A Probabilistic Approach to Compressed Sensing

Robust Algorithms

Eric W. Tramel

28 August 2014

Contributors & SPARCS Team

Andre MANOEL
Univ. Sao Paulo

Florent KRZAKALA
ENS, Univ. P. & M. Curie

Lenka ZDEBOROVÁ
CNRS & CEA-Saclay

Angéliqué DREMEAU
ENS

Eric W. TRAMEL
ENS

Francesco CALTAGIRONE
CNRS & CEA-Saclay

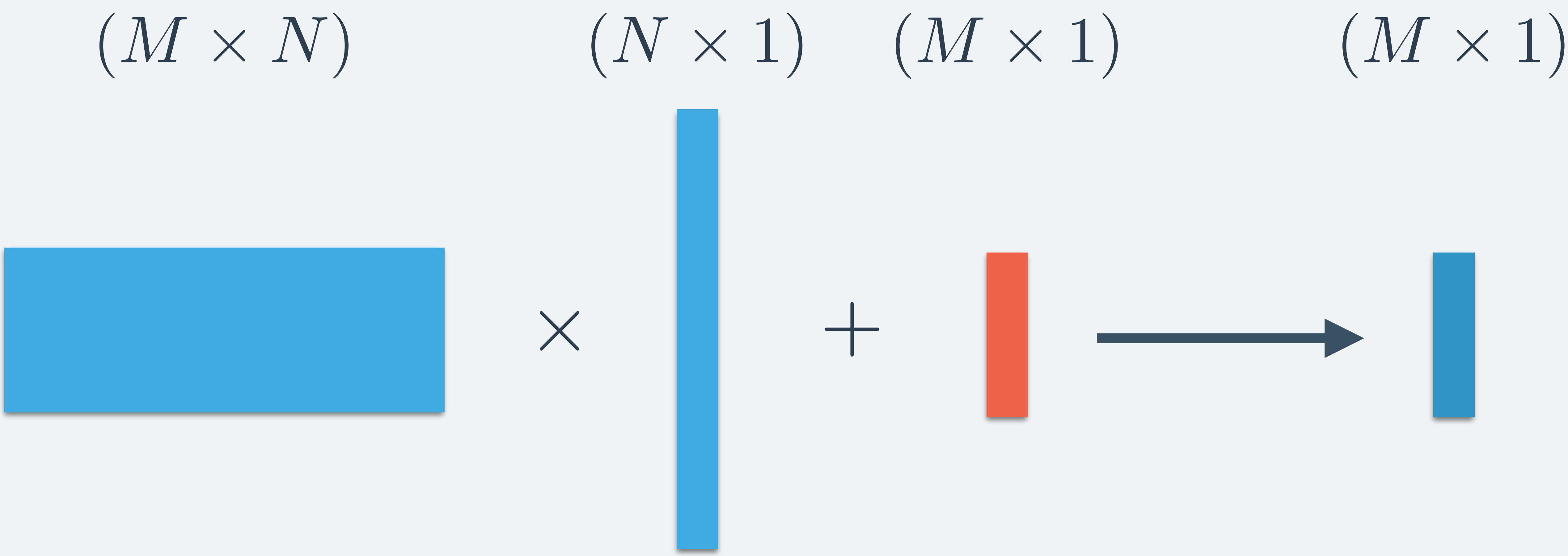
Alaa SAADE
ENS

Jean BARBIER
ENS

Christophe SCHÜLKE
CNRS & CEA-Saclay

Compressed Sensing

General CS Problem: $\mathbf{y} = F\mathbf{x} + \mathbf{w}$



Projection Matrix
Underdetermined

Signal
Sparse

Noise
Additive

Measurements
Observed Data

Probabilistic CS

Question: How do we obtain **x** from **y** and **F**?

Deterministic

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad ||\mathbf{x}||_0 \quad \text{s.t.} \quad ||\mathbf{y} - F\mathbf{x}||_2^2 \leq \epsilon \qquad \text{(Greedy)}$$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad ||\mathbf{y} - F\mathbf{x}||_2^2 + \lambda ||\mathbf{x}||_1 \qquad \text{(LASSO)}$$

Probabilistic

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \quad P(\mathbf{x}|\mathbf{y}, F) \qquad \text{(MAP)}$$

$$\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}] = \int d\mathbf{x} \;\; \mathbf{x} \; P(\mathbf{x}|\mathbf{y}, F) \qquad \text{(MMSE)}$$

Probabilistic CS

Bayes Rule

$$P(\mathbf{x}|\mathbf{y}, F) = \frac{1}{Z} P(\mathbf{y}|\mathbf{x}, F) P_0(\mathbf{x})$$

AWGN Likelihood

$$P(\mathbf{y}|\mathbf{x}, F) = \prod_{\mu} \mathcal{N}\left(y_\mu - \sum_i F_{\mu i} x_i, \Delta\right)$$

Probabilistic CS

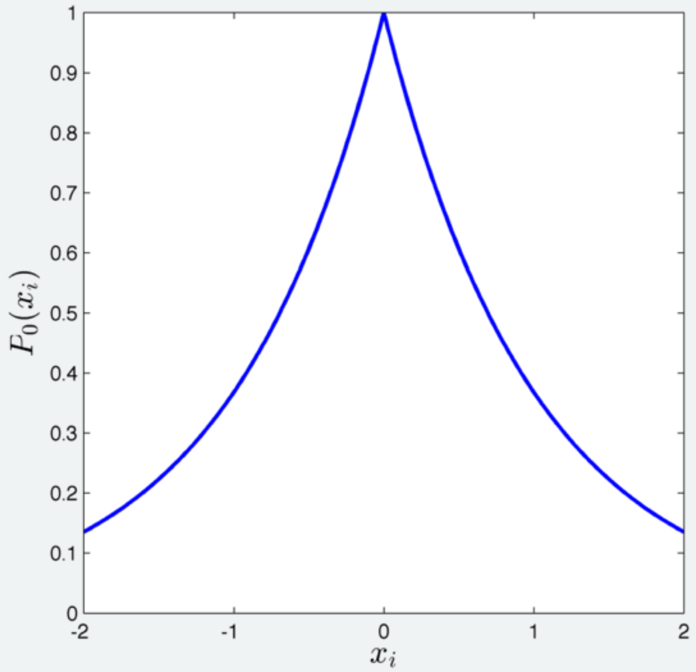
Factorized Prior

$$P_0(\mathbf{x}) = \prod_i P_0(x_i)$$

For K-Sparse Signals…

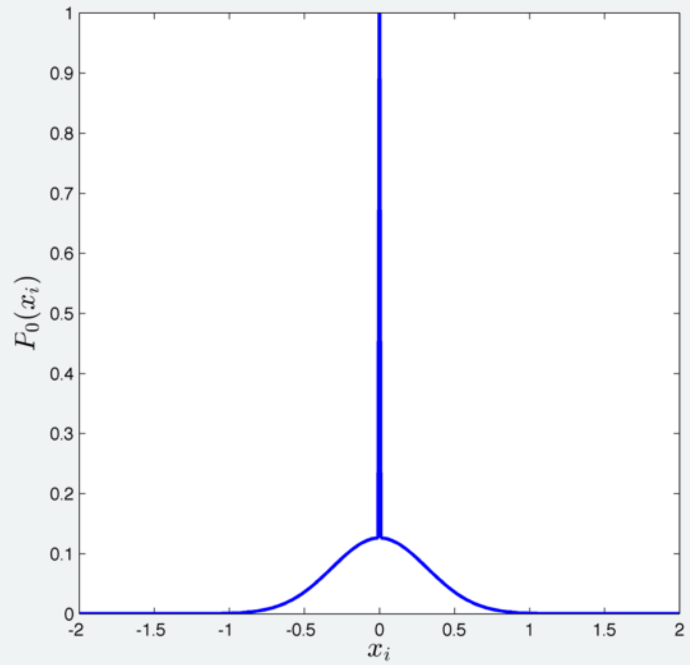
L1/Laplace

$$P_0(x_i) \propto \exp\{-|x_i|\}$$



"Sparse" Bernoulli-ϕ

$$P_0(x_i) = (1-\rho)\delta(x_i) + \rho\phi(x_i)$$



An Unwieldy Posterior

Full Posterior

$$P(\mathbf{x}|\mathbf{y}, F) = \frac{1}{Z} \prod_i P_0(x_i) \prod_\mu \frac{1}{\sqrt{2\pi\Delta}} \exp\left\{ -\frac{1}{2\Delta} \left(y_\mu - \sum_i F_{\mu i} x_i \right)^2 \right\}$$

Oh Buddy, That Partition…

$$Z = \int \mathrm{d}x_1 \int \mathrm{d}x_2 \ldots \int \mathrm{d}x_N \prod_i P_0(x_i) \prod_\mu \frac{1}{\sqrt{2\pi\Delta}} \exp\left\{ -\frac{1}{2\Delta} \left(y_\mu - \sum_i F_{\mu i} x_i \right)^2 \right\}$$

An Unwieldy Posterior

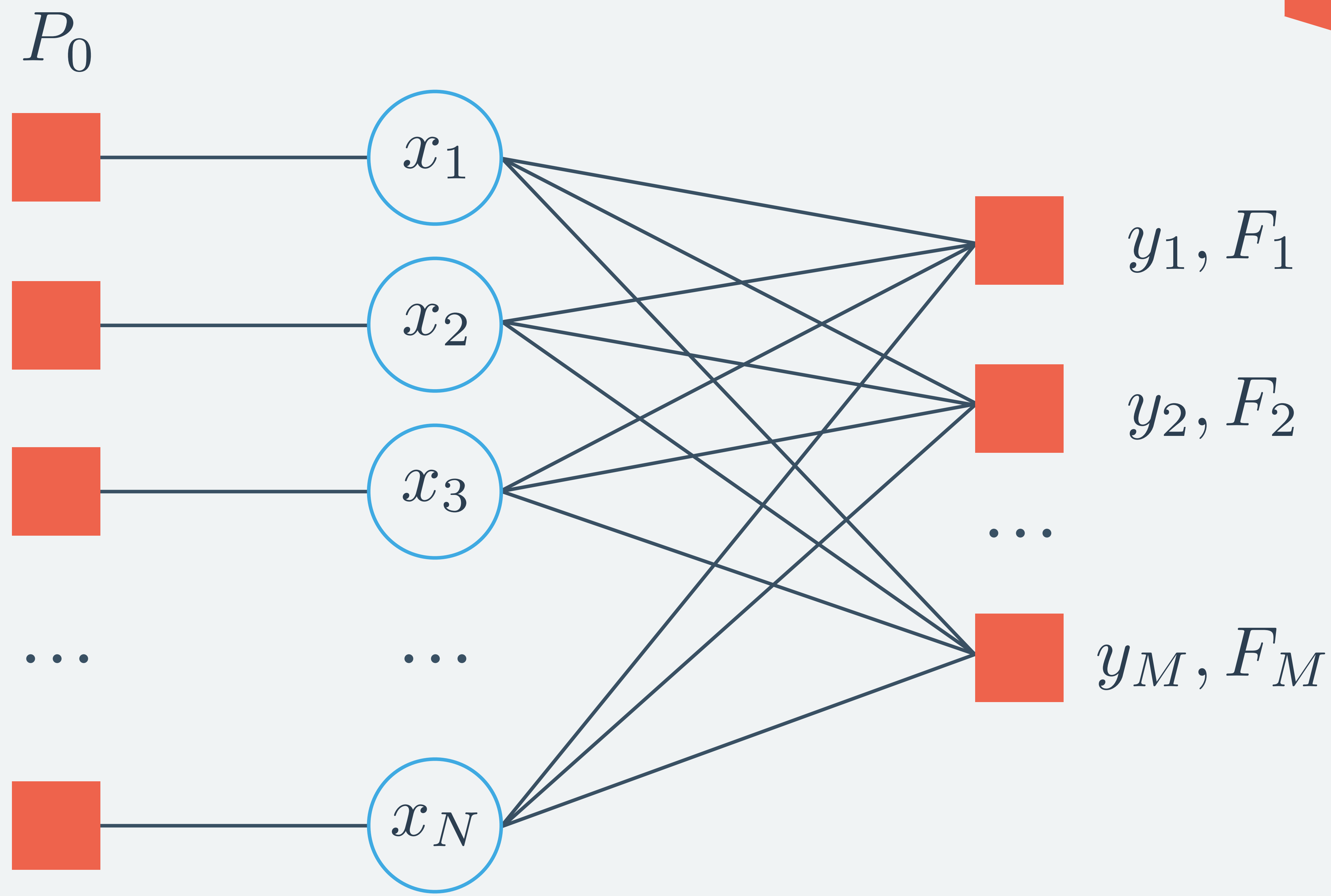
Much Nicer: A Factorized Posterior

$$P(\mathbf{x}|\mathbf{y}, F) \propto \prod_i Q(x_i|\mathbf{y}, F)$$

How do we find Q?

Sum-Product Belief Propagation on the Factor Graph!

