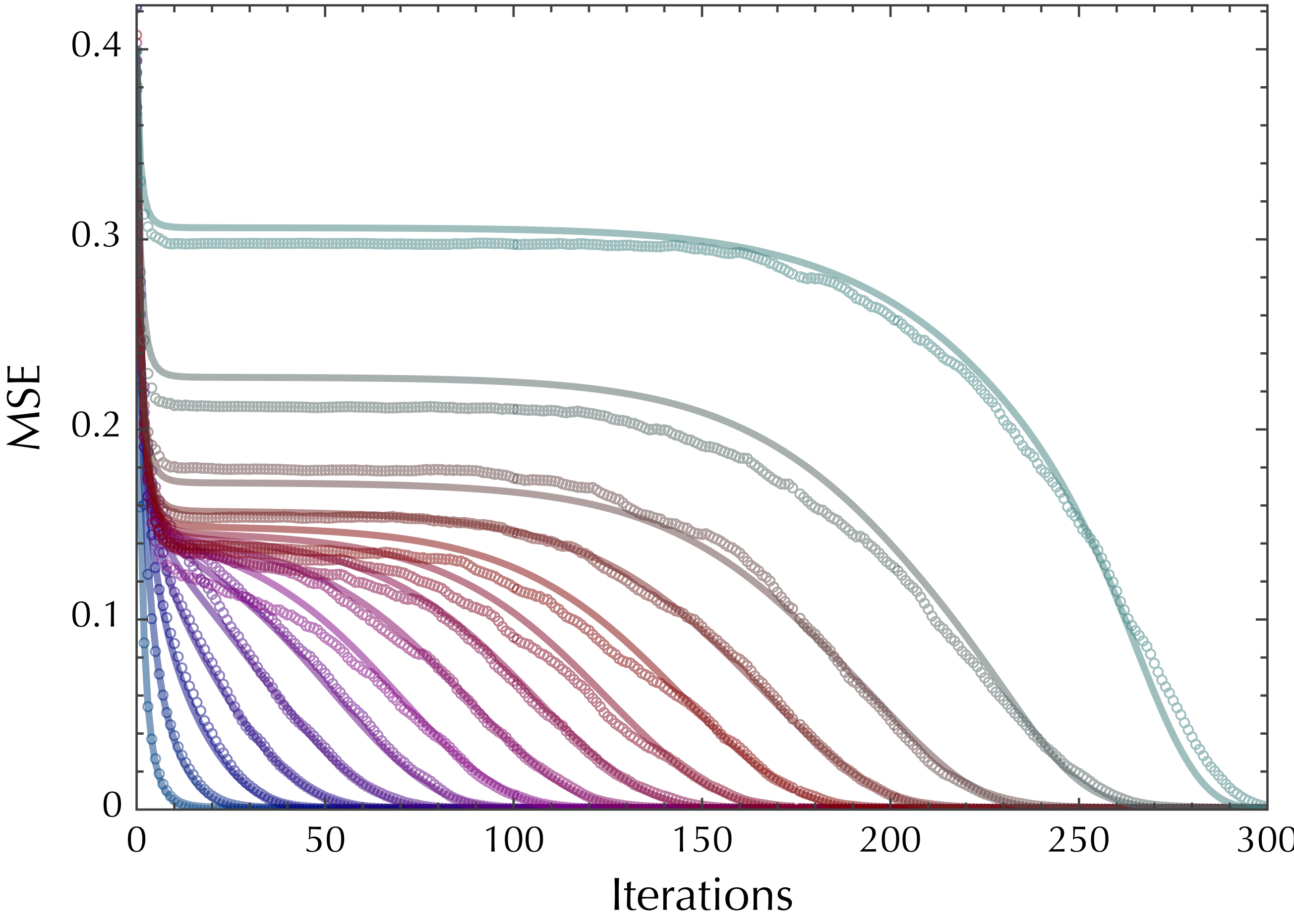
BP reconstruction time for
for seeded and non seeded matrices



