Swept Approximate Message Passing for Sparse Estimation

Andre MANOEL
Univ. Sao Paulo

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

Eric W. Tramel

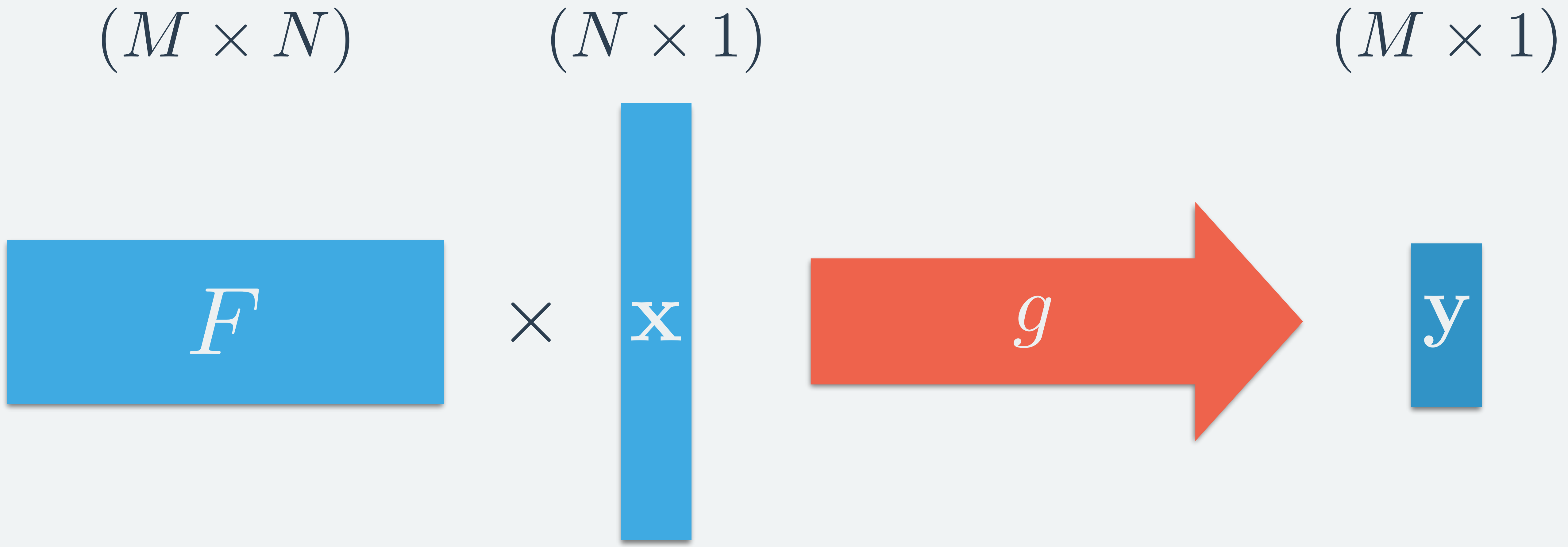
7 July 2015

Eric W. TRAMEL
ENS

Lenka ZDEBOROVÁ
CNRS & CEA-Saclay

Inverse Problems

General Linear Problem: $\mathbf{y} = g\left(F\mathbf{x}\right)$



Projection Matrix

- *iid* Random ?
- Underdetermined?
- Low Rank?
- Sparse?

Signal

Prior Model?

Channel

- Corruption
- Information Loss
- Noise Model?

Measurements

Observed Data

Ex: Compressed Sensing

$$\boxed{\mathbf{y} = F\mathbf{x} + \mathbf{w} \qquad w_\mu \sim \mathcal{N}(0, \Delta)}$$

CS Problem: How do we obtain **x** from **y** and **F** knowing **g = AWGN** & **x** is K-Sparse?

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

(Greedy)

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

(LASSO)

Deterministic

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

(MAP)

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

(MMSE)

Probabilistic

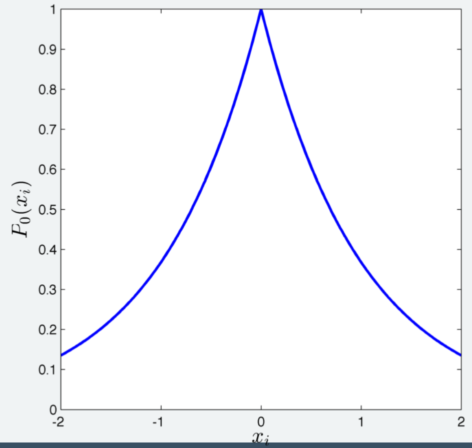
An Unwieldy Posterior

Full Posterior

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

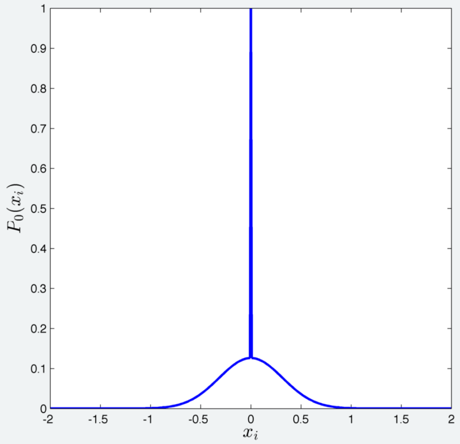
L1/Laplace

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



"Sparse" Bernoulli-Gaussain

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



AWGN Noise Variance

Oh Buddy, That Partition…

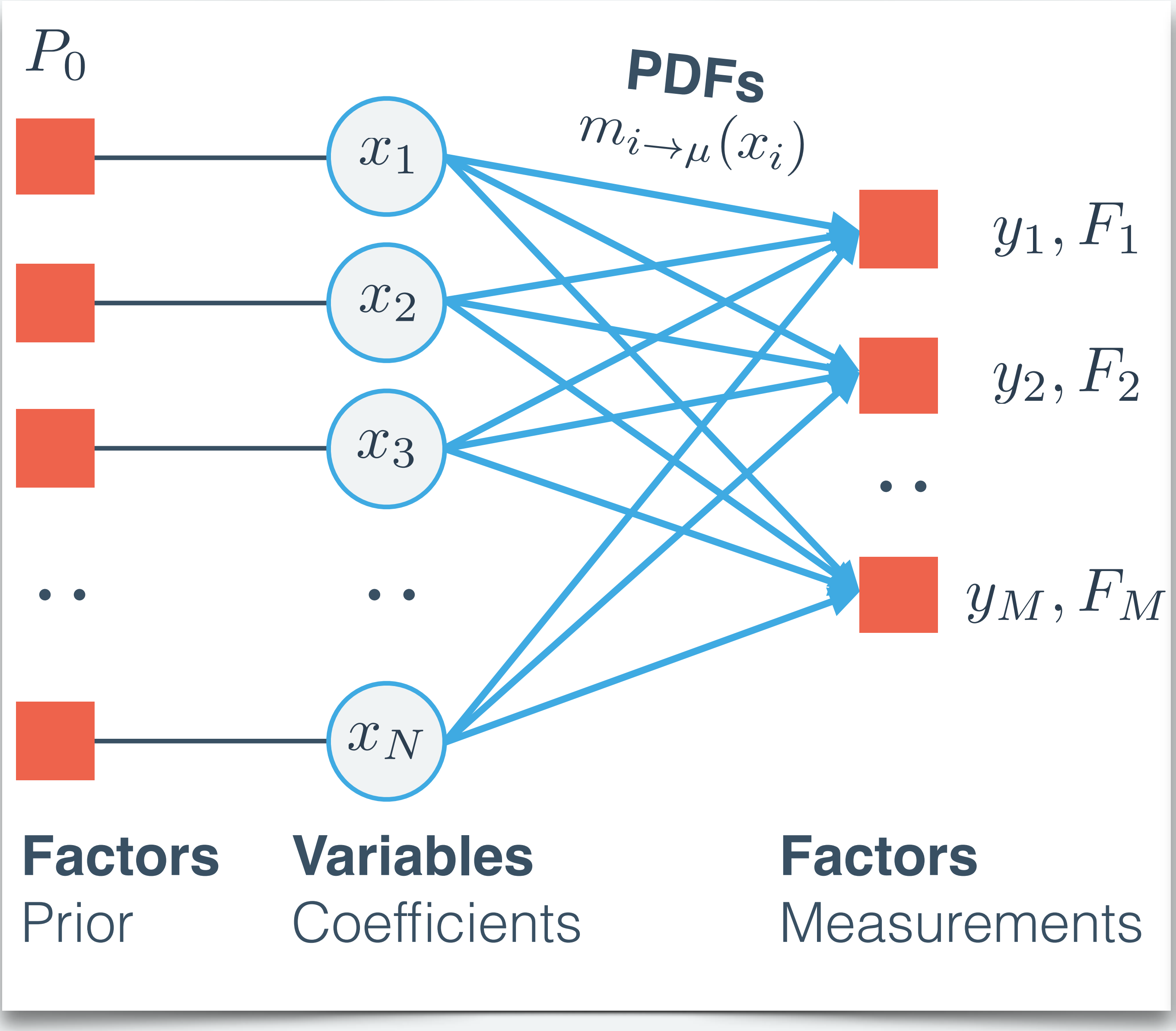
$$Z = \int dx_1 \int dx_2 \ldots \int dx_N \prod_i P_0(x_i)\prod_\mu \frac{1}{\sqrt{2\pi\Delta}}\exp\left\{-\frac{1}{2\Delta}\left(y_\mu - \sum_i F_{\mu i}x_i\right)^2\right\}$$