Graphical Model for Factorization

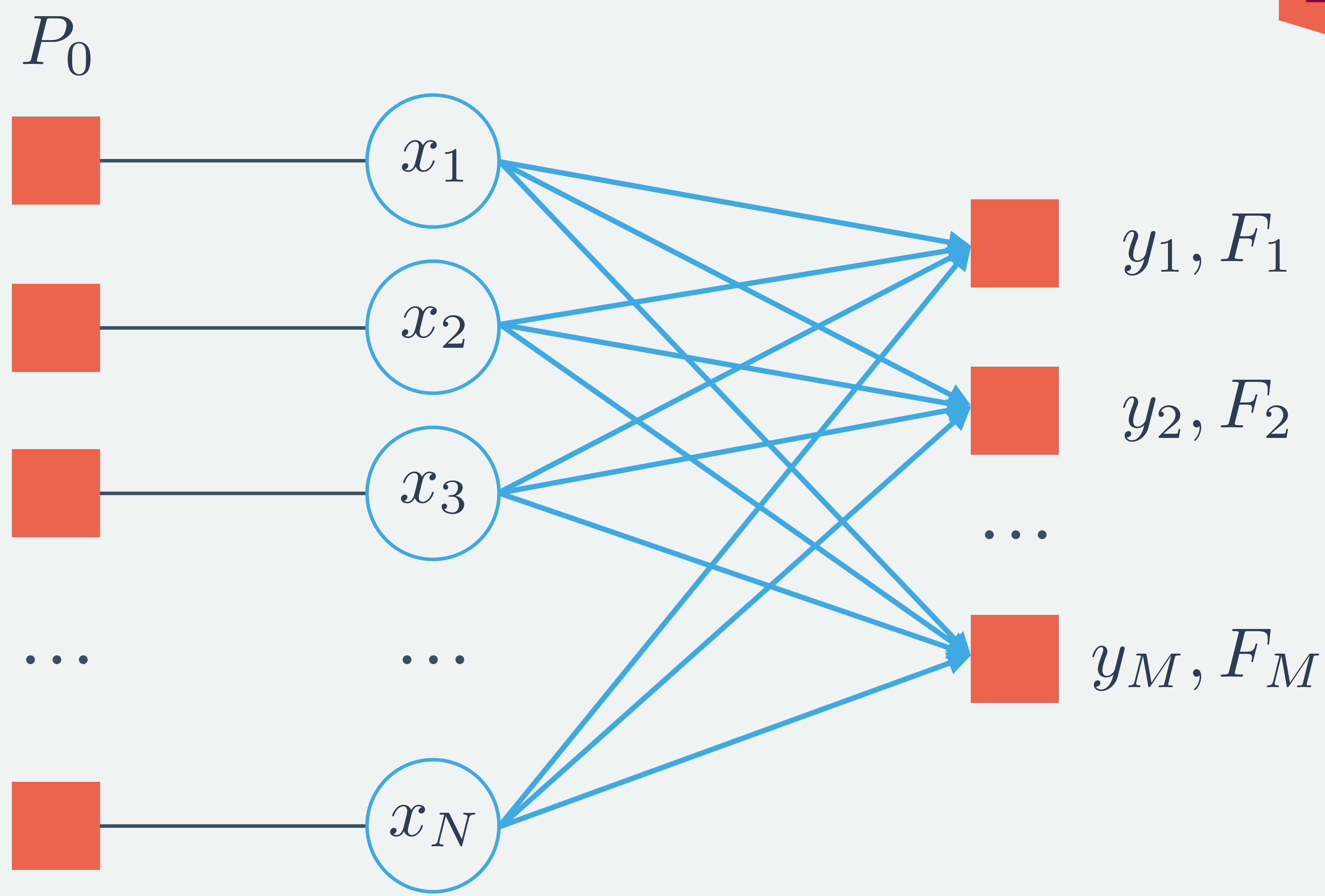


Factors
Prior

Variables
Coefficients

Factors
Measurements

Graphical Model for Factorization

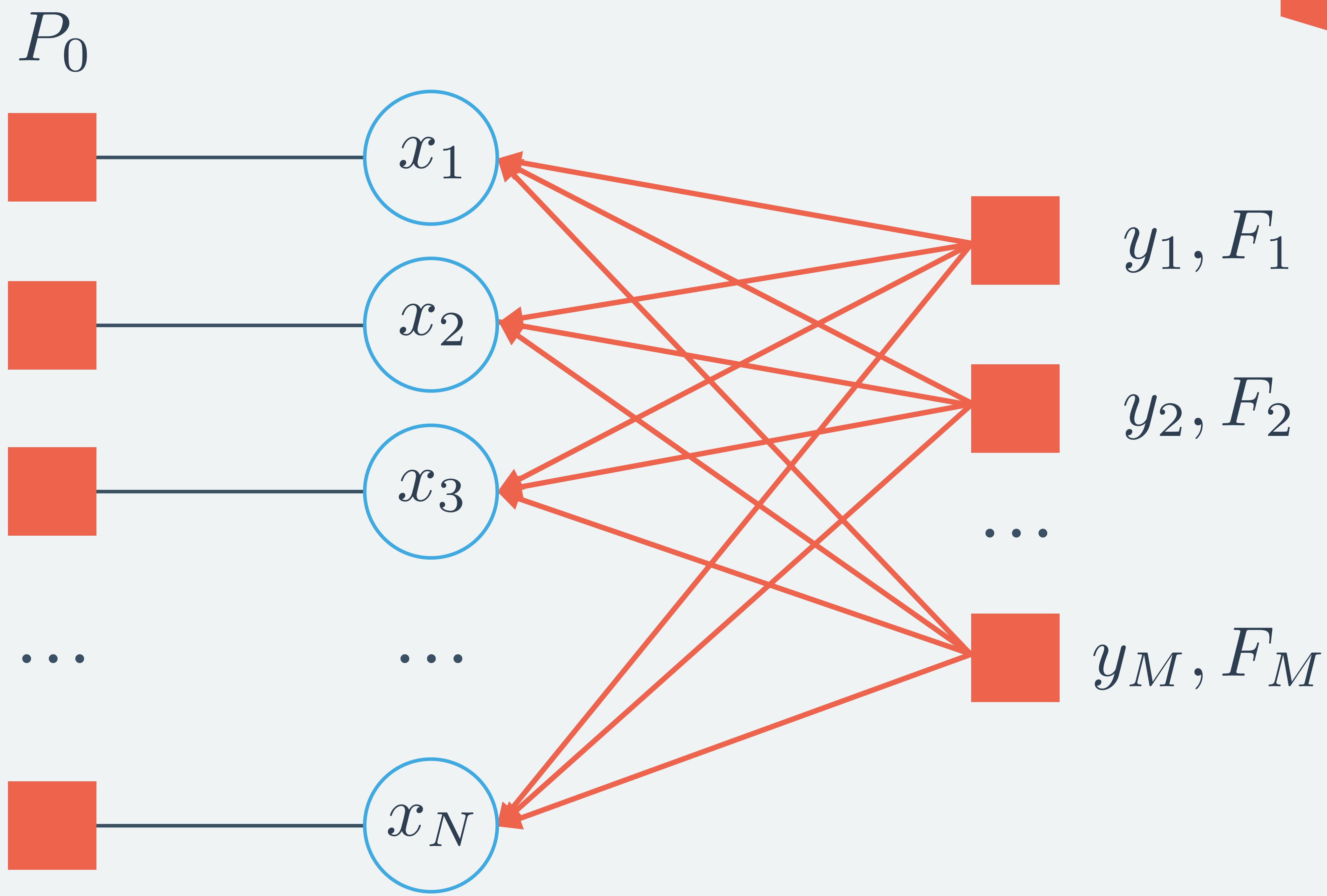


Factors
Prior

Variables
Coefficients

Factors
Measurements

Graphical Model for Factorization



Factors
Prior

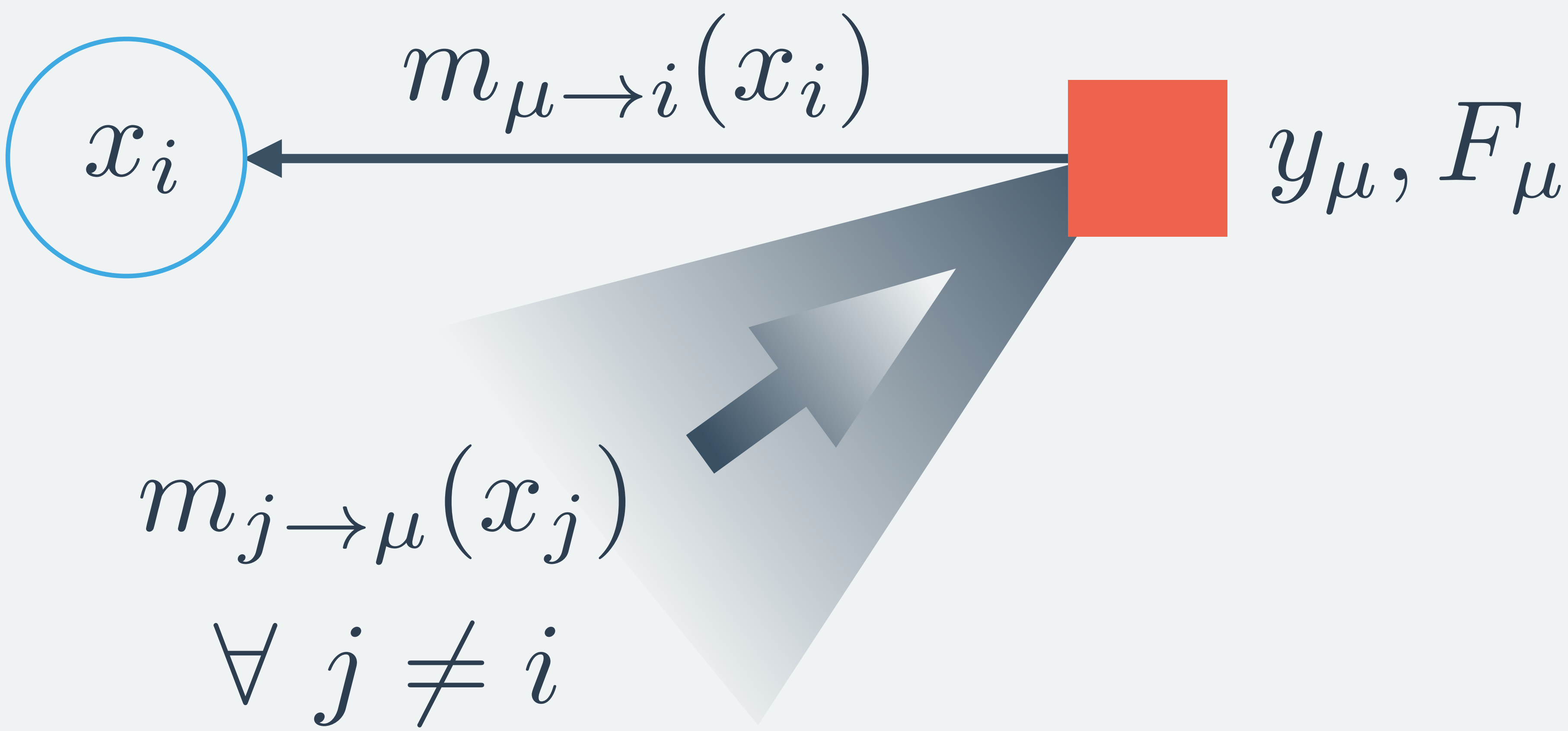
Variables
Coefficients

Factors
Measurements

BP on the Factor Graph

Factor to Variable Messages

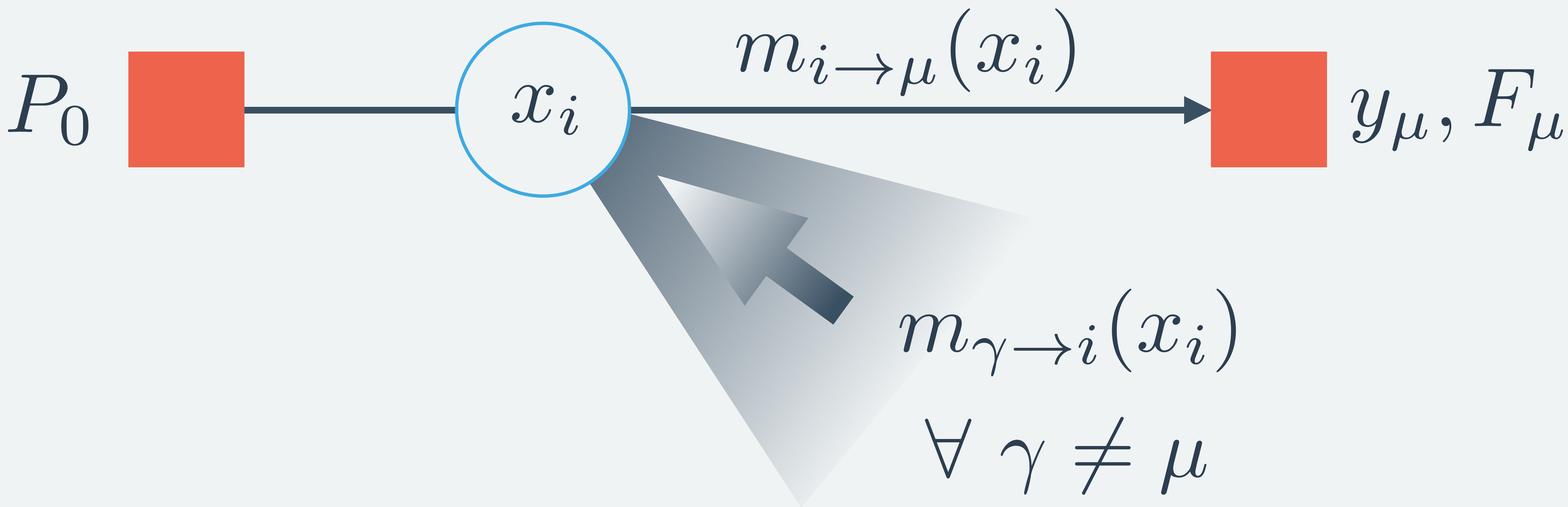
$$m_{\mu\to i}(x_i) = \frac{1}{Z_{\mu\to i}} \int \prod_{j\neq i} \mathrm{d}x_j \;\; \exp\left\{ -\frac{1}{2\Delta} \left(\sum_{j\neq i} F_{\mu j} x_j + F_{\mu i} x_i - y_\mu \right)^2 \right\} m_{j\to\mu}(x_j)$$



BP on the Factor Graph

Variable to Factor Messages

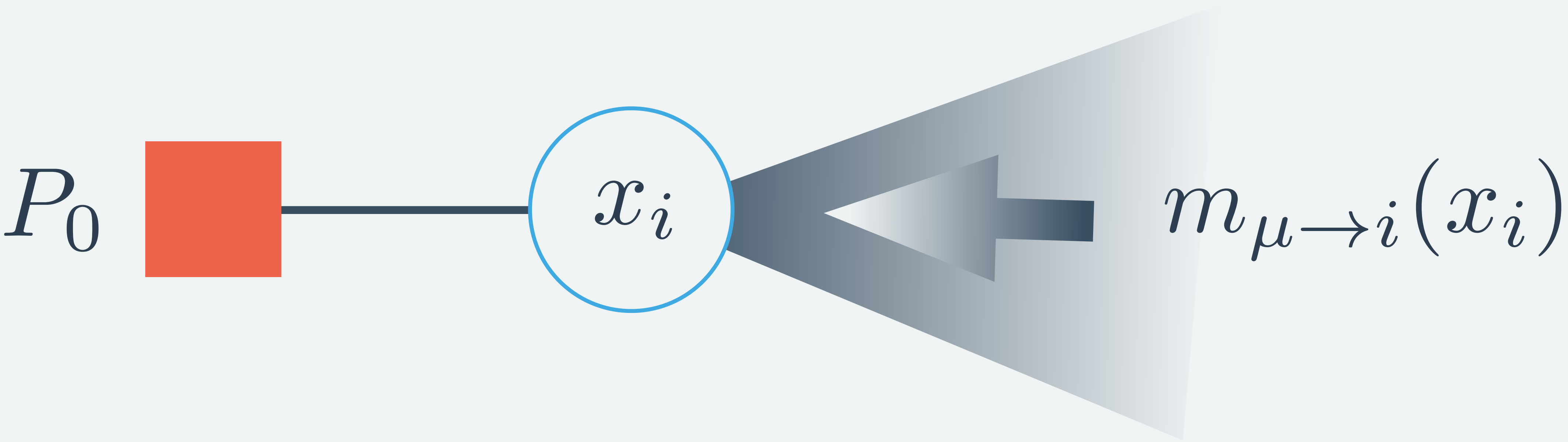
$$m_{i\to\mu}(x_i) = \frac{1}{Z_{i\to\mu}} P_0(x_i) \prod_{\gamma\neq\mu} m_{\gamma\to i}(x_i)$$



BP on the Factor Graph

Final Factorization

$$Q(x_i|\mathbf{y},F) = \frac{1}{Z_i} P_0(x_i) \prod_{\mu} m_{\mu\to i}(x_i)$$



Real Valued BP…

Issue

Messages are continuous functions !

Computability

Can we parameterize these messages somehow ?

Making BP Tractable

Relaxed BP (r-BP)
Retain the mean and variance of the messages.
*Assumes values of **F** are small w.r.t. **N**: **O(1/√N)***

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

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

Messages can be written in terms of these two moments
via Hubbard-Stratonovich transform & a second-order Taylor expansion about 0.