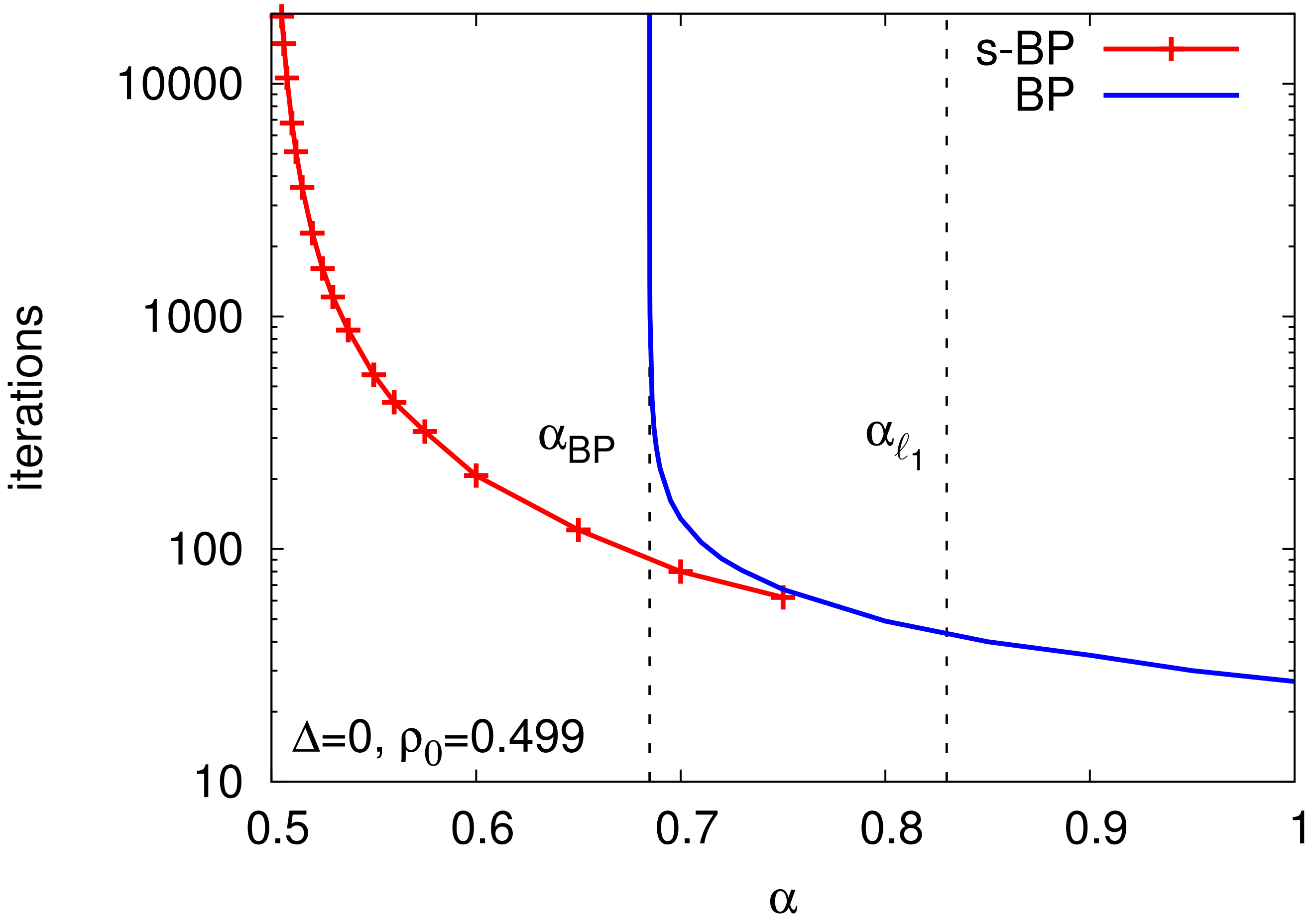
This strategy allows an Optimal reconstruction
(up to $\alpha=\rho$) in the limit of large signals

Comparing the algorithm and replica theory

BP analyzed by density evolution versus
an actual test with N=40000
(MSE in the different block versus time)