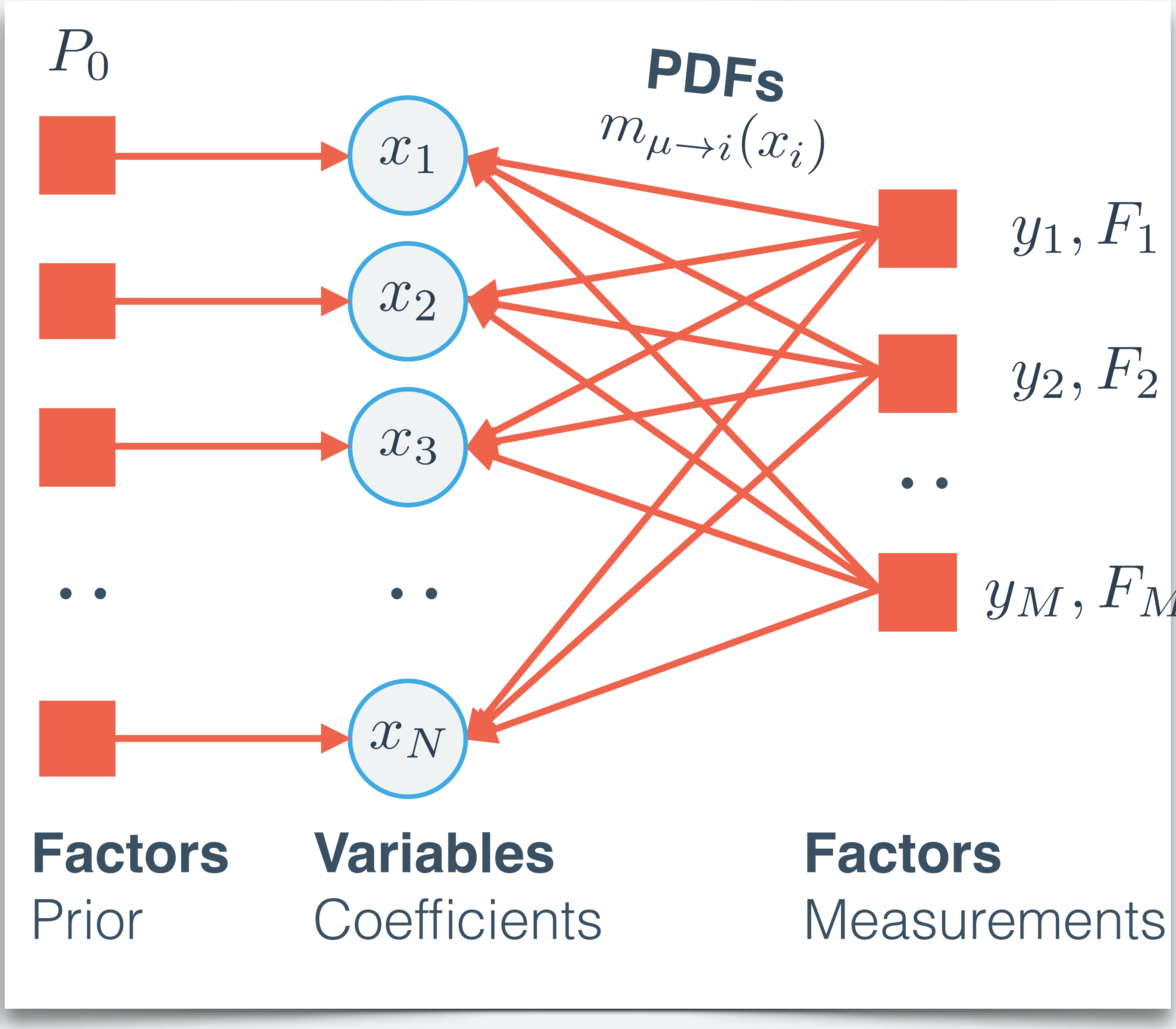
Graphical Model for Factorization

ENS

Variable to Factor Messages

Factor to Variable Messages





Goal: Produce

$$P(x_i) \propto P_0(x_i) \prod m_{\mu\to i}(x_i)$$

Relaxed BP

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

Parallel Edge Iteration

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

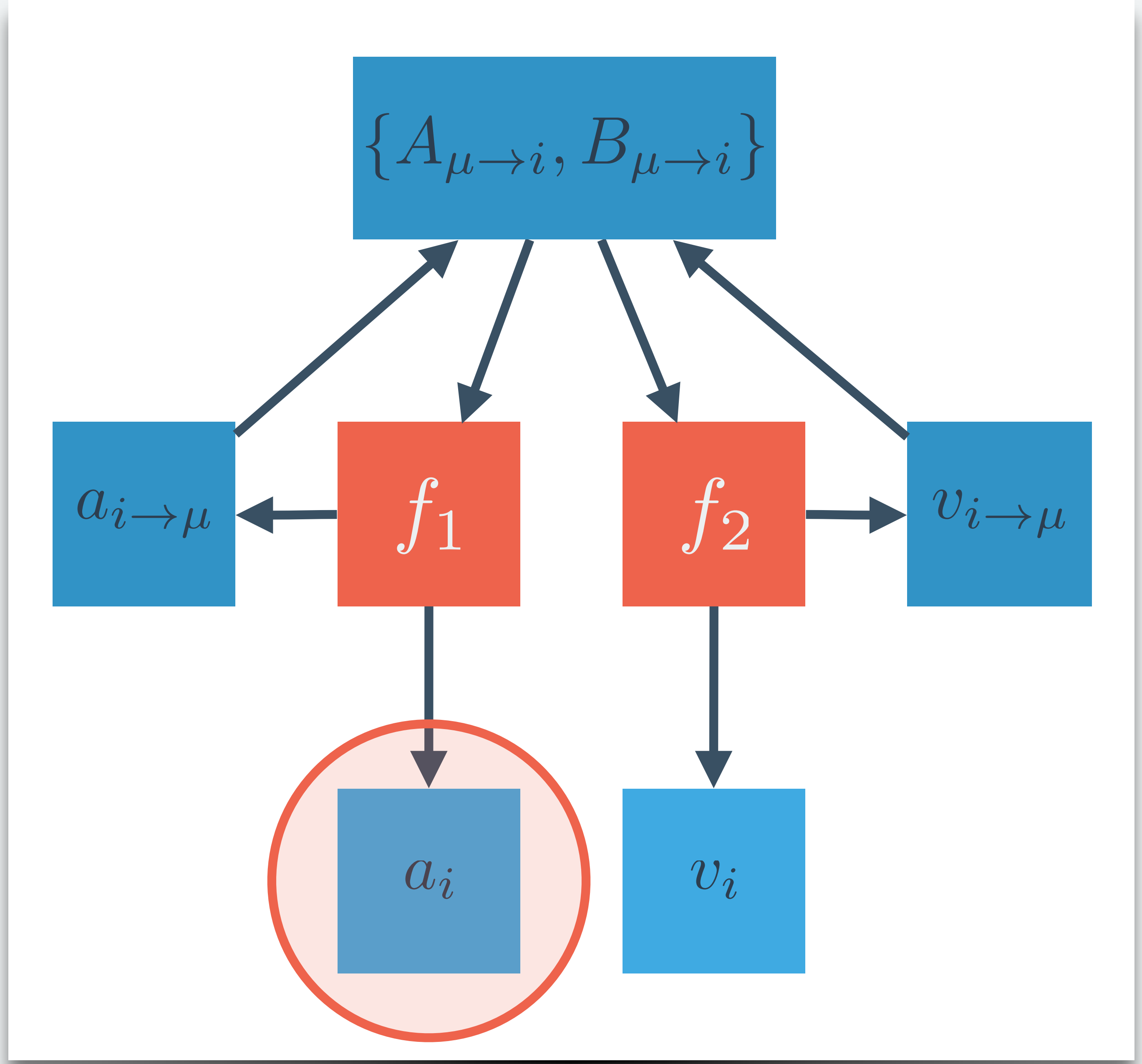