Relaxed BP

r-BP Equations

$$A_{\mu\to i} = \frac{F_{\mu i}^2}{\Delta + \sum_{j\neq i} F_{\mu j}^2 v_{j\to\mu}}$$

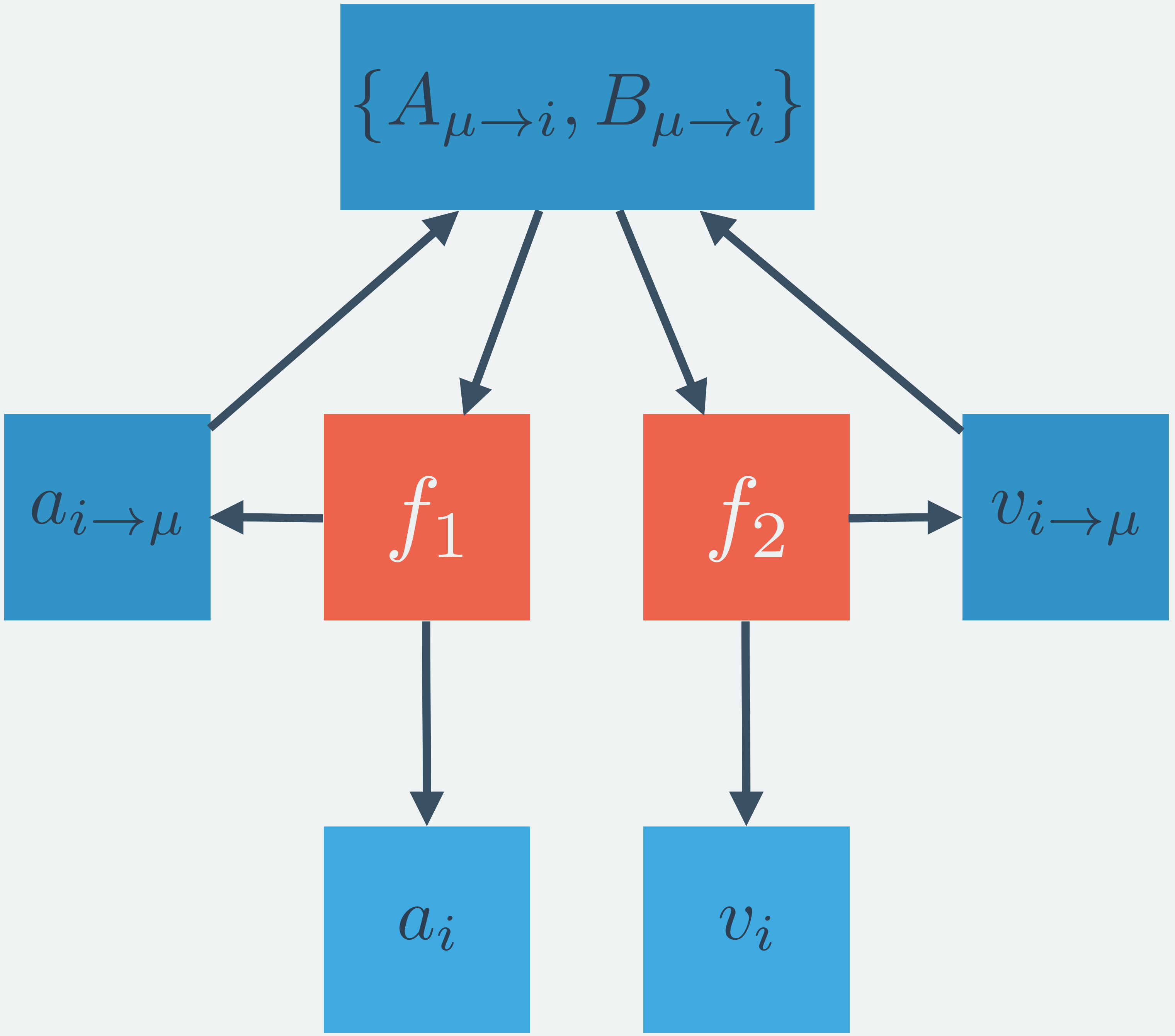
$$B_{\mu\to i} = \frac{F_{\mu i}(y_\mu - \sum_{j\neq i} F_{\mu j} a_{j\to\mu})}{\Delta + \sum_{j\neq i} F_{\mu j}^2 v_{j\to\mu}}$$

$$a_{i\to\mu} = f_1\left(\frac{1}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}, \frac{\sum_{\gamma\neq\mu} B_{\gamma\to i}}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}\right)$$

$$v_{i\to\mu} = f_2\left(\frac{1}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}, \frac{\sum_{\gamma\neq\mu} B_{\gamma\to i}}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}\right)$$

$$a_i = f_1\left(\frac{1}{\sum_\mu A_{\mu\to i}}, \frac{\sum_\mu B_{\mu\to i}}{\sum_\mu A_{\mu\to i}}\right)$$

$$v_i = f_2\left(\frac{1}{\sum_\mu A_{\mu\to i}}, \frac{\sum_\mu B_{\mu\to i}}{\sum_\mu A_{\mu\to i}}\right)$$



Relaxed BP

r-BP Equations

$$A_{\mu\to i} = \frac{F_{\mu i}^2}{\Delta + \sum_{j\neq i} F_{\mu j}^2 v_{j\to\mu}}$$

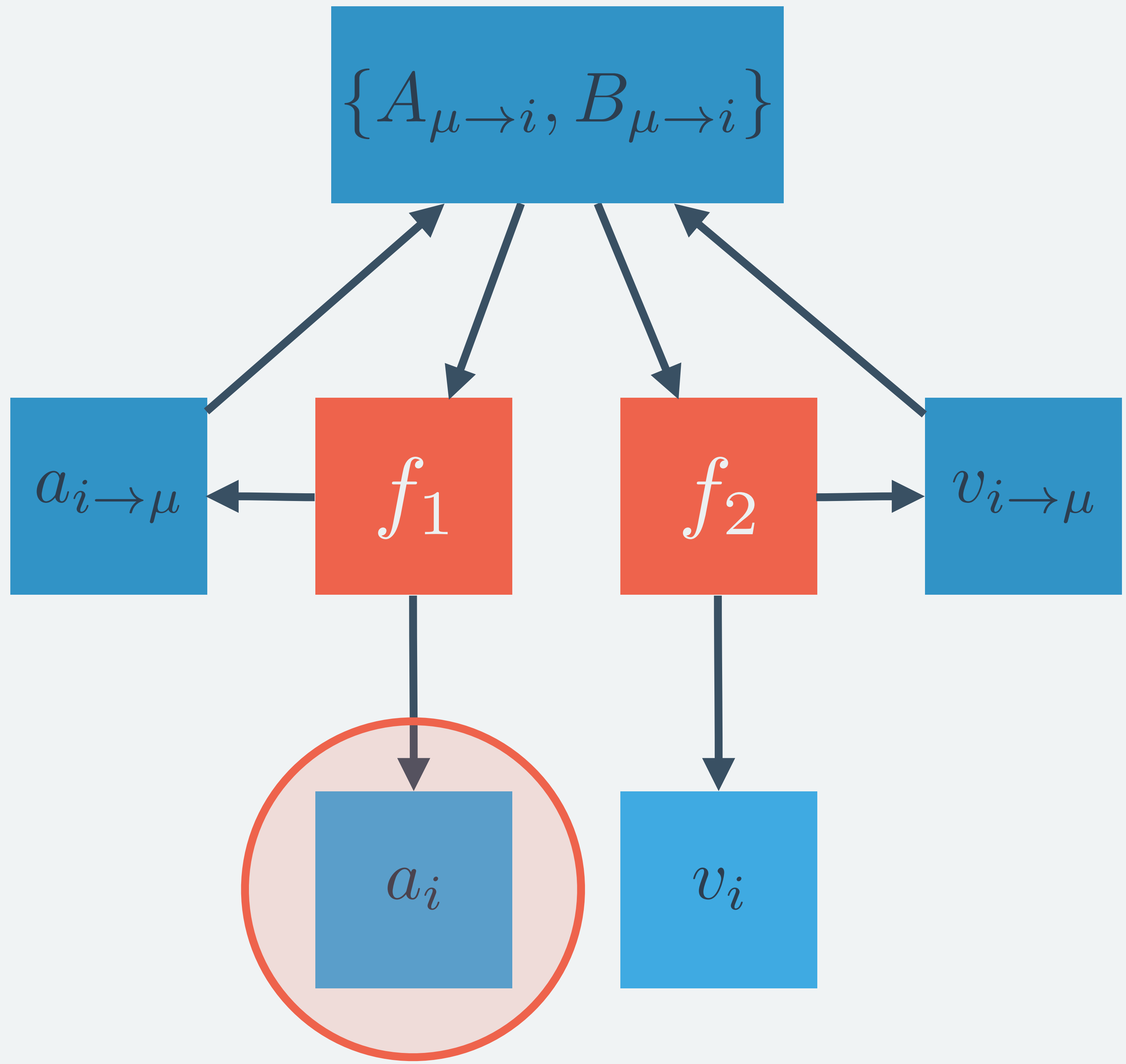
$$B_{\mu\to i} = \frac{F_{\mu i}(y_\mu - \sum_{j\neq i} F_{\mu j} a_{j\to\mu})}{\Delta + \sum_{j\neq i} F_{\mu j}^2 v_{j\to\mu}}$$

$$a_{i\to\mu} = f_1\left(\frac{1}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}, \frac{\sum_{\gamma\neq\mu} B_{\gamma\to i}}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}\right)$$

$$v_{i\to\mu} = f_2\left(\frac{1}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}, \frac{\sum_{\gamma\neq\mu} B_{\gamma\to i}}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}\right)$$

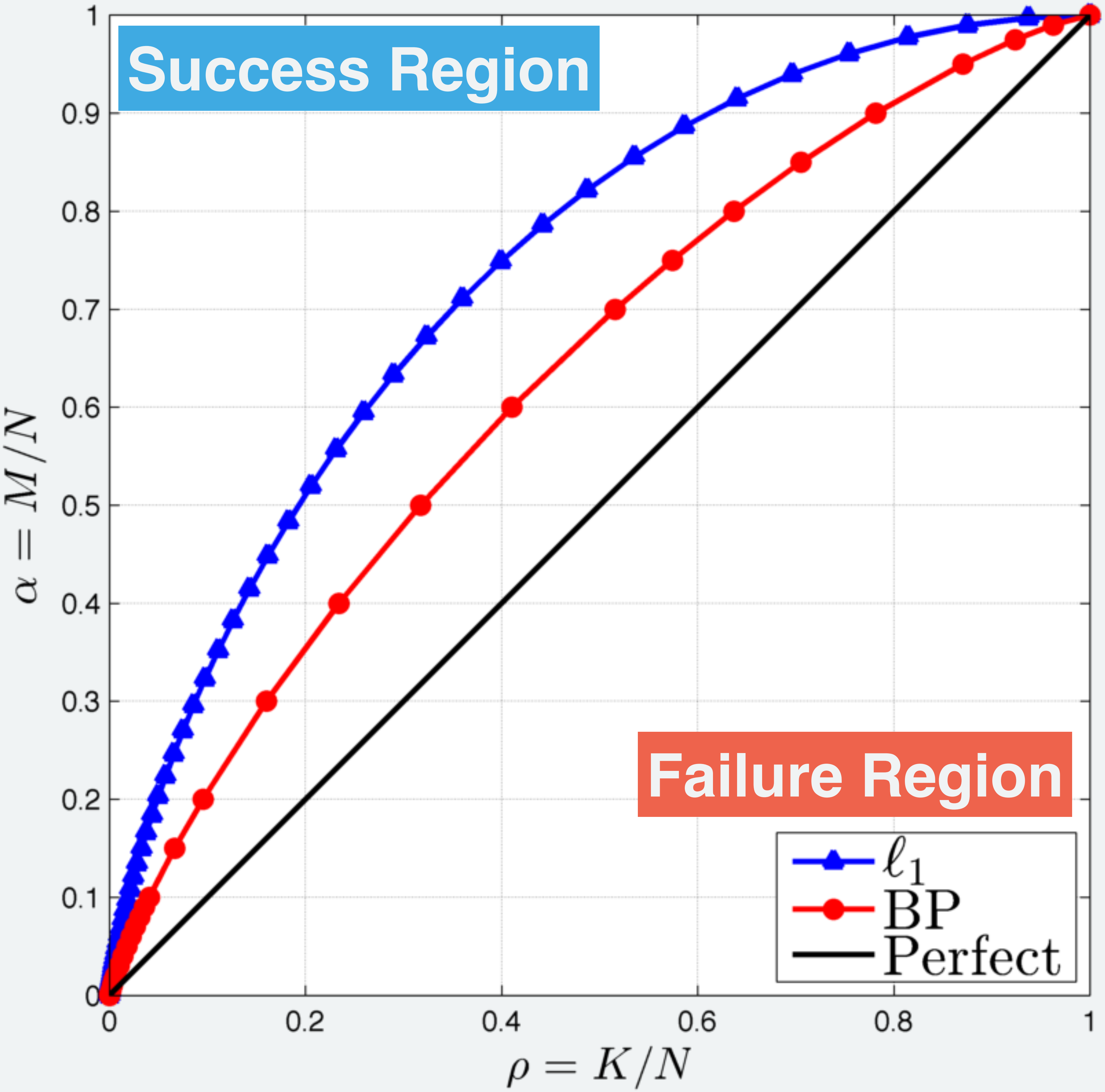
$$a_i = f_1\left(\frac{1}{\sum_\mu A_{\mu\to i}}, \frac{\sum_\mu B_{\mu\to i}}{\sum_\mu A_{\mu\to i}}\right)$$

$$v_i = f_2\left(\frac{1}{\sum_\mu A_{\mu\to i}}, \frac{\sum_\mu B_{\mu\to i}}{\sum_\mu A_{\mu\to i}}\right)$$



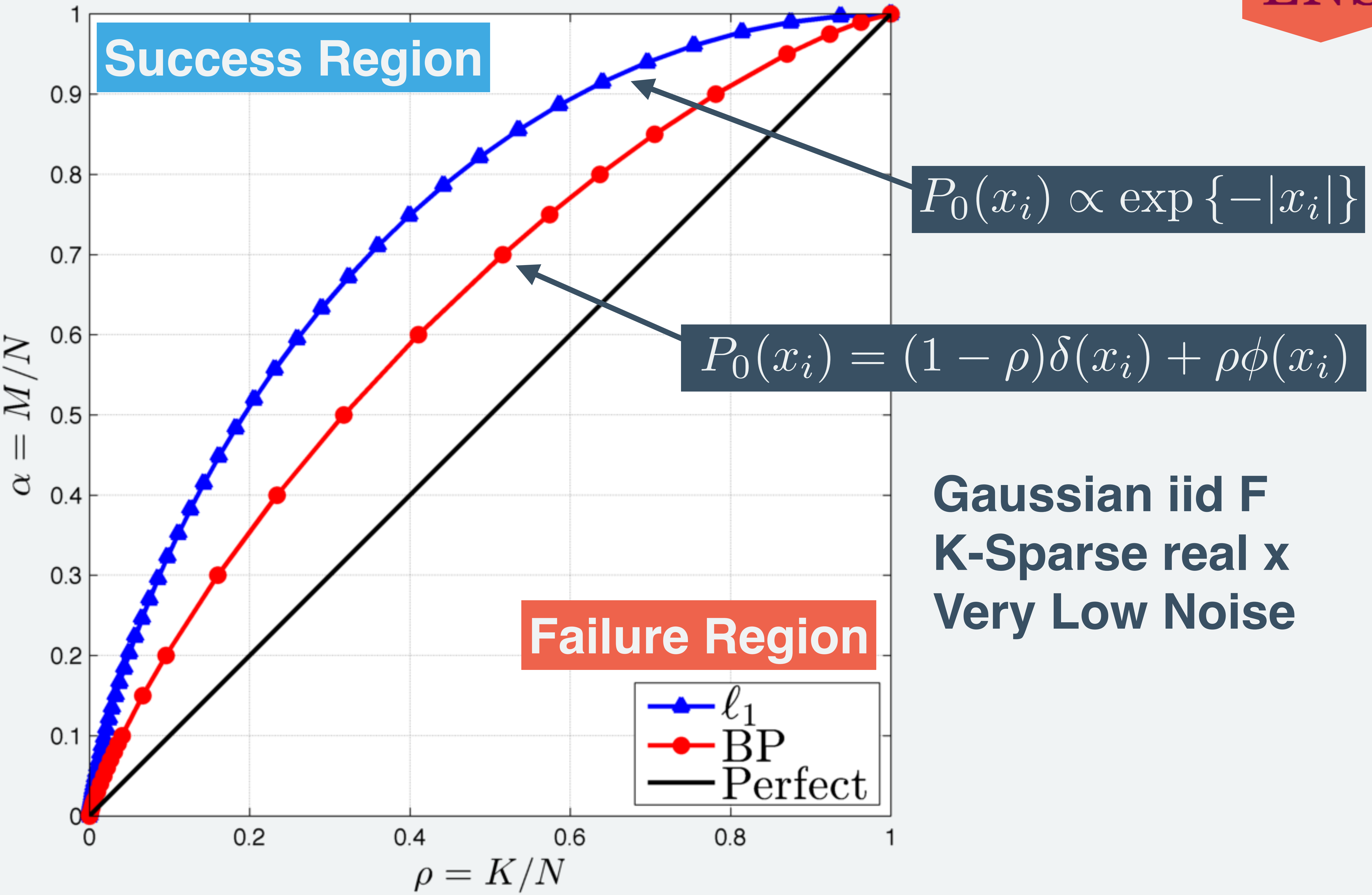
MMSE Signal Reconstruction!