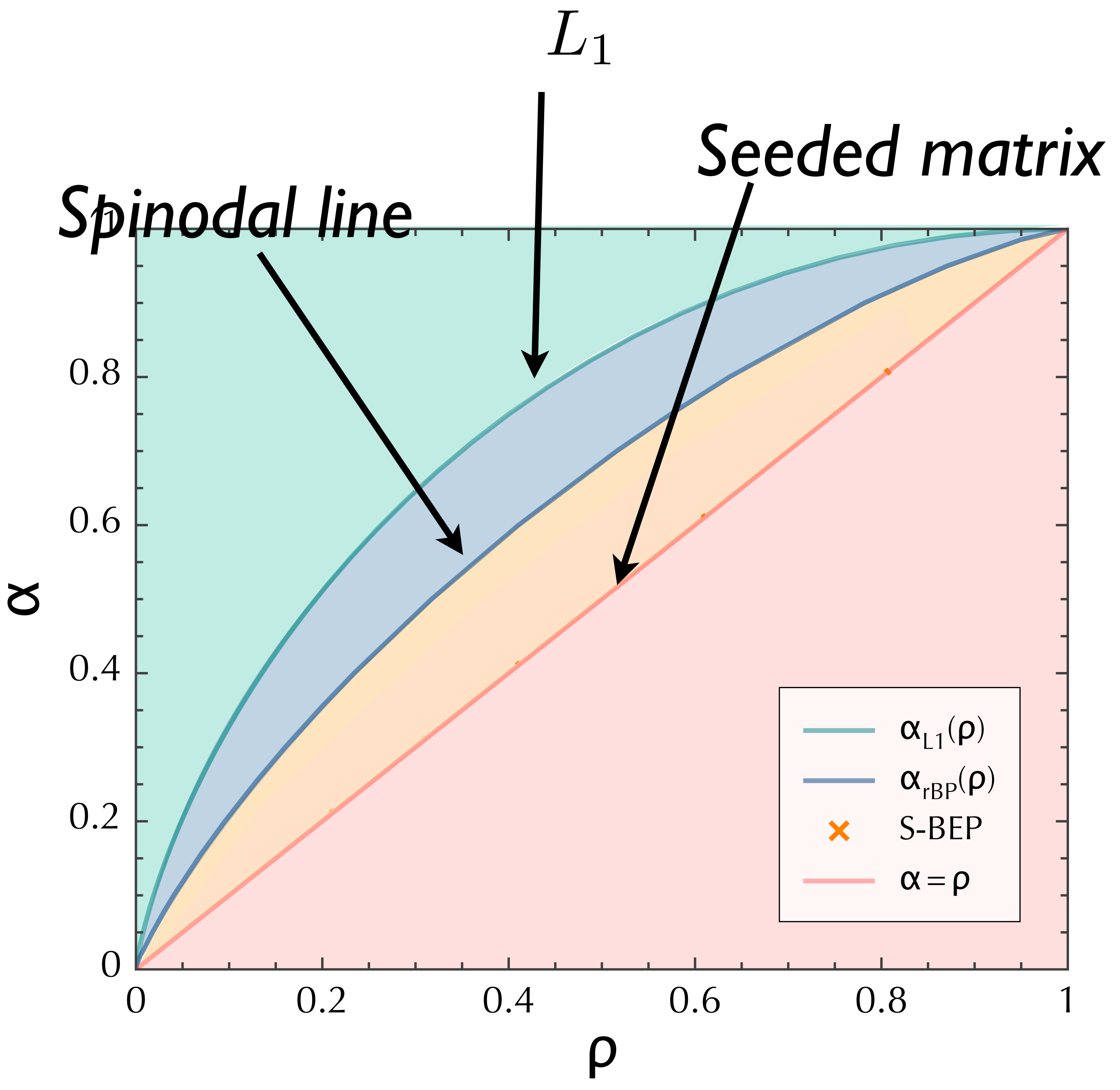
BP reconstruction time for
for seeded and non seeded matrices



This strategy allows an Optimal reconstruction
(up to $\alpha=\rho$) in the limit of large signals

Generic proof for optimal reconstruction
(when the prior matches the signal):
D. Donoho, A. Javanmard, & A. Montanari, '11

Best measurement rates reached!