$$B_{\mu\to i} = \frac{F_{\mu i}(y_\mu - \sum_{j\neq i} F_{\mu j} a_{j\to\mu})}{\Delta + \sum_{j\neq i} F_{\mu j}^2 v_{j\to\mu}}$$

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

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

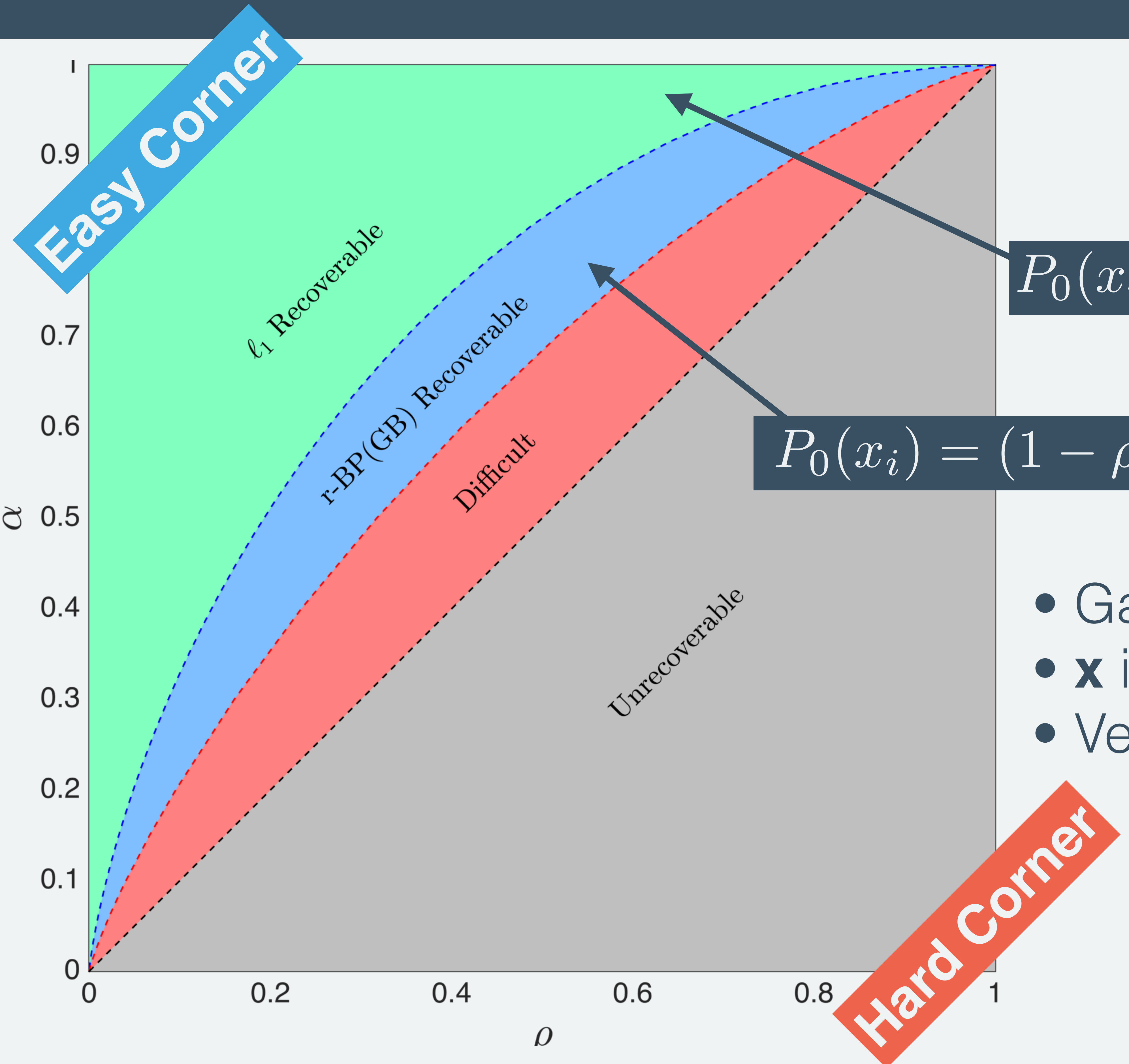
$$a_i = f_1\left(\frac{1}{\sum_{\mu} A_{\mu\to i}}, \frac{\sum_{\mu} B_{\mu\to i}}{\sum_{\mu} A_{\mu\to i}}\right)$$

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



MMSE Signal Reconstruction!