BP Transition Performance



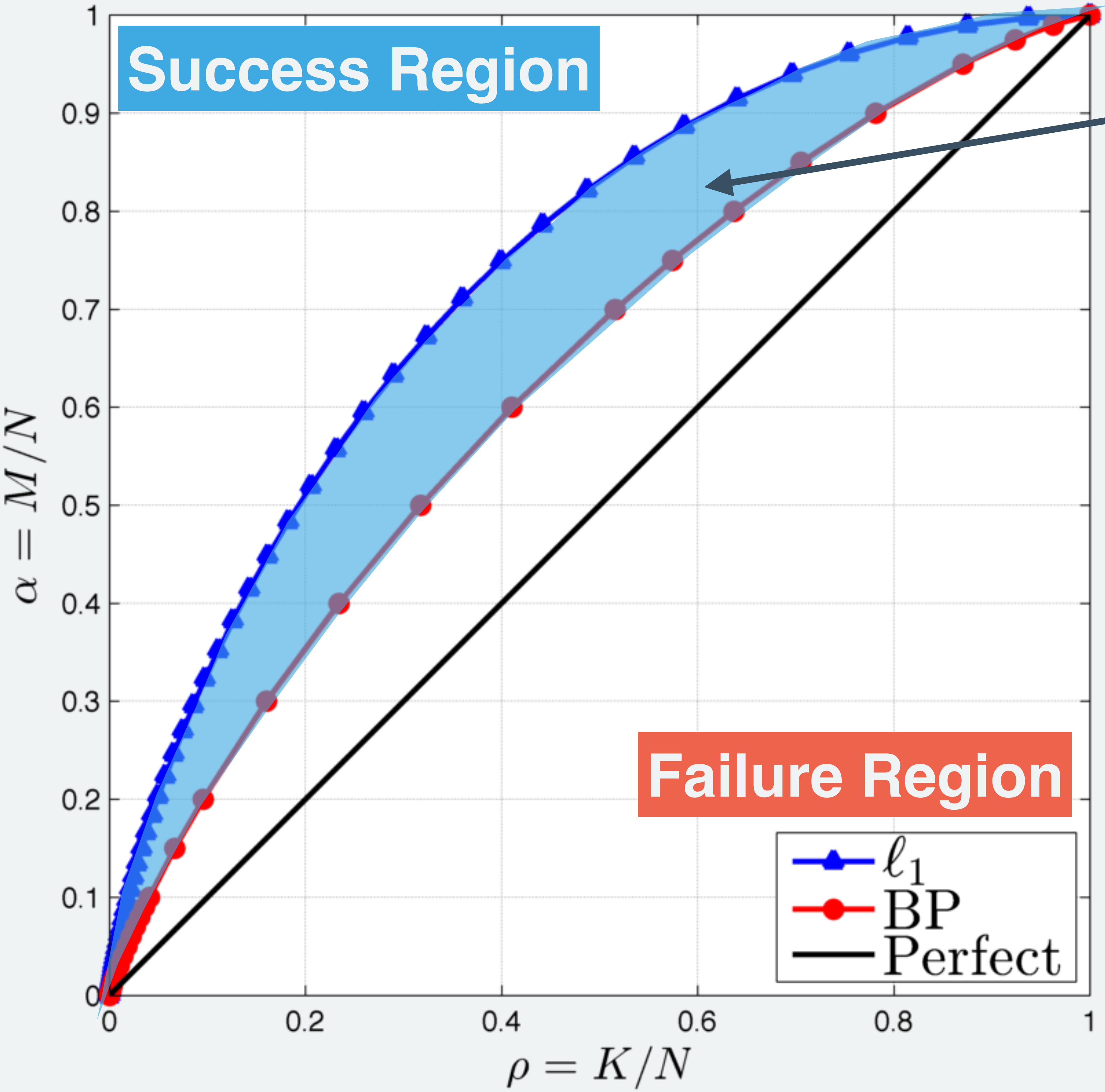
Gaussian iid F
K-Sparse real x
Very Low Noise

BP Transition Performance



Gaussian iid F
K-Sparse real x
Very Low Noise

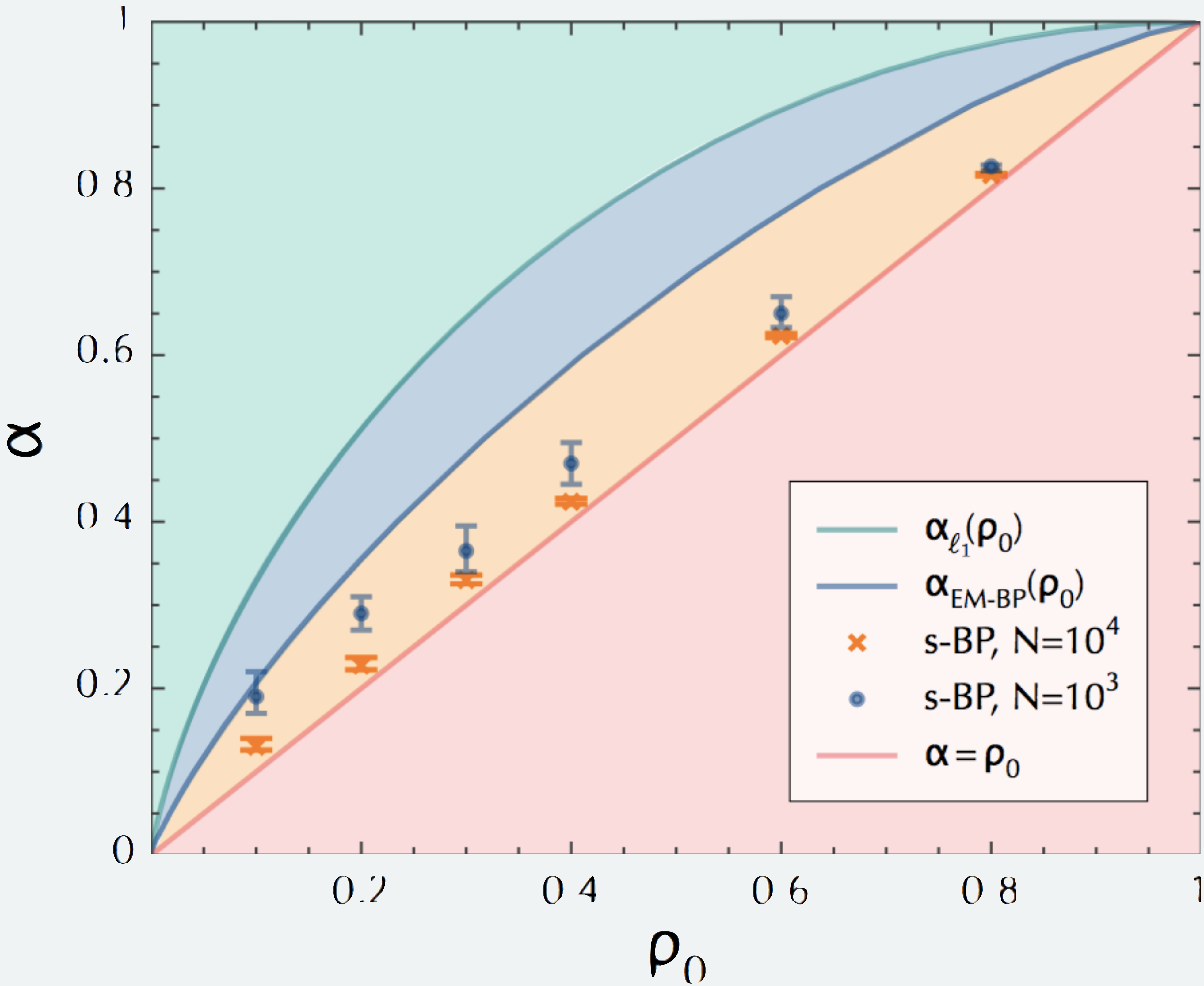
BP Transition Performance



Increase in set of solvable problems over L1-based Prior.

Gaussian iid F
K-Sparse real x
Very Low Noise

Optimal Transition w/ Seeds



Seeded F
K-Sparse real x
Very low noise

from Krzakala et al, "Statistical-physics-based reconstruction in compressed sensing," 2012.

r-BP to AMP via TAP

r-BP
Great performance for CS reconstruction of K-Sparse signals

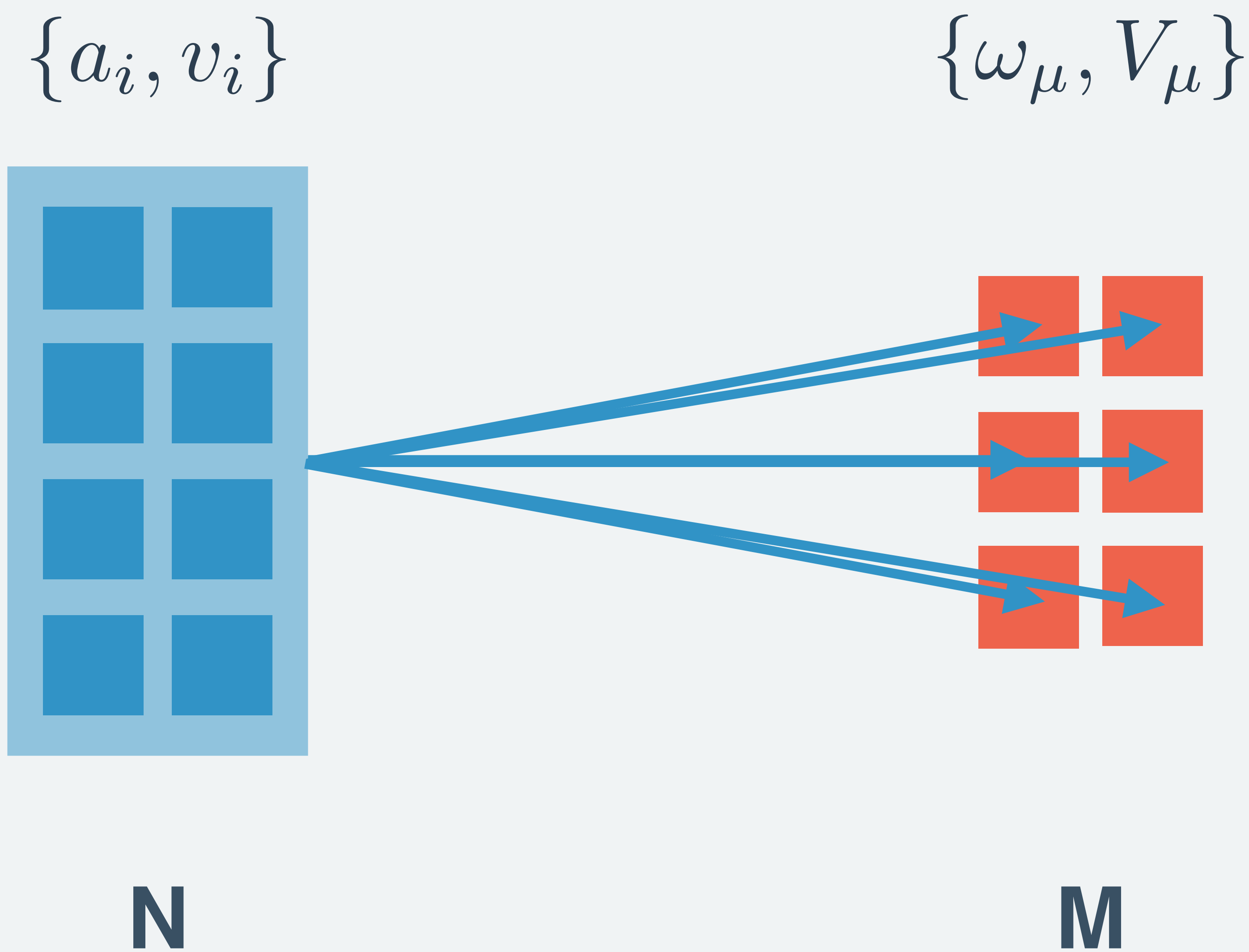
Issue
Computational & memory requirements scale with edges.