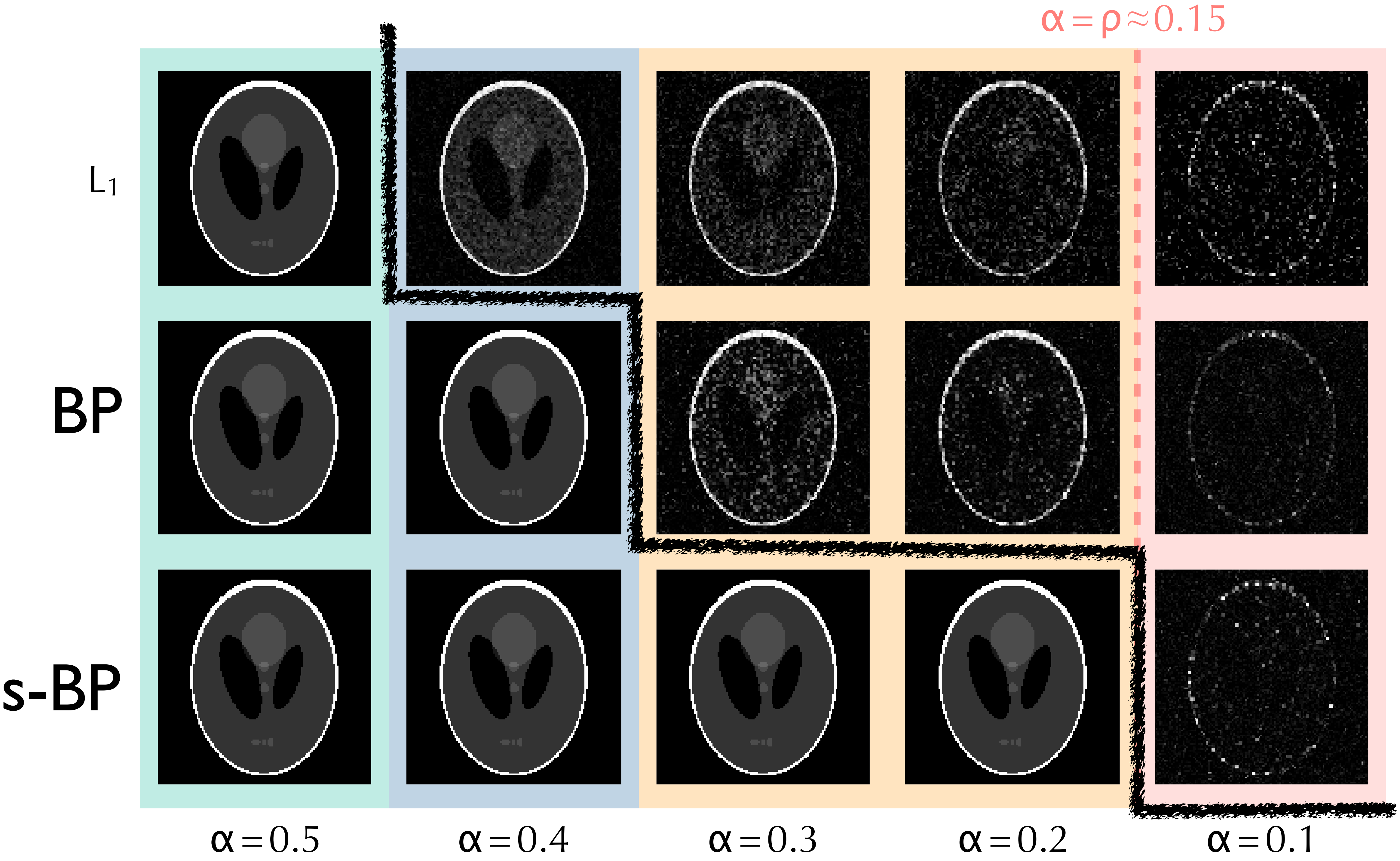
L_1

Seeded matrix

Spinodal line

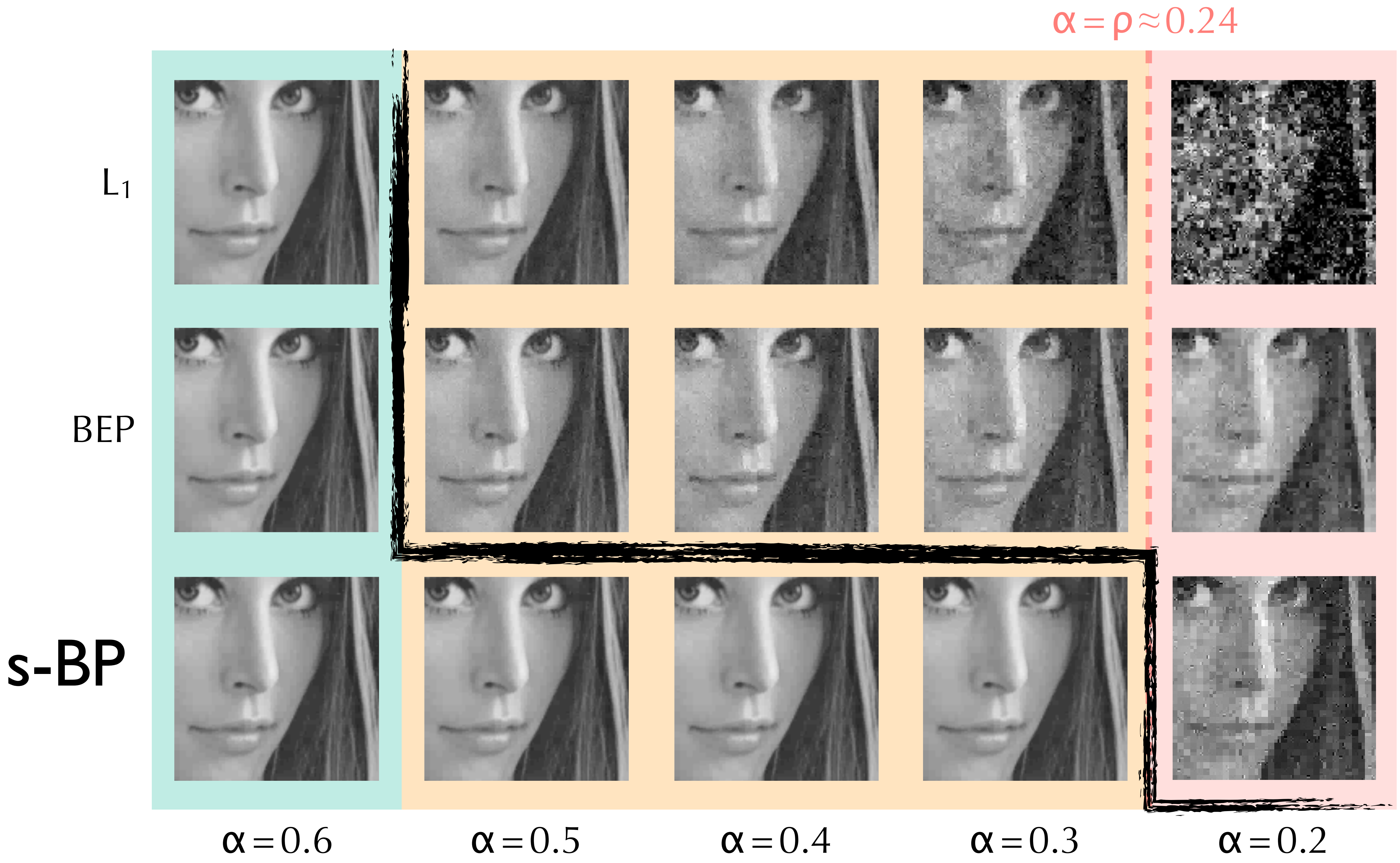
A combination of Statistical physics technics (Bethe-Peierls, Replica) and concepts (dynamics, nucleation and growth) has allowed to solve a major problem in signal processing theory

An example



Shepp-Logan phantom, in the Haar-wavelet representation

A more interesting example



The Lena picture in the Haar-wavelet representation

Conclusions...

- A probabilistic approach to reconstruction
- Analysis of best possible reconstruction for different class of signals
- The Belief Propagation algorithm
- Optimality achieving seeded measurements matrices

... and perspectives:

- More information in the prior (Correlated measurement, wavelets, etc...)
- Other matrices with asymptotic measurements?
- Non-random matrix (e.g. Radon operator in Tomography, Fourier, etc..)
- Additive and multiplicative noise, Quasi-sparsity, etc... ?
- Calibration, and matrix/dictionary learning?
- Applications ?

http://leshouches2013.krzakala.org

SPECIAL ANNOUNCEMENTS

http://leshouches2013.krzakala.org

SPECIAL ANNOUNCEMENTS

2 Post-doc openings on these topics for 2013

If you work in Statistical physics, Information science, Signal processing, etc...

ASPICS

Project

Applying Statistical Physics to Inference in Compressed Sensing

http://krzakala.org

http://leshouches2013.krzakala.org

SPECIAL ANNOUNCEMENTS

2 Post-doc openings on these topics for 2013

If you work in Statistical physics, Information science, Signal processing, etc...

ASPICS

Project

Applying Statistical Physics to Inference in Compressed Sensing

http://krzakala.org

COMING SOON: An interdisciplinary school on these topics:

Les Houches, October 2013, Organizers F. Krzakala & L. Zdeborová

http://leshouches2013.krzakala.org