r-BP Transition Performance



Easy Corner

Hard Corner

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

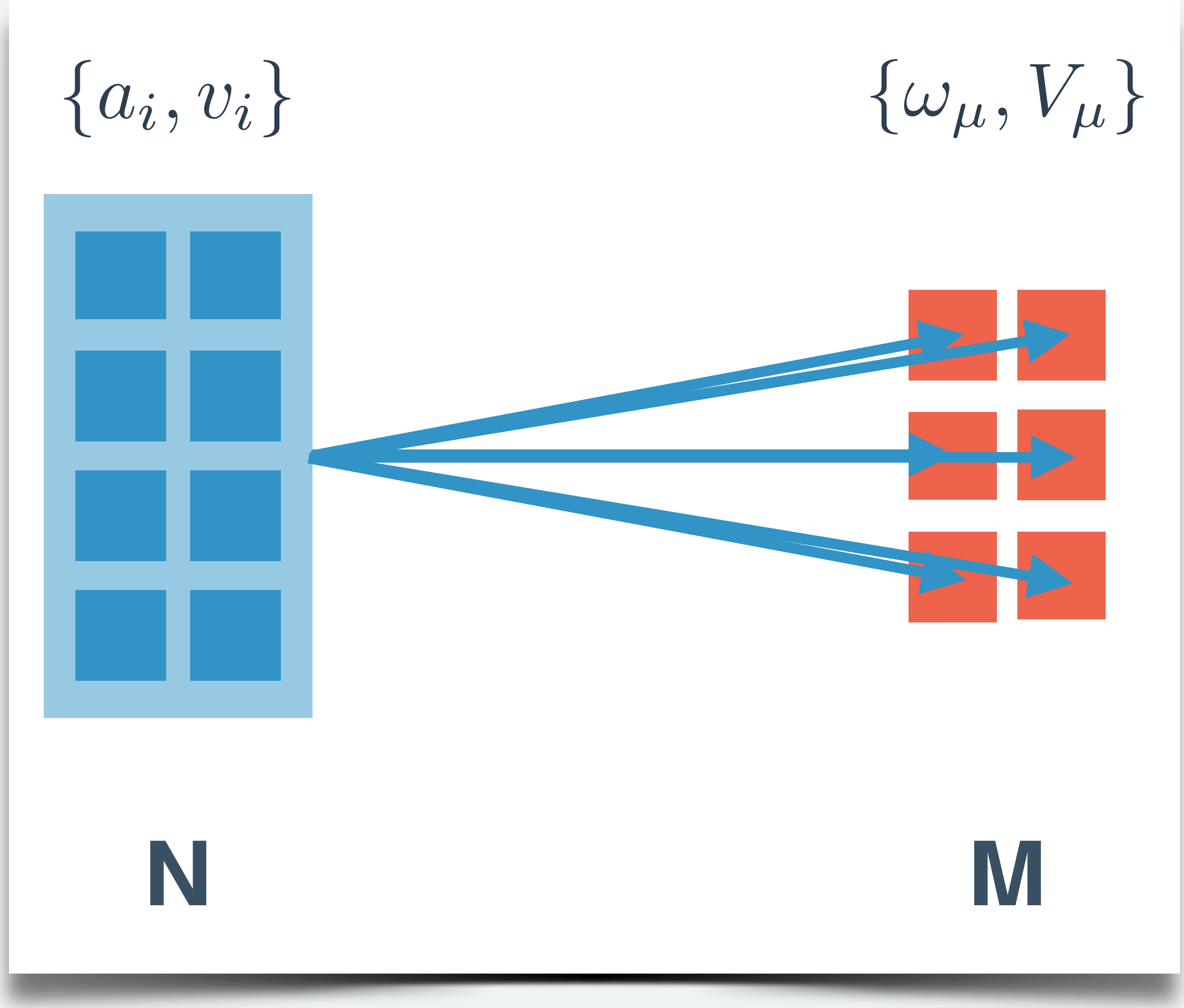
$$P_0(x_i) = (1-\rho)\delta(x_i) + \rho\phi(x_i)$$

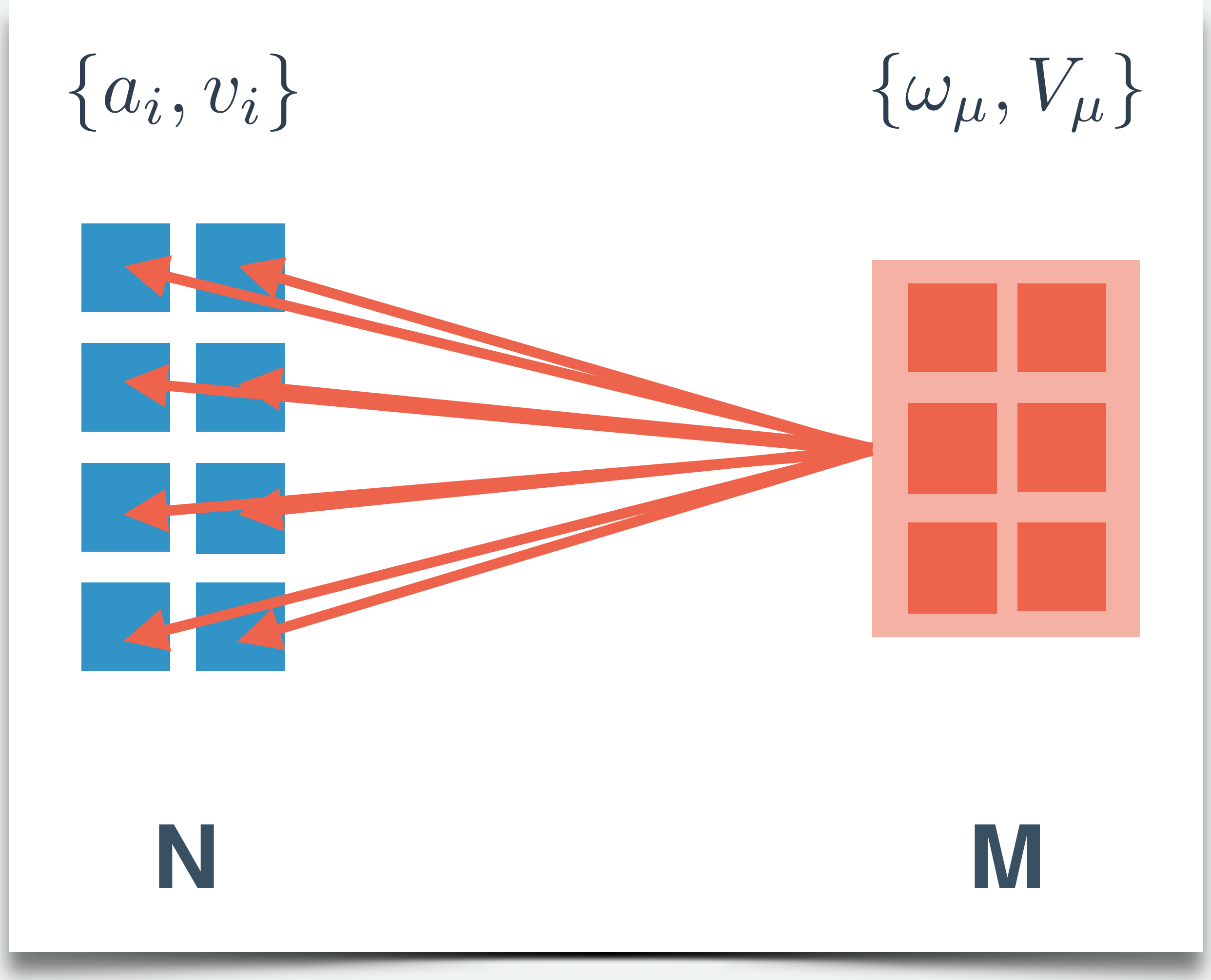
- Gaussian *iid* **F**
- **x** is K-sparse
- Very Low Noise

r-BP to AMP via TAP

TAP Intuition (Extended Mean-Field)