Solution
Use Thouless-Anderson-Palmer (TAP) approach to define algorithm on variables rather than edges !

r-BP to AMP via TAP

TAP Intuition

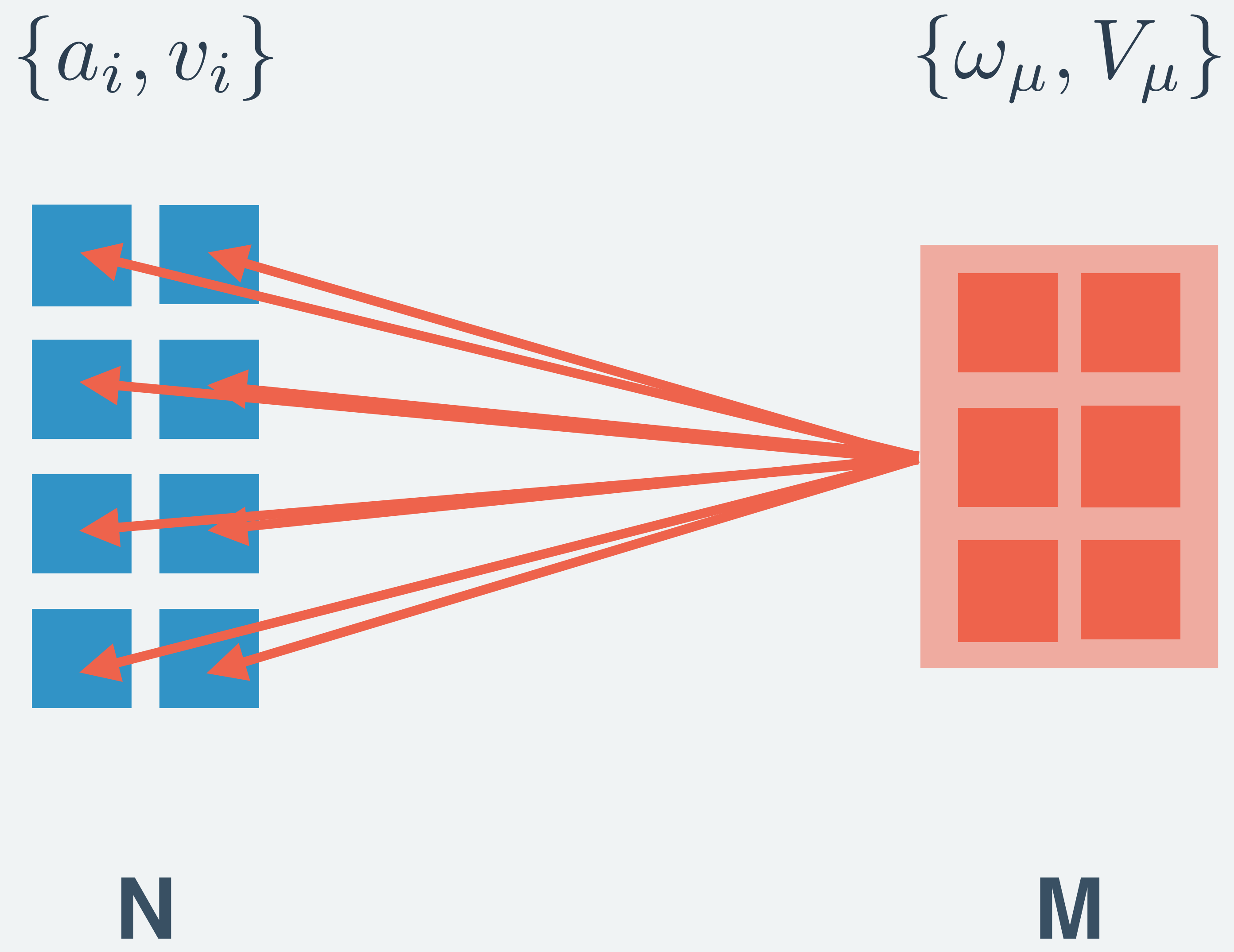
If F is ***not sparse*** and if its entries scale ***O(1/√N)***, then message means and variances are ***nearly independent*** of any one factor ***μ*** in the limit ***N→∞***.



r-BP to AMP via TAP

TAP Intuition

If F is ***not sparse*** and if its entries scale ***O(1/√N)***, then message means and variances are ***nearly independent*** of any one factor ***μ*** in the limit ***N→∞***.



r-BP to AMP via TAP

r-BP

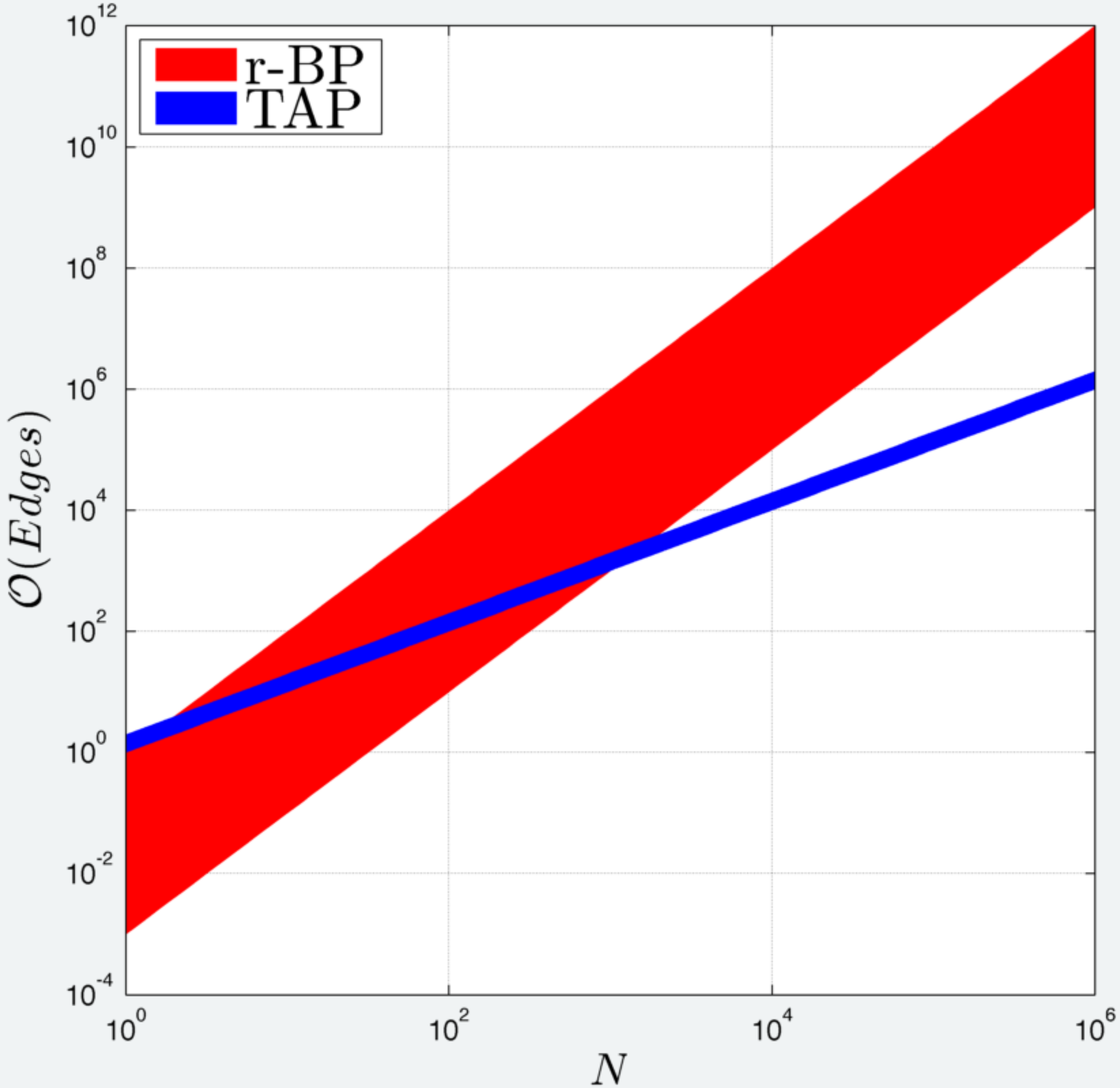
Messages/Edges scale as

$$\mathcal{O}(MN) = (\alpha N^2)$$

TAP on r-BP

Messages/Edges scale as

$$\mathcal{O}(M + N) = ((1 + \alpha)N)$$



Sum-Product AMP Algorithm

In Its Totality...

$$V_\mu^{t+1} = \sum_i F_{\mu i}^2 v_i^t$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

$$(\Sigma_i^{t+1})^2 = \left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = a_i^t + (\Sigma_i^{t+1})^2 \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu^{t+1})}{\Delta + V_\mu^{t+1}}$$

$$a_i^{t+1} = f_1((\Sigma_i^{t+1})^2, R_i)$$

$$v_i^{t+1} = f_2((\Sigma_i^{t+1})^2, R_i)$$

Sum-Product AMP Algorithm

Simplification

If we assume a Gaussian iid projector **F**, then we can say

$$F_{\mu i}^2 \approx \frac{1}{N}$$

Giving us…

$$\omega = Fa - \frac{\langle v \rangle}{\Delta + \langle v \rangle}(y - \omega)$$

$$R = a + \frac{1}{\alpha} F^T (y - \omega)$$

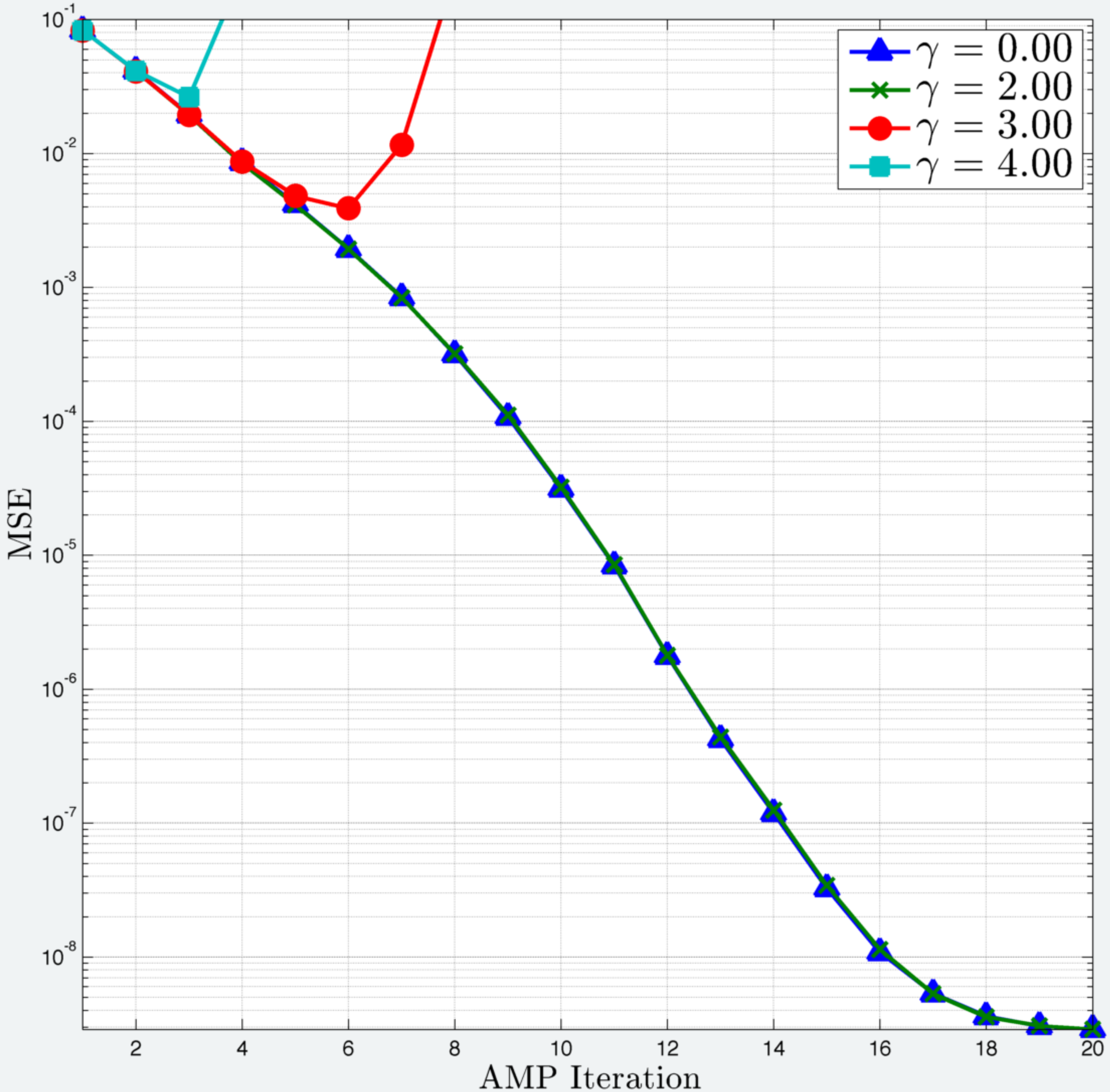
$$a = f_1\left(\frac{1}{\alpha}(\Delta + \langle v \rangle), R\right)$$

$$v = f_2\left(\frac{1}{\alpha}(\Delta + \langle v \rangle), R\right)$$

AMP Divergence

The Big Obstacle
AMP diverges when **F** strays from zero-mean Gaussian iid !

Pathological Case: NZ Mean



$$F_{\mu i} \sim \mathcal{N}\left(\frac{\gamma}{N}, \frac{1}{N}\right)$$

$$N = 2048$$

$$\Delta = 10^{-8}$$

$$\alpha = 0.4$$

$$\rho_0 = 0.1$$

$$\phi \sim \mathcal{N}(0,1)$$


Everything Cool
Sparse Matrices
Low-Rank Matrices
Super-resolution
Deblurring & Deconvolution

Some Approaches

Damping

Known to practitioners for a while, though without rigor.

$$(\Sigma_i^{t+1})^2 = \theta(\Sigma_i^t)^2 + (1-\theta)\left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = \theta R_i^t + (1-\theta)\left[a^t + \frac{1}{\alpha}F^T(y - \omega^{t+1})\right]$$

Issue

How do we choose θ to ensure convergence ?

Some Approaches

Damping Generalized AMP (GAMP)

Rangan et al, *On the convergence of approximate message passing with arbitrary matrices*, 2014

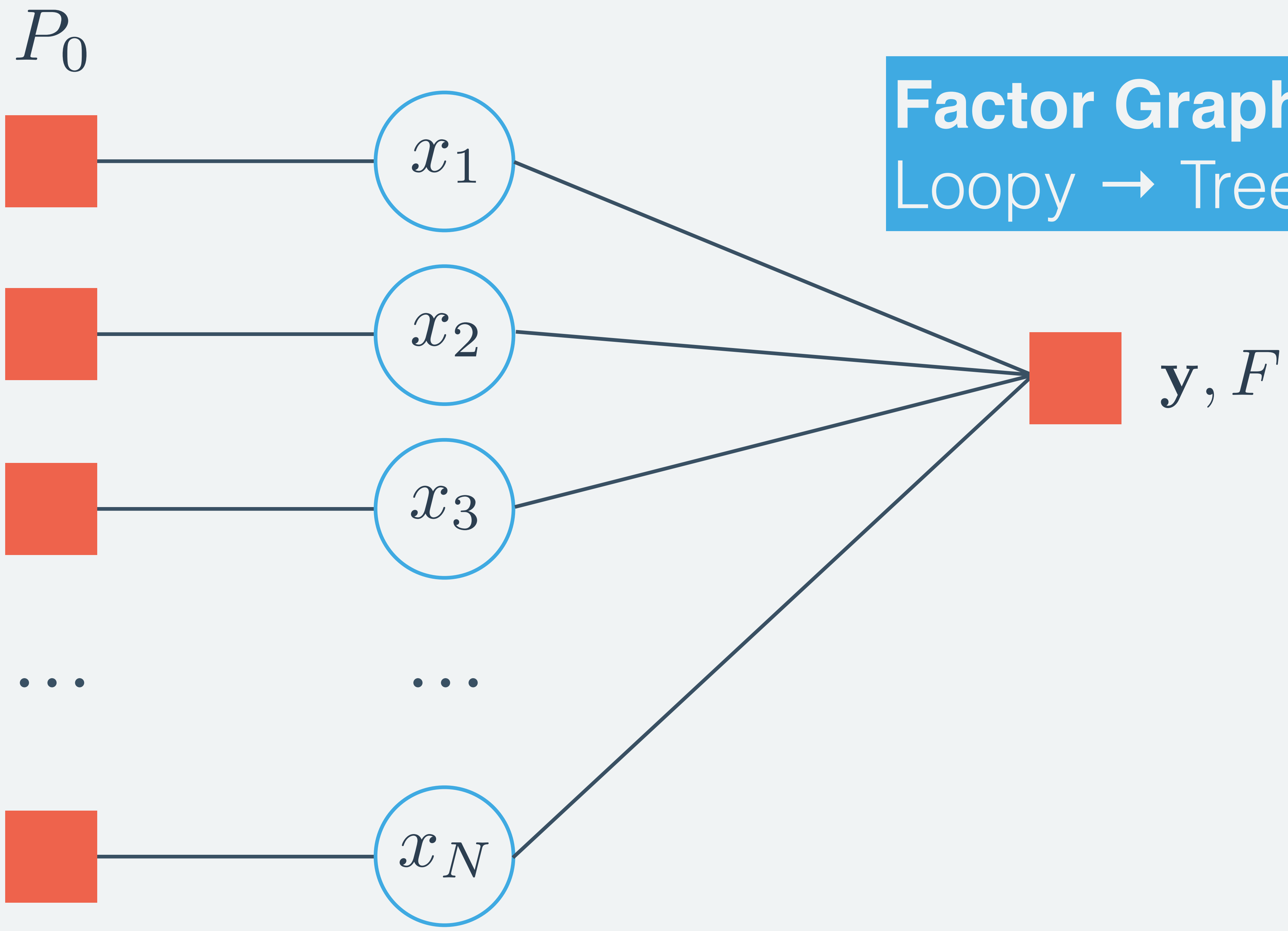
For Gaussian Prior…

Convergence-ensuring damping given by singular values of **F**.