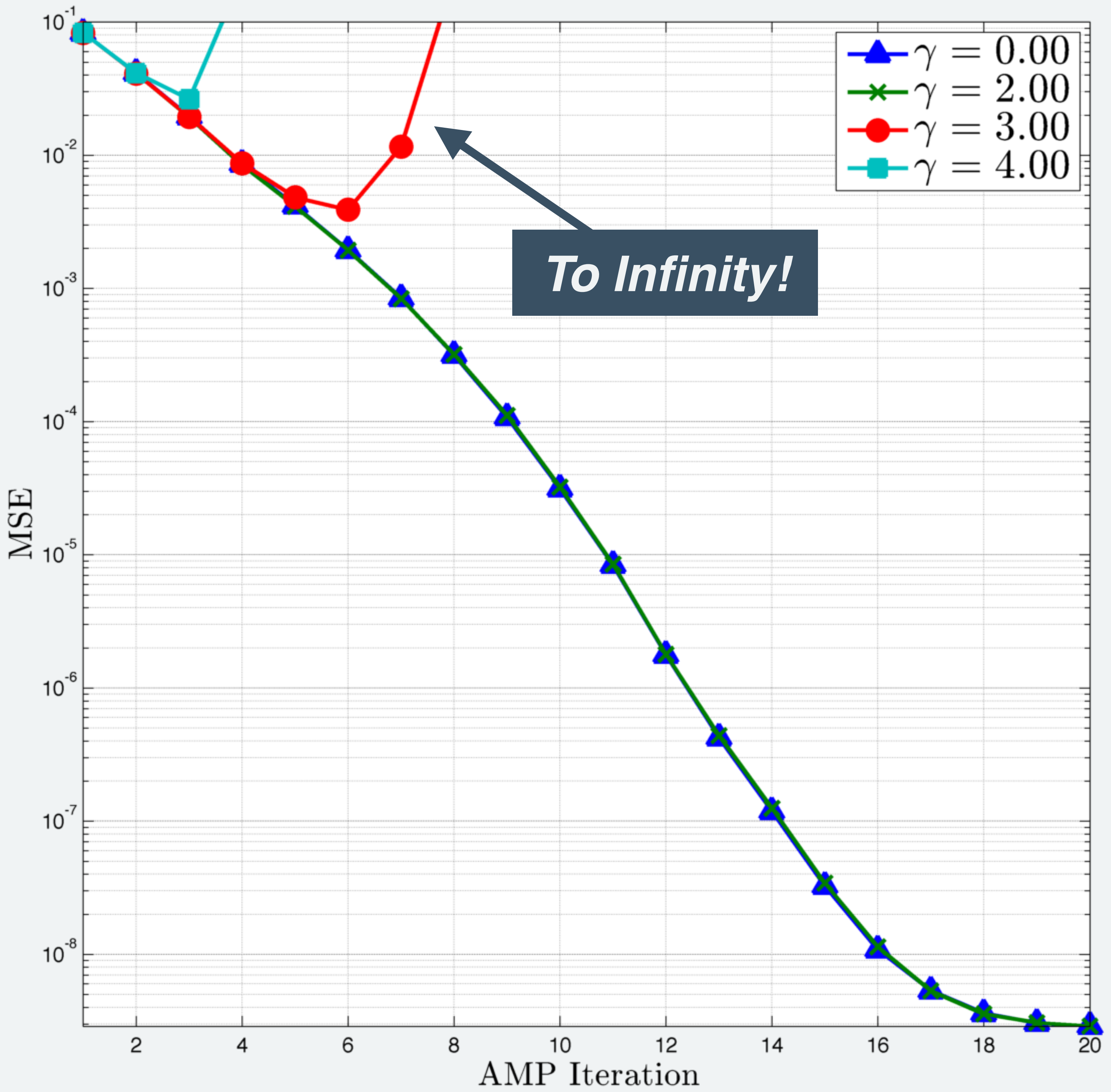
If **F** is ***not sparse*** and if its entries scale ***O(1/√N)***, then message means and variances are ***nearly independent*** of any single edge message in the limit ***N→∞***.





Big Savings: Compute Burden $\mathrm{O}(\alpha N^2) \to \mathrm{O}((1+\alpha)N)$

Breaking AMP & r-BP



To Infinity!

Adding a slight mean

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

$$N = 2048$$

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

$$\alpha = 0.4$$

$$\rho_0 = 0.1$$

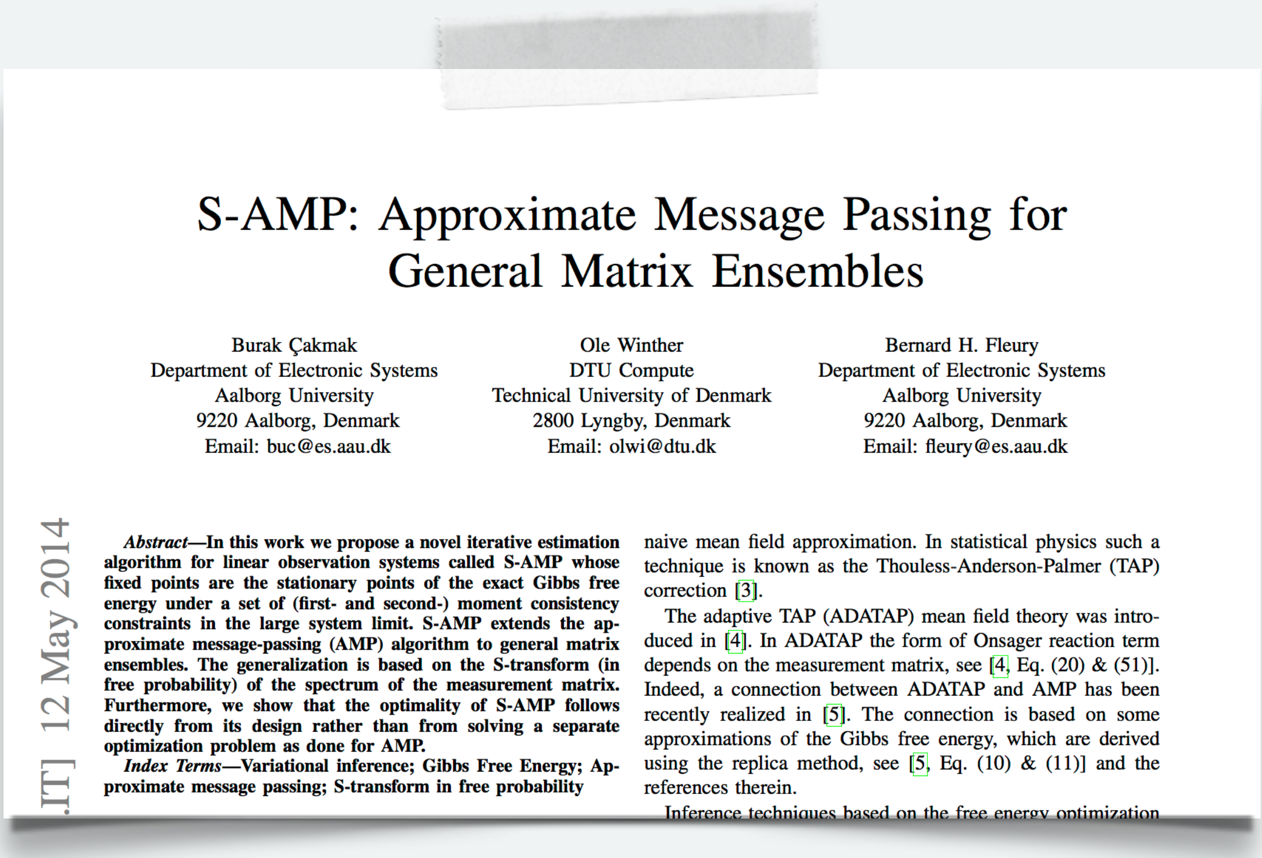
$$\phi \sim \mathcal{N}(0, 1)$$

The Big Obstacle

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

Everything Cool
Sparse Matrices
Low-Rank Matrices
Debluring & Deconvolution
Super-resolution
Feature Selection