S-AMP

S-Transform AMP (S-AMP)

Çakmak et al, *S-AMP: Approximate Message Passing for General Matrix Ensembles*, 2014



Factor Graph

Loopy → Tree

S-AMP

S-Transform AMP (S-AMP)

Çakmak et al, *S-AMP: Approximate Message Passing for General Matrix Ensembles*, 2014

$$s_F^{t-1} \triangleq S_F\left(-\left\langle \eta'_{t-1}(F^\dagger \mathbf{z}^{t-1} + \boldsymbol{\mu}^{t-1})\right\rangle\right)$$

$$\mathbf{z}^t = \mathbf{y} - F\boldsymbol{\mu}^t + \left(1 - \frac{1}{s_F^{t-1}}\right)\mathbf{z}^{t-1}$$

$$\mu^{t+1} = \eta(F^\dagger \mathbf{z}^t + \boldsymbol{\mu}^t)$$

S-Transform in Free Probability of the AED of $F^\dagger F$

Swept Coordinate Update

Sequential r-BP for non-zero mean F

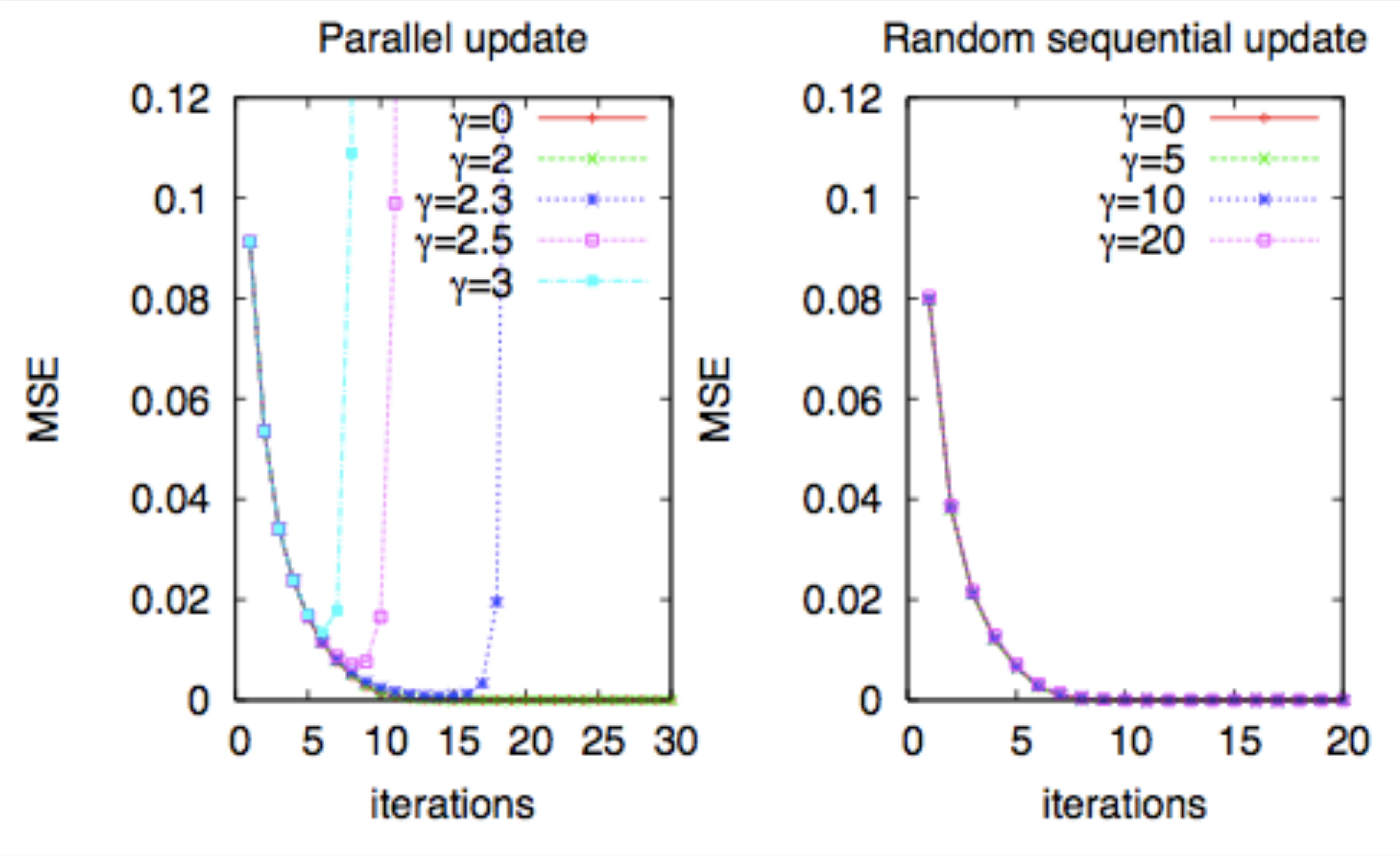
Caltagirone et al, *On Convergence of Approximate Message Passing*, 2014

$$F_{\mu,i} \sim \mathcal{N}\left(\frac{\gamma}{N}, \frac{1}{N}\right)$$

Nishimori Line Stability via State Evolution

Solution path can absorb small amounts of instability in the parallel update…*but too much prevents convergence!*

Swept Coordinate Update



from Caltagirone et al, On Convergence of Approximate Message Passing, 2014

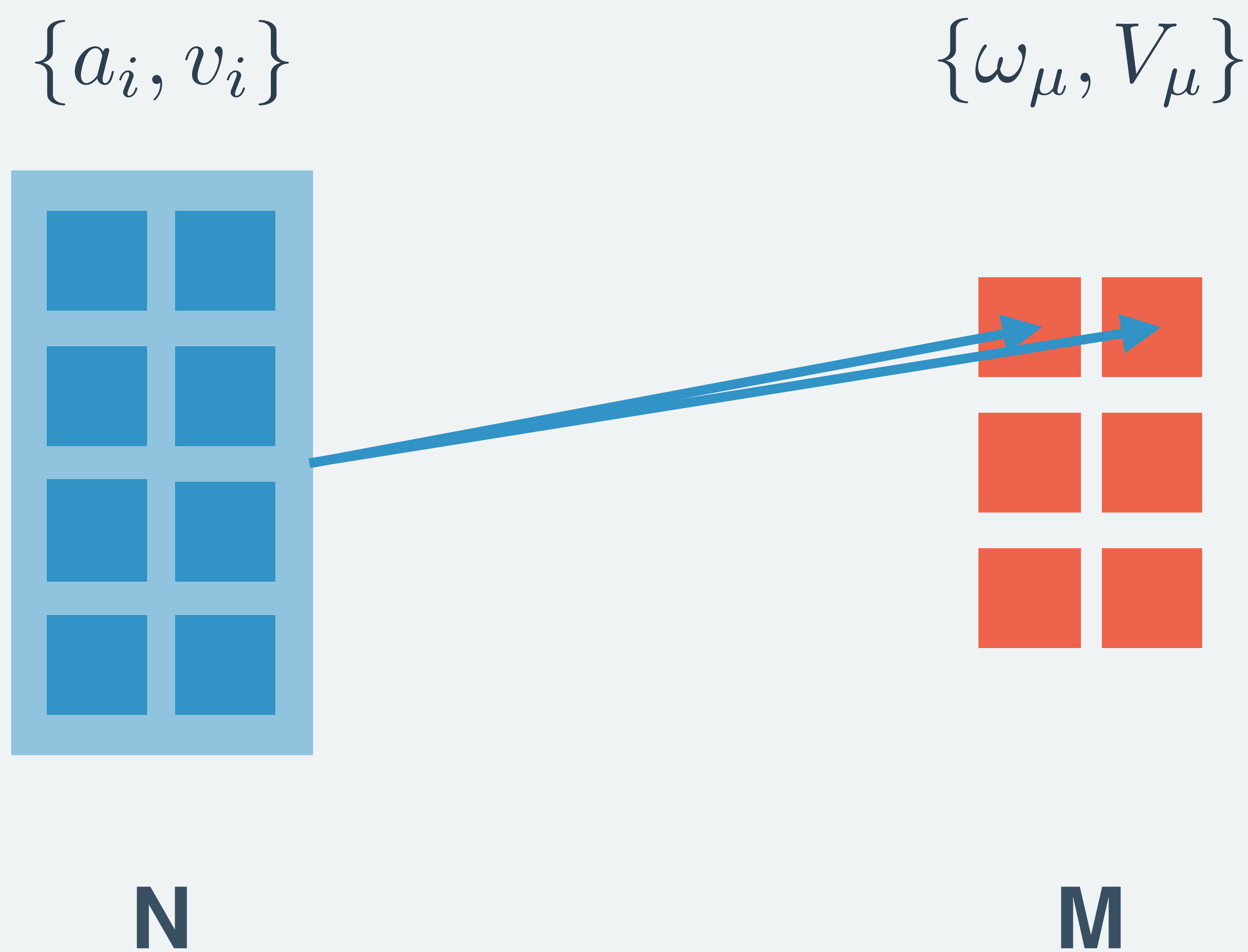
Swept Coordinate Update

Swept AMP (SwAMP)

Manoel et al, *Sparse Estimation with the Swept Approximated Message-Passing Algorithm*, 2014

Idea

Apply TAP to Sequential r-BP

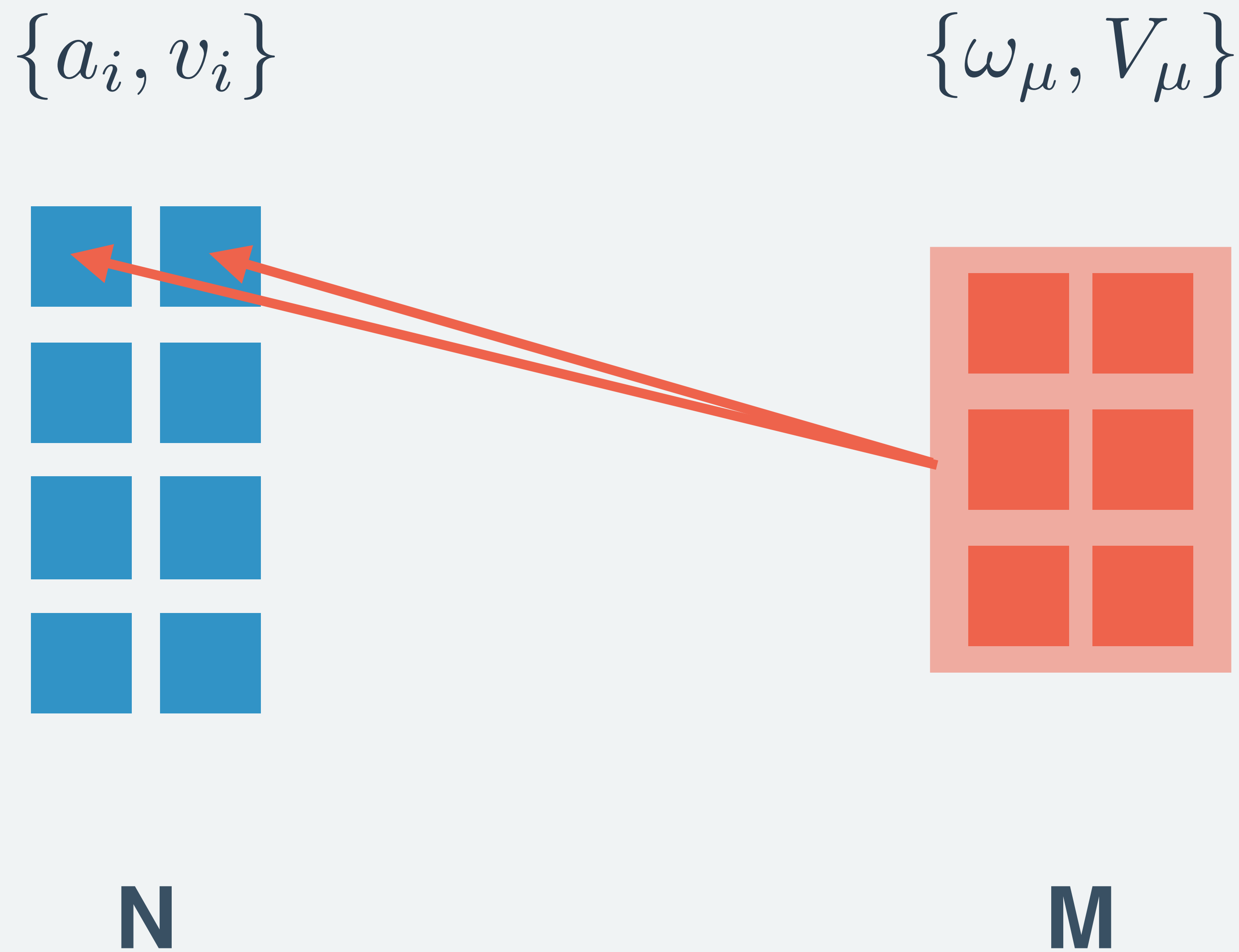


Swept Coordinate Update

Swept AMP (SwAMP)

Manoel et al, *Sparse Estimation with the Swept Approximated Message-Passing Algorithm*, 2014

Idea
Apply TAP to Sequential r-BP



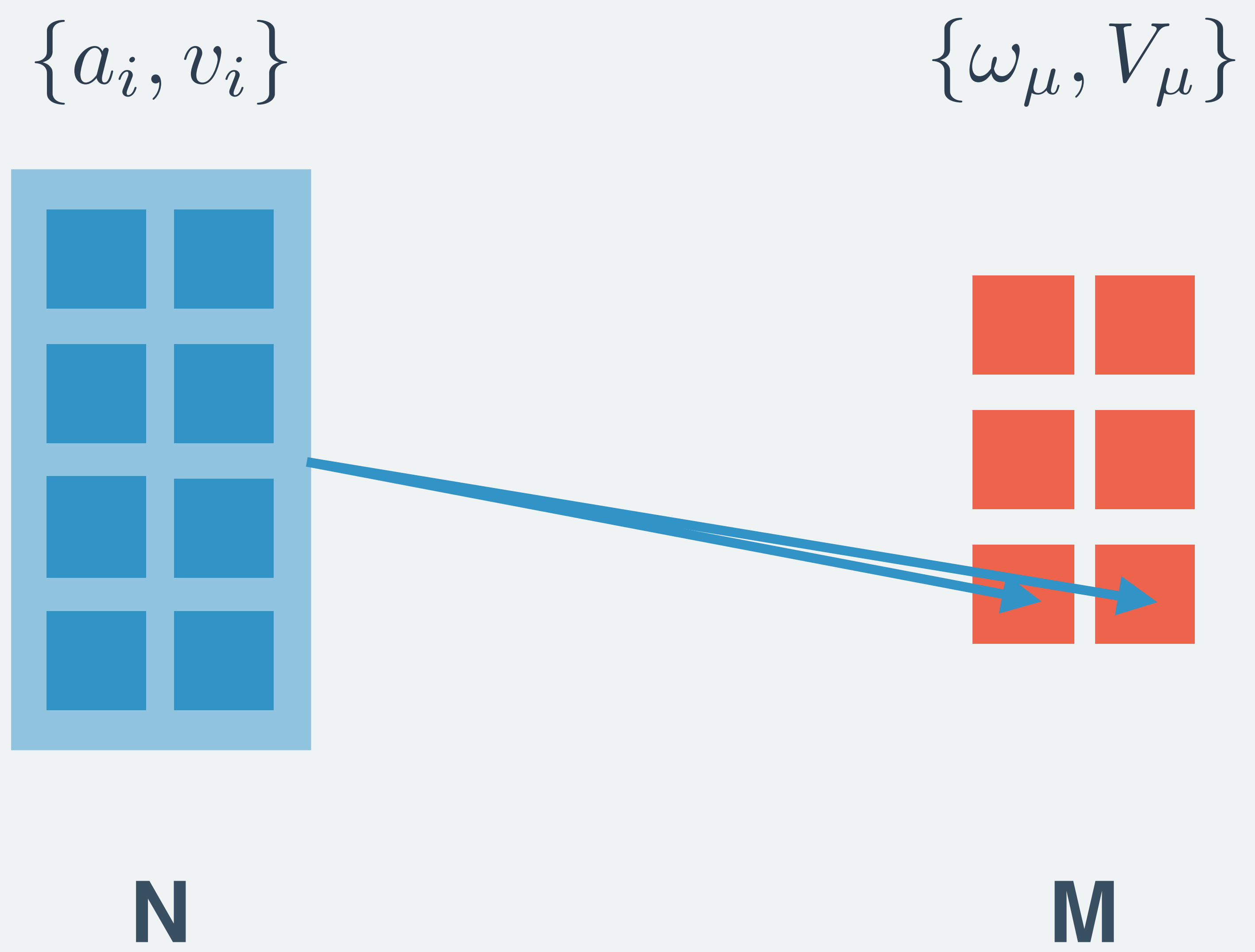
Swept Coordinate Update

Swept AMP (SwAMP)

Manoel et al, *Sparse Estimation with the Swept Approximated Message-Passing Algorithm*, 2014

Idea

Apply TAP to Sequential r-BP

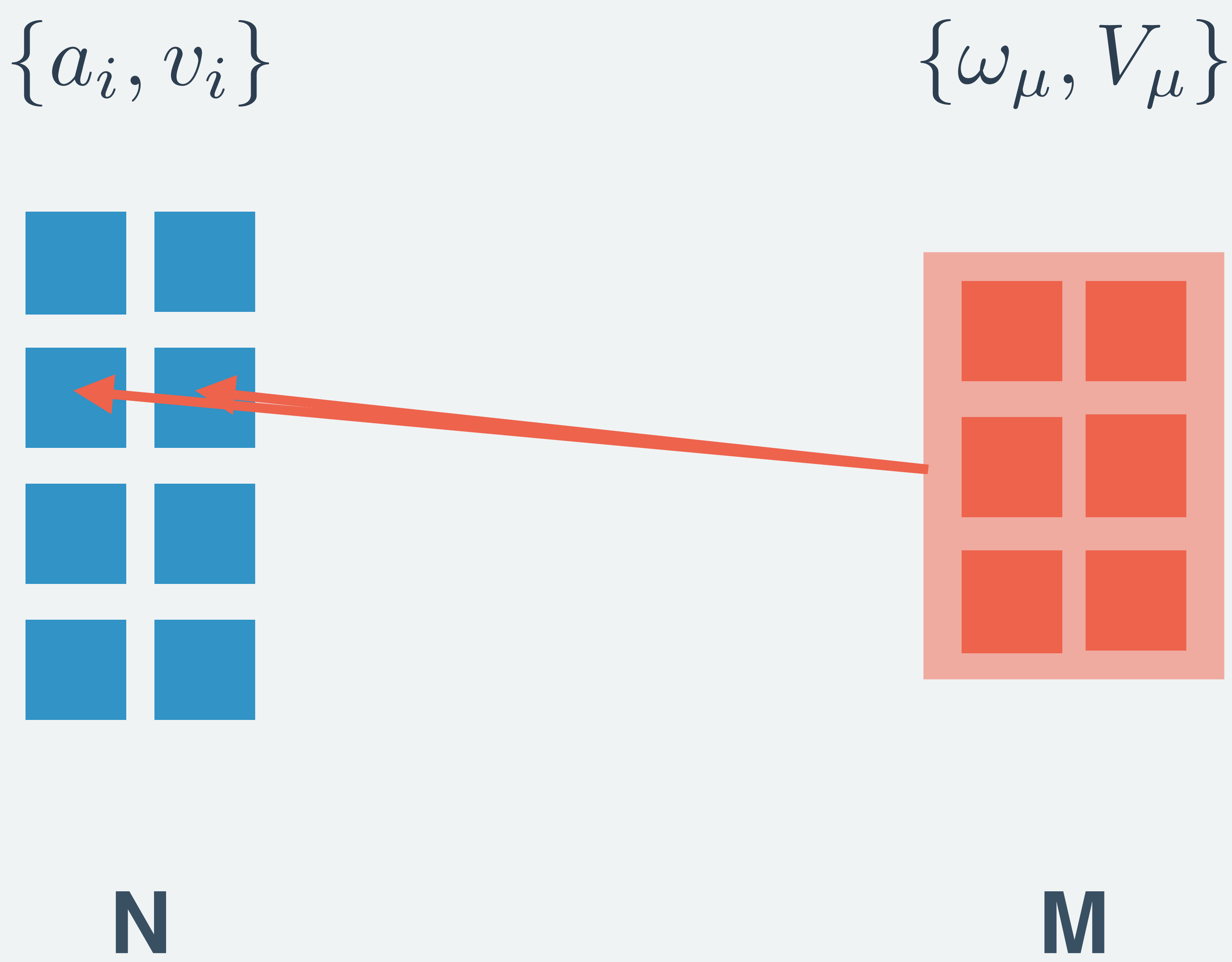


Swept Coordinate Update

Swept AMP (SwAMP)

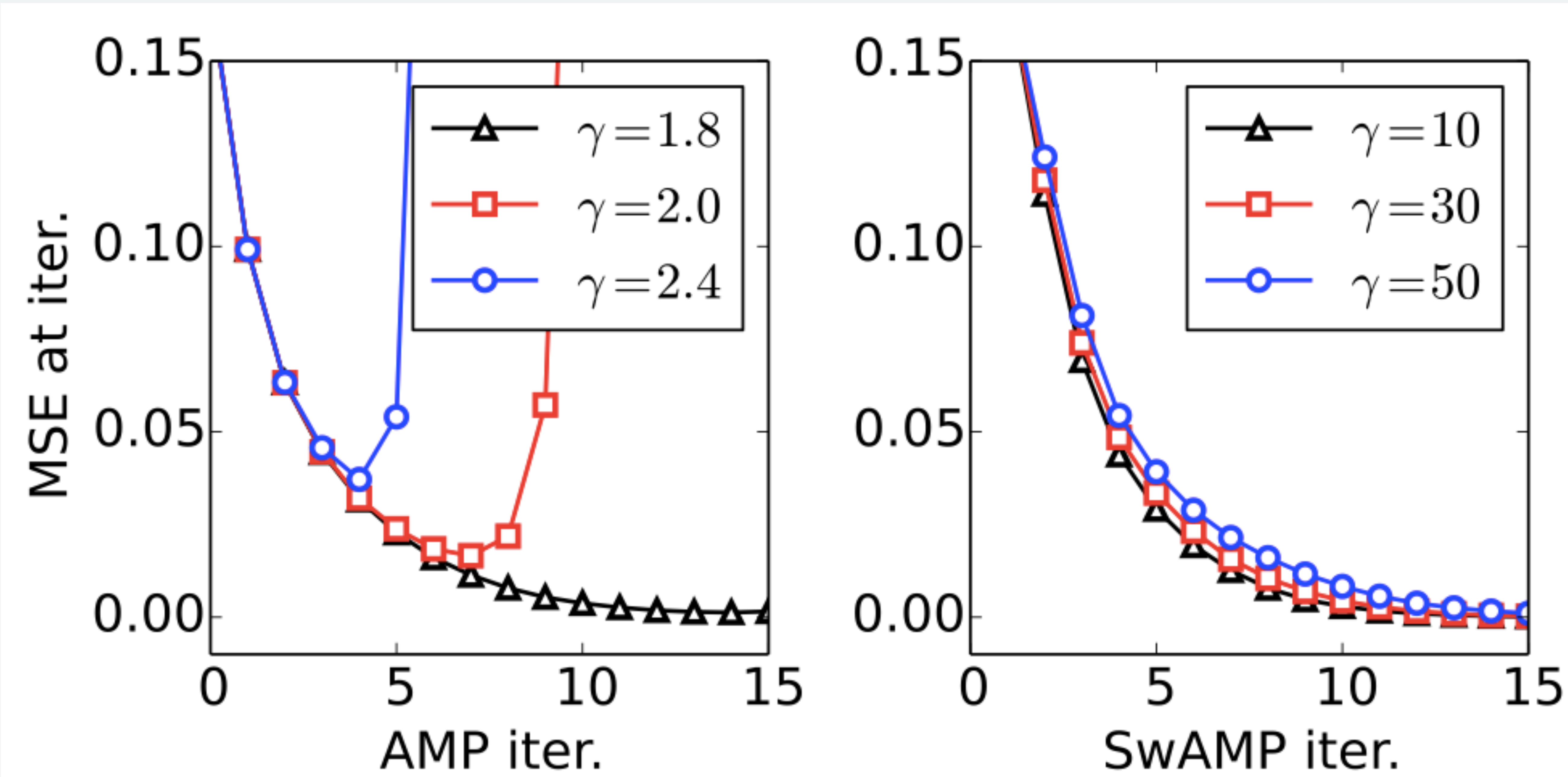
Manoel et al, *Sparse Estimation with the Swept Approximated Message-Passing Algorithm*, 2014

Idea
Apply TAP to Sequential r-BP



SwAMP

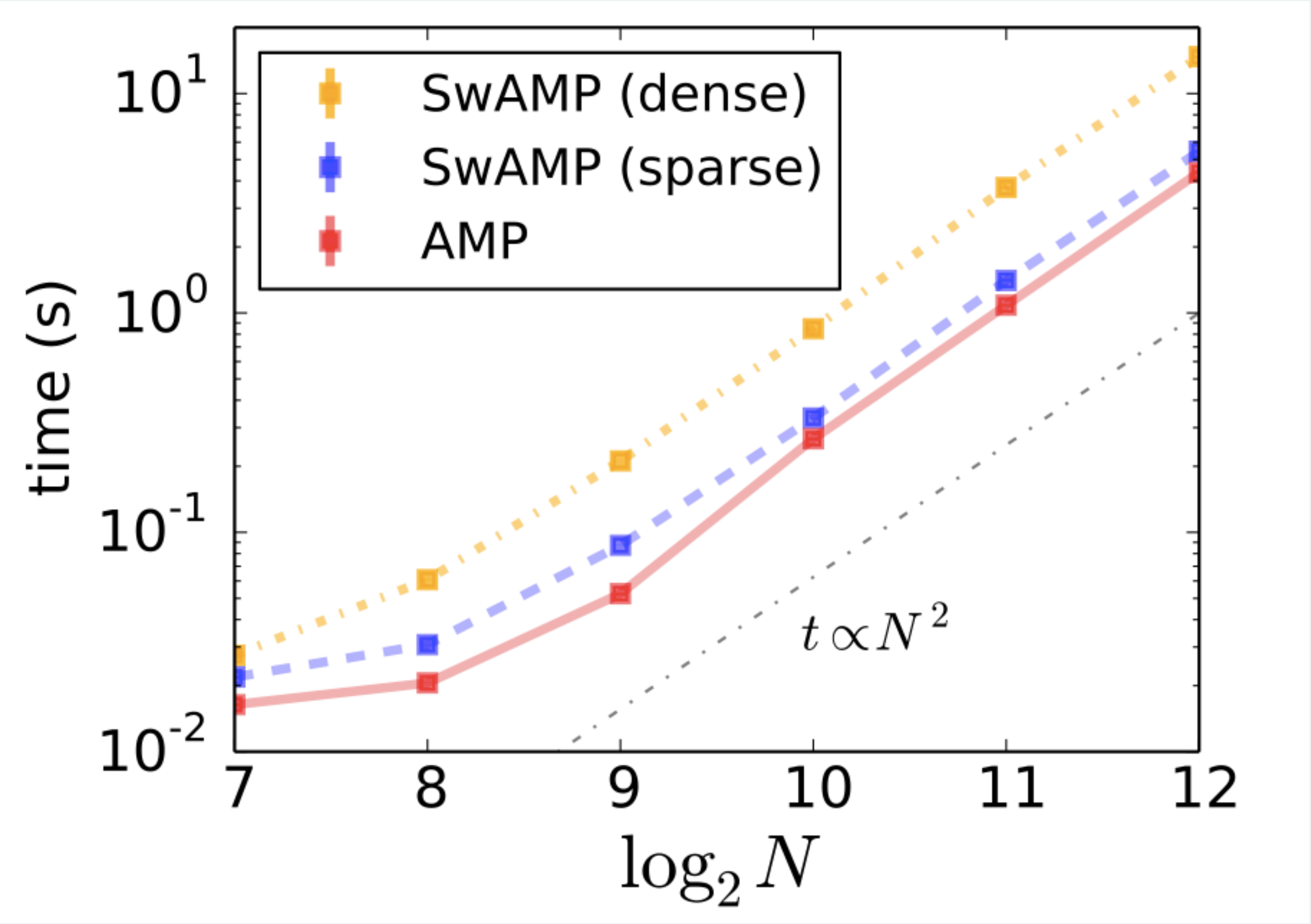
Matches sequential r-BP results !



from Manoel et al, Sparse Estimation with the Swept Approximated Message-Passing Algorithm, 2014