Some Approaches…



S-AMP: Approximate Message Passing for General Matrix Ensembles

Burak Çakmak
Department of Electronic Systems
Aalborg University
9220 Aalborg, Denmark
Email: buc@es.aau.dk

Ole Winther
DTU Compute
Technical University of Denmark
2800 Lyngby, Denmark
Email: olwi@dtu.dk

Bernard H. Fleury
Department of Electronic Systems
Aalborg University
9220 Aalborg, Denmark
Email: fleury@es.aau.dk

IT] 12 May 2014

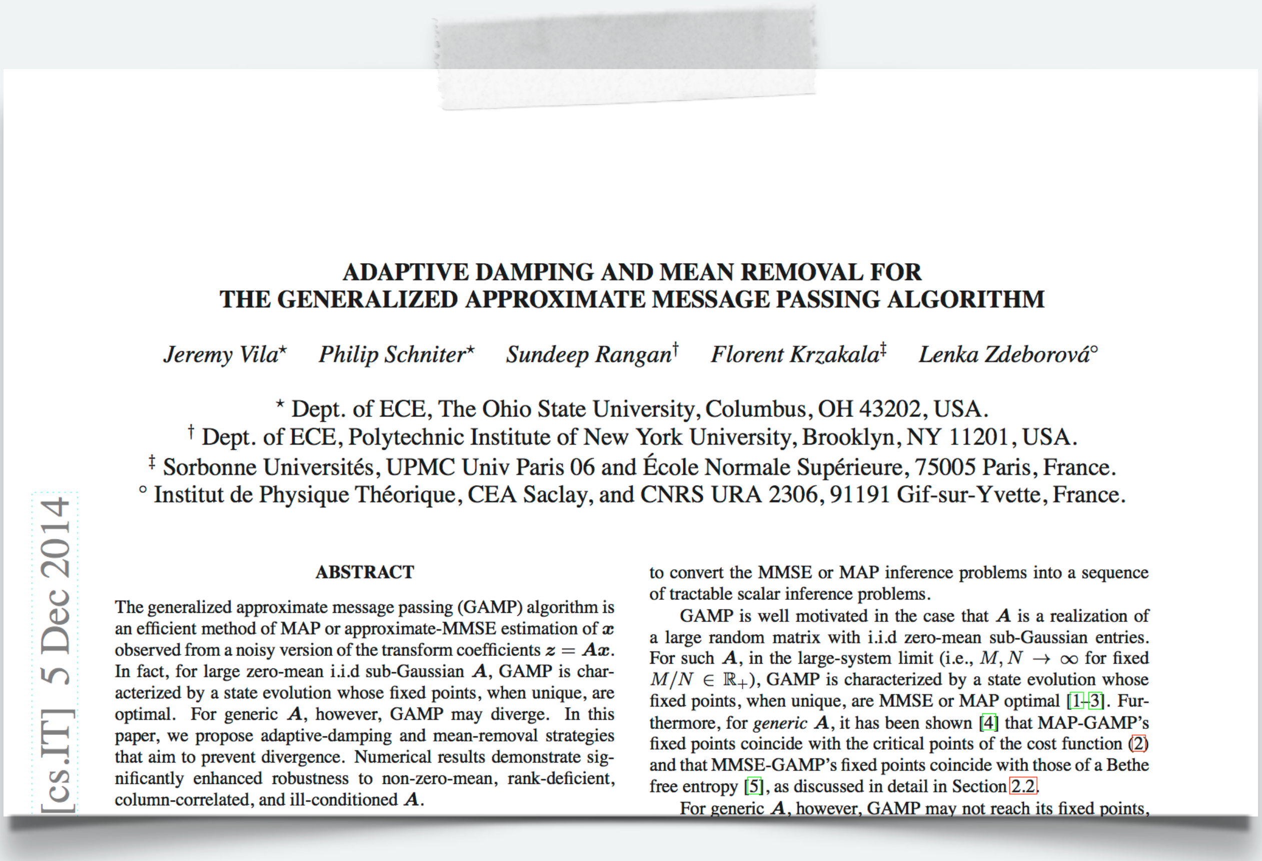
***Abstract*—In this work we propose a novel iterative estimation algorithm for linear observation systems called S-AMP whose fixed points are the stationary points of the exact Gibbs free energy under a set of (first- and second-) moment consistency constraints in the large system limit. S-AMP extends the approximate message-passing (AMP) algorithm to general matrix ensembles. The generalization is based on the S-transform (in free probability) of the spectrum of the measurement matrix. Furthermore, we show that the optimality of S-AMP follows directly from its design rather than from solving a separate optimization problem as done for AMP.**

***Index Terms*—Variational inference; Gibbs Free Energy; Approximate message passing; S-transform in free probability**

naive mean field approximation. In statistical physics such a technique is known as the Thouless-Anderson-Palmer (TAP) correction [3].

The adaptive TAP (ADATAP) mean field theory was introduced in [4]. In ADATAP the form of Onsager reaction term depends on the measurement matrix, see [4, Eq. (20) & (51)]. Indeed, a connection between ADATAP and AMP has been recently realized in [5]. The connection is based on some approximations of the Gibbs free energy, which are derived using the replica method, see [5, Eq. (10) & (11)] and the references therein.

Inference techniques based on the free energy optimization



ADAPTIVE DAMPING AND MEAN REMOVAL FOR THE GENERALIZED APPROXIMATE MESSAGE PASSING ALGORITHM

Jeremy Vila[*] *Philip Schniter*[*] *Sundeep Rangan*[†] *Florent Krzakala*[‡] *Lenka Zdeborová*[°]

[*] Dept. of ECE, The Ohio State University, Columbus, OH 43202, USA.
[†] Dept. of ECE, Polytechnic Institute of New York University, Brooklyn, NY 11201, USA.
[‡] Sorbonne Universités, UPMC Univ Paris 06 and École Normale Supérieure, 75005 Paris, France.
[°] Institut de Physique Théorique, CEA Saclay, and CNRS URA 2306, 91191 Gif-sur-Yvette, France.

[cs.IT] 5 Dec 2014

ABSTRACT

The generalized approximate message passing (GAMP) algorithm is an efficient method of MAP or approximate-MMSE estimation of $\boldsymbol{x}$ observed from a noisy version of the transform coefficients $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$. In fact, for large zero-mean i.i.d sub-Gaussian $\boldsymbol{A}$, GAMP is characterized by a state evolution whose fixed points, when unique, are optimal. For generic $\boldsymbol{A}$, however, GAMP may diverge. In this paper, we propose adaptive-damping and mean-removal strategies that aim to prevent divergence. Numerical results demonstrate significantly enhanced robustness to non-zero-mean, rank-deficient, column-correlated, and ill-conditioned $\boldsymbol{A}$.

to convert the MMSE or MAP inference problems into a sequence of tractable scalar inference problems.

GAMP is well motivated in the case that $\boldsymbol{A}$ is a realization of a large random matrix with i.i.d zero-mean sub-Gaussian entries. For such $\boldsymbol{A}$, in the large-system limit (i.e., $M, N \to \infty$ for fixed $M/N \in \mathbb{R}_+$), GAMP is characterized by a state evolution whose fixed points, when unique, are MMSE or MAP optimal [1–3]. Furthermore, for *generic* $\boldsymbol{A}$, it has been shown [4] that MAP-GAMP's fixed points coincide with the critical points of the cost function (2) and that MMSE-GAMP's fixed points coincide with those of a Bethe free entropy [5], as discussed in detail in Section 2.2.