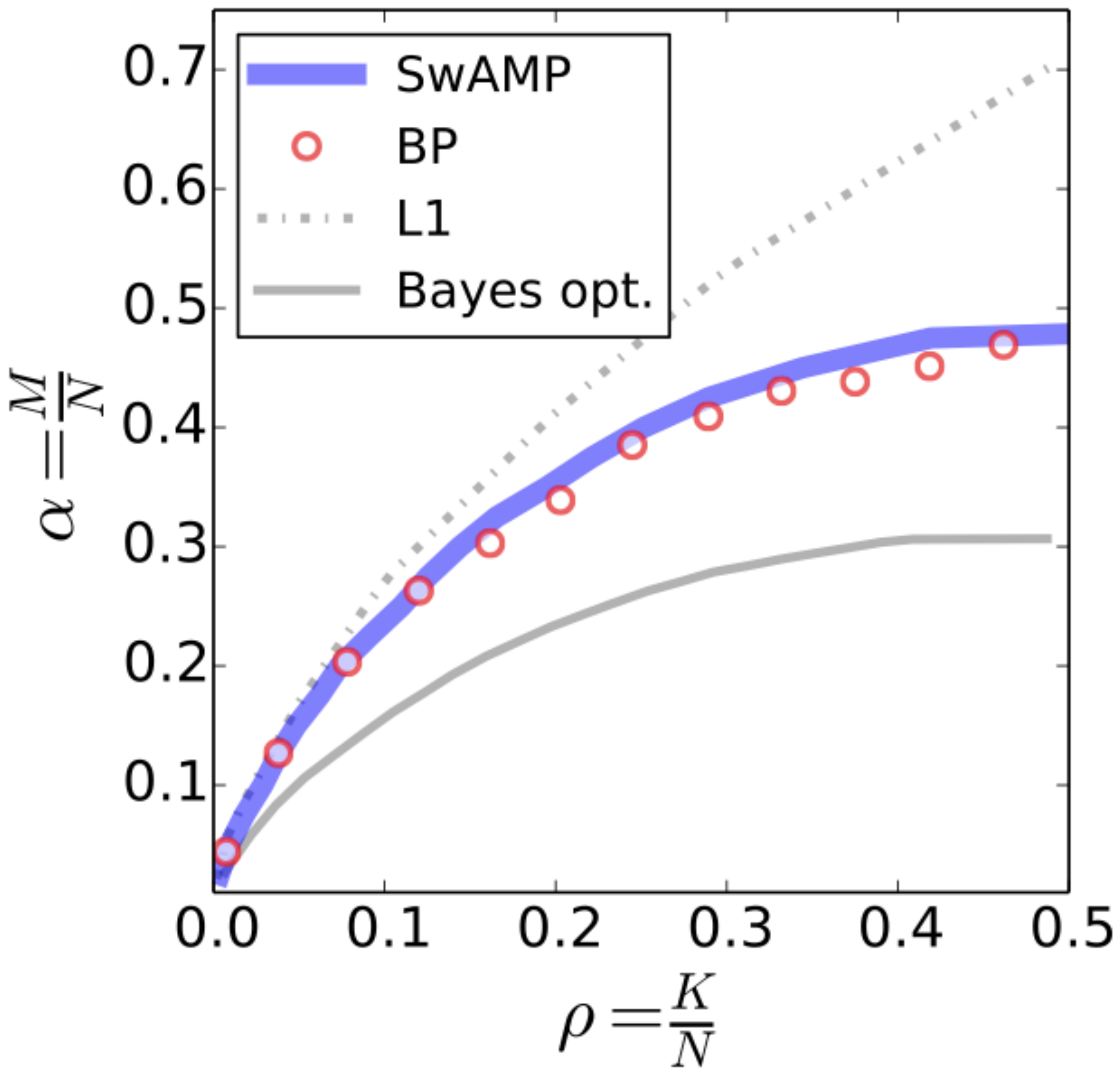
SwAMP

Performance Impact: *not so bad*



from Manoel et al, Sparse Estimation with the Swept Approximated Message-Passing Algorithm, 2014

SwAMP



Sparse Matrices
Works for
group testing, too!

$$F_{\mu i} \in \{0, 1\}$$
$$\sum_i F_{\mu i} = 7$$
$$x_i \in \{0, 1\}$$
$$\phi \sim \delta(x - 1)$$

from Manoel et al, Sparse Estimation with the Swept Approximated Message-Passing Algorithm, 2014

SwAMP

Sequential Approach to AMP

+ Avoids divergence in many cases
+ Works even with TAP assumptions explicitly violated
- Some cost in efficiency over parallel AMP

Open Questions

* What is the set of problems for which it ***doesn't* work**?
* Equivalent with r-BP?
* Is parallel FP-iteration doomed for wide class of problems w/o fundamental changes (i.e. S-AMP) ?

SwAMP

Available Online ! Try it out !

+ https://github.com/eric-tramel/SwAMP-Demo

Direct Free Energy Minimization

A Variational Ansatz

In Krzakala et al, *Variational Free Energies for Compressed Sensing*, 2014

$$\mathcal{F}(\{R_i\}, \{\Sigma_i\}) \triangleq$$

$$\frac{1}{2\Delta}\sum_{\mu}\left(y_\mu - \sum_i F_{\mu i} a_i\right)^2 + \frac{1}{2}\sum_\mu \log\left[1 + \sum_i F_{\mu i}^2 v_i/\Delta\right]$$

$$+ D_{KL}(Q||P_0) + \frac{M}{2}\log 2\pi\Delta$$

Equivalence

This ansatz has the same fixed-points as sum-product AMP (TAP + r-BP).

Direct Free Energy Minimization

A Variational Ansatz

In Krzakala et al, *Variational Free Energies for Compressed Sensing*, 2014

$$\mathcal{F}(\{R_i\},\{\Sigma_i\}) \triangleq$$

$a_i \triangleq f_1(\Sigma_i, R_i)$

$$\frac{1}{2\Delta}\sum_\mu \left(y_\mu - \sum_i F_{\mu i} a_i\right)^2 + \frac{1}{2}\sum_\mu \log\left[1 + \sum_i F_{\mu i}^2 v_i/\Delta\right]$$

$$+ D_{KL}(Q||P_0) + \frac{M}{2}\log 2\pi\Delta$$

$v_i \triangleq f_2(\Sigma_i, R_i)$

Function of **R**,**Σ**,**a**, and **v**

Direct Free Energy Minimization

A Variational Ansatz

In Krzakala et al, *Variational Free Energies for Compressed Sensing*, 2014

$$\mathcal{F}(\{R_i\},\{\Sigma_i\}) \triangleq$$

$$\frac{1}{2\Delta}\sum_\mu\left(y_\mu-\sum_i F_{\mu i}a_i\right)^2+\frac{1}{2}\sum_\mu \log\left[1+\sum_i F_{\mu i}^2 v_i/\Delta\right]$$

$$+D_{KL}(Q||P_0)+\frac{M}{2}\log 2\pi\Delta$$

Strictly Positive Terms

Shift Term & Lower Bound

Direct Free Energy Minimization

A Variational Ansatz

In Krzakala et al, *Variational Free Energies for Compressed Sensing*, 2014

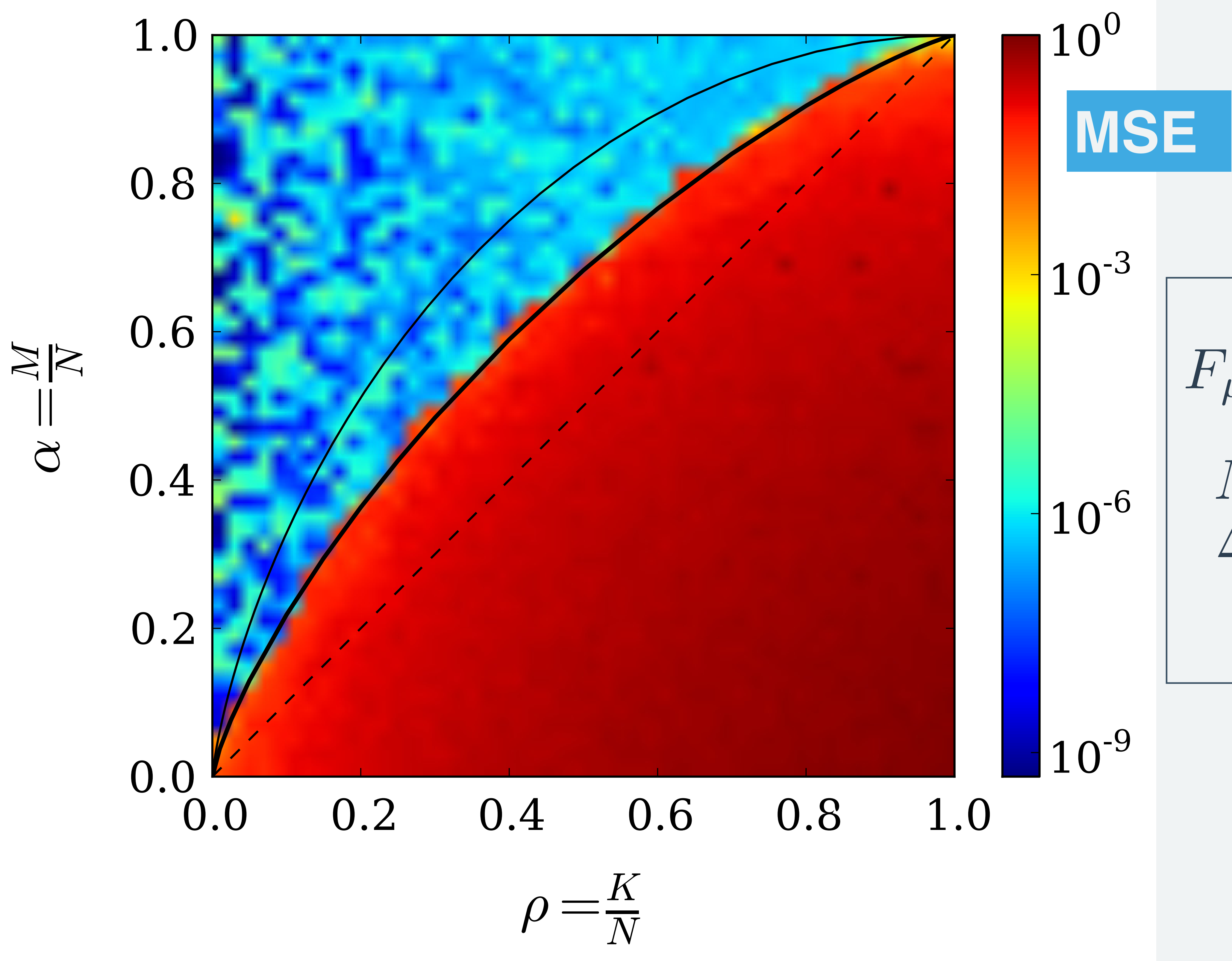
$$\mathcal{F}(\{R_i\},\{\Sigma_i\}) \triangleq$$

$$\frac{1}{2\Delta}\sum_{\mu}\left(y_\mu - \sum_i F_{\mu i} a_i\right)^2 + \frac{1}{2}\sum_{\mu}\log\left[1 + \sum_i F_{\mu i}^2 v_i/\Delta\right]$$

$$+ D_{KL}(Q||P_0) + \frac{M}{2}\log 2\pi\Delta$$

Idea
Minimize over this free energy directly!

Minimization…works!



$$F_{\mu i} \sim \mathcal{N}\left(0, \frac{1}{N}\right)$$

$$N = 2048$$
$$\Delta = 10^{-8}$$
$$\phi \sim \mathcal{N}(0,1)$$

from Krzakala et al, "Variational free energies for compressed sensing," 2014.

Direct Free Energy Minimization

Main Points

+ Opens a new approach to BP performance ***sans explicit message passing.***
- Perhaps not as efficient as FP-iteration

Open Questions

* Can it work for *any* choice of **F**?
* Equivalent with r-BP always?
* How can this minimization be approached more efficiently?
* How does this functional tie in with known optimization methods?

Thanks! Questions? Thanks!

Questions? Thanks! Questions?

Thanks! Questions? Thanks!

Questions? Thanks! Questions?

Thanks! Questions? Thanks!

Questions? Thanks! Questions?

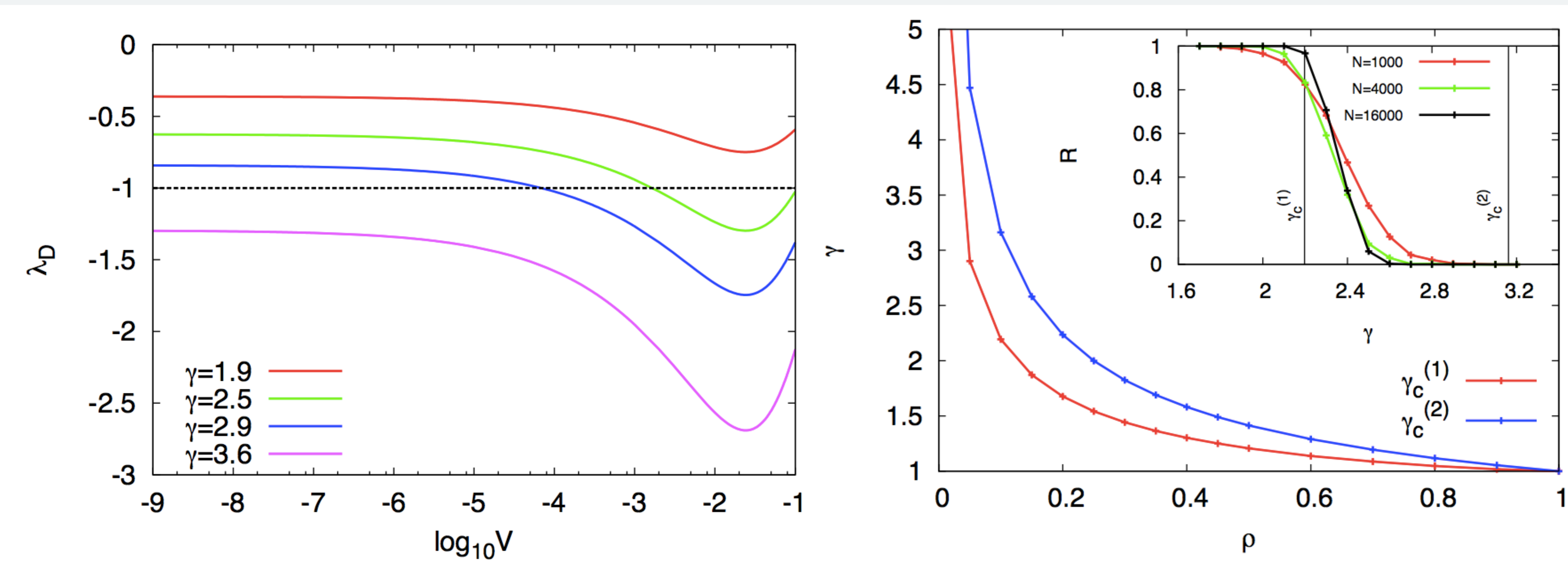
Swept Coordinate Update

ENS

Identification of Critical Means

$\gamma_c^{(1)}$ — Portions of the Nishimori Line are unstable

$\gamma_c^{(2)}$ — All of the Nishimori Line is unstable



from Caltagirone et al, On Convergence of Approximate Message Passing, 2014