For generic $\boldsymbol{A}$, however, GAMP may not reach its fixed points,

F as a random matrix ensemble in Free Probability.

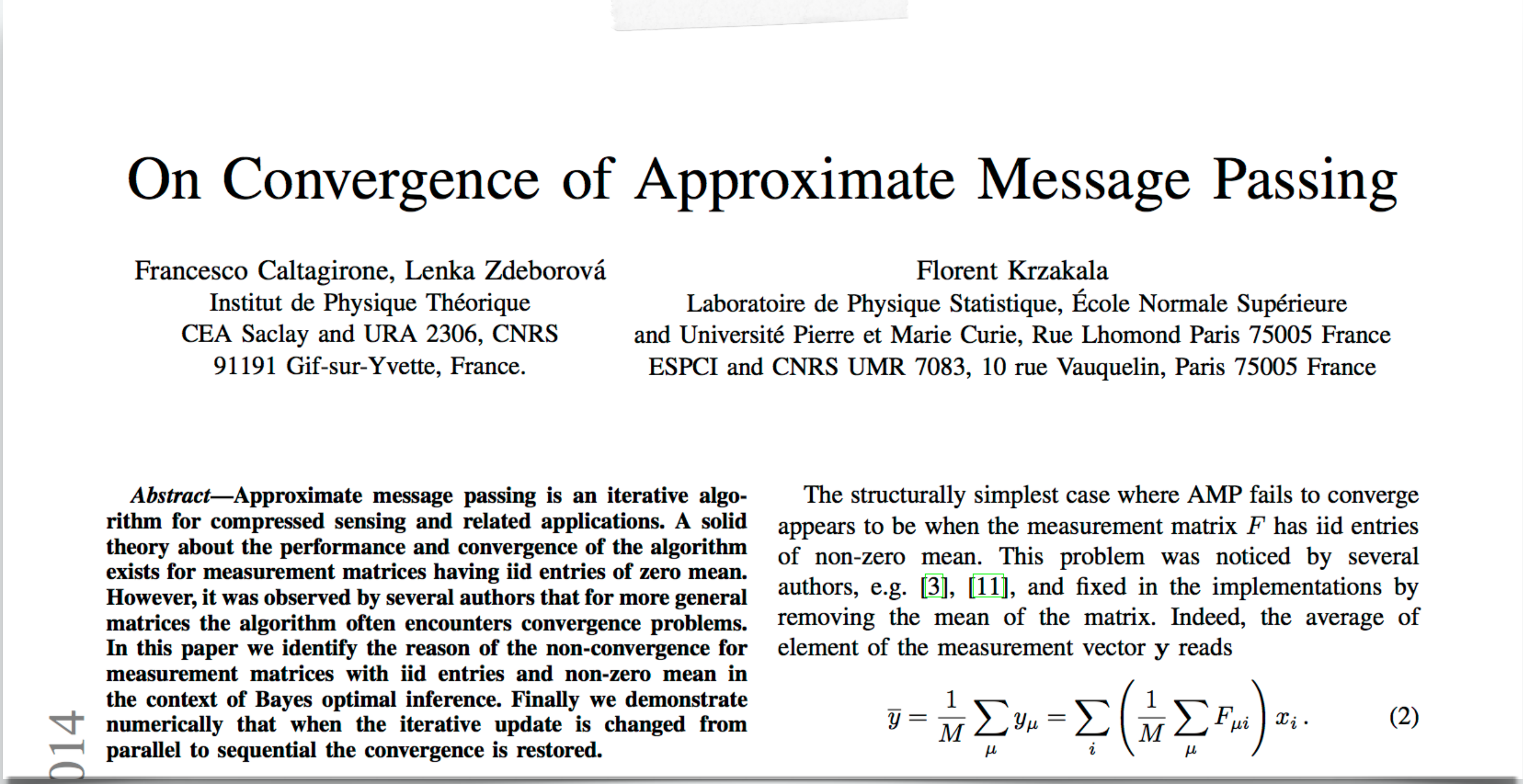
➡ Restricted to random matrices.

Slow FPI, force minimization of Bethe Free Energy*.

➡ Requires cost calculation (non-trivial) at each iteration.

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

Sequential r-BP Update



On Convergence of Approximate Message Passing

Francesco Caltagirone, Lenka Zdeborová
Institut de Physique Théorique
CEA Saclay and URA 2306, CNRS
91191 Gif-sur-Yvette, France.

Florent Krzakala
Laboratoire de Physique Statistique, École Normale Supérieure
and Université Pierre et Marie Curie, Rue Lhomond Paris 75005 France
ESPCI and CNRS UMR 7083, 10 rue Vauquelin, Paris 75005 France

***Abstract*—Approximate message passing is an iterative algorithm for compressed sensing and related applications. A solid theory about the performance and convergence of the algorithm exists for measurement matrices having iid entries of zero mean. However, it was observed by several authors that for more general matrices the algorithm often encounters convergence problems. In this paper we identify the reason of the non-convergence for measurement matrices with iid entries and non-zero mean in the context of Bayes optimal inference. Finally we demonstrate numerically that when the iterative update is changed from parallel to sequential the convergence is restored.**

The structurally simplest case where AMP fails to converge appears to be when the measurement matrix F has iid entries of non-zero mean. This problem was noticed by several authors, e.g. [3], [11], and fixed in the implementations by removing the mean of the matrix. Indeed, the average of element of the measurement vector $\mathbf{y}$ reads

$$\overline{y} = \frac{1}{M}\sum_{\mu} y_\mu = \sum_i \left(\frac{1}{M}\sum_{\mu} F_{\mu i}\right) x_i\,. \qquad (2)$$

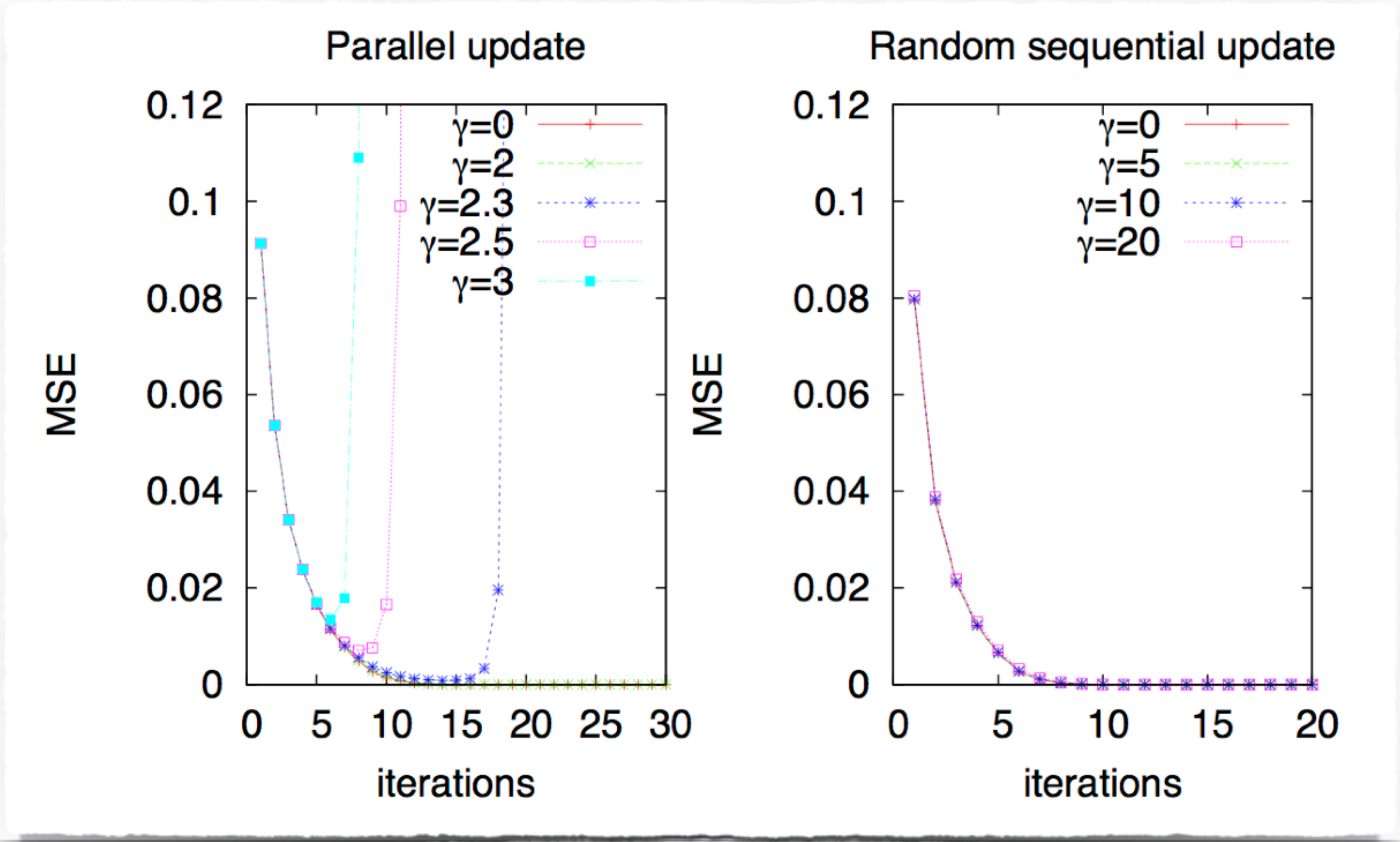
$$F_{\mu,i} \sim \mathcal{N}\left(\frac{\gamma}{N}, \frac{1}{N}\right)$$

Why does parallel update fail?

CZK2014: Solution path can absorb **small** amounts of instability in the parallel update…*but too much prevents convergence!*

Sequential r-BP Update

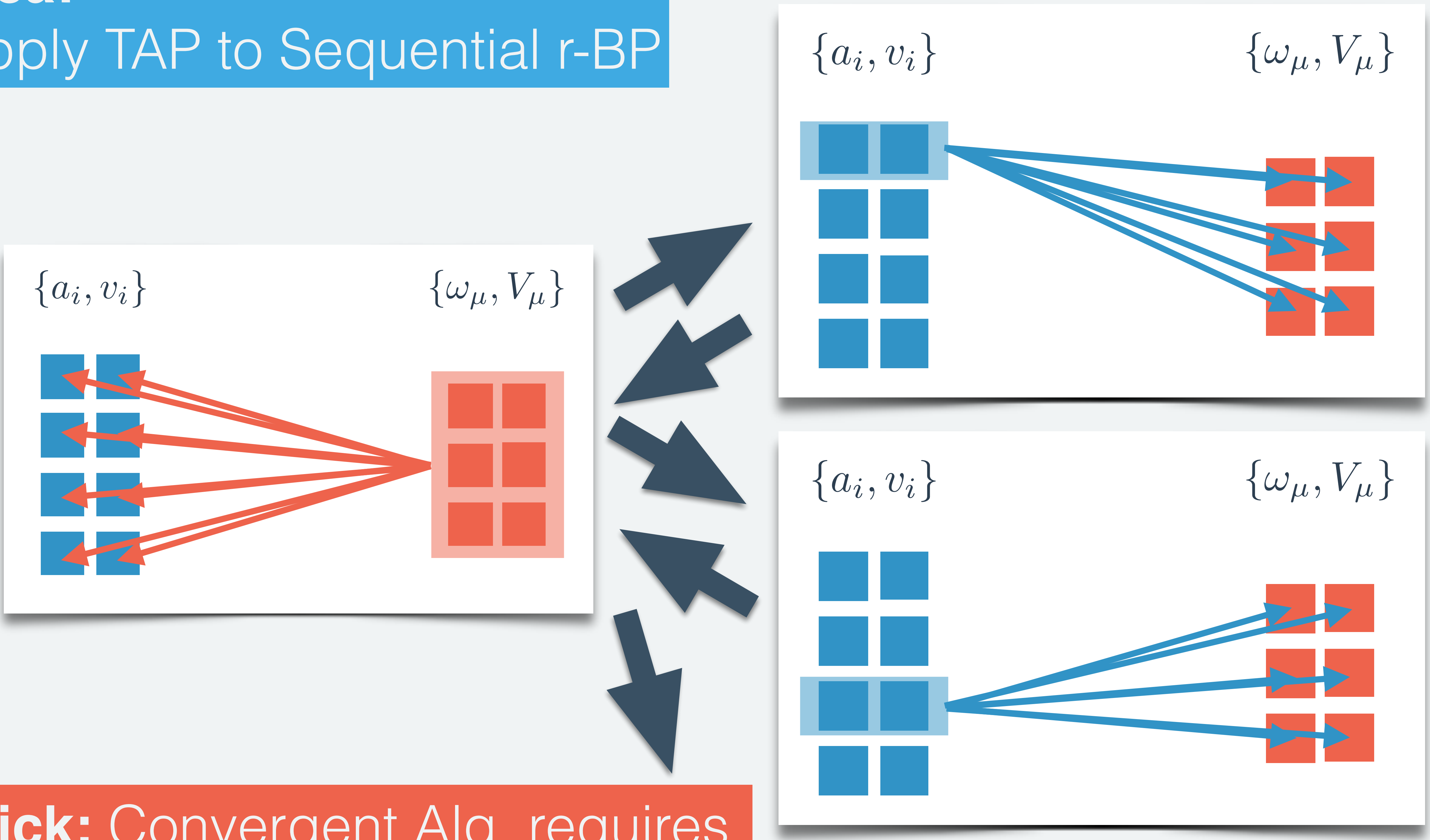
Updating one random edge at a time…



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

Our Proposal: Swept AMP (SwAMP)

Idea:
Apply TAP to Sequential r-BP



Trick: Convergent Alg. requires re-derivation of time indices!

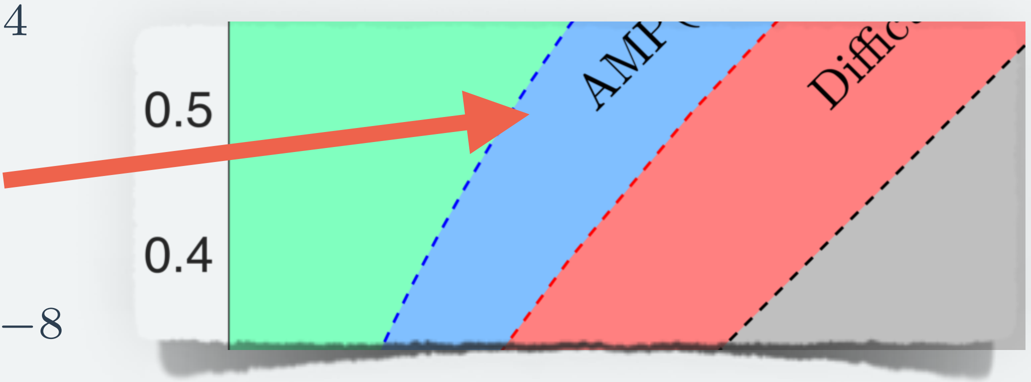
Case I: Non-zero Mean

$N = 10^4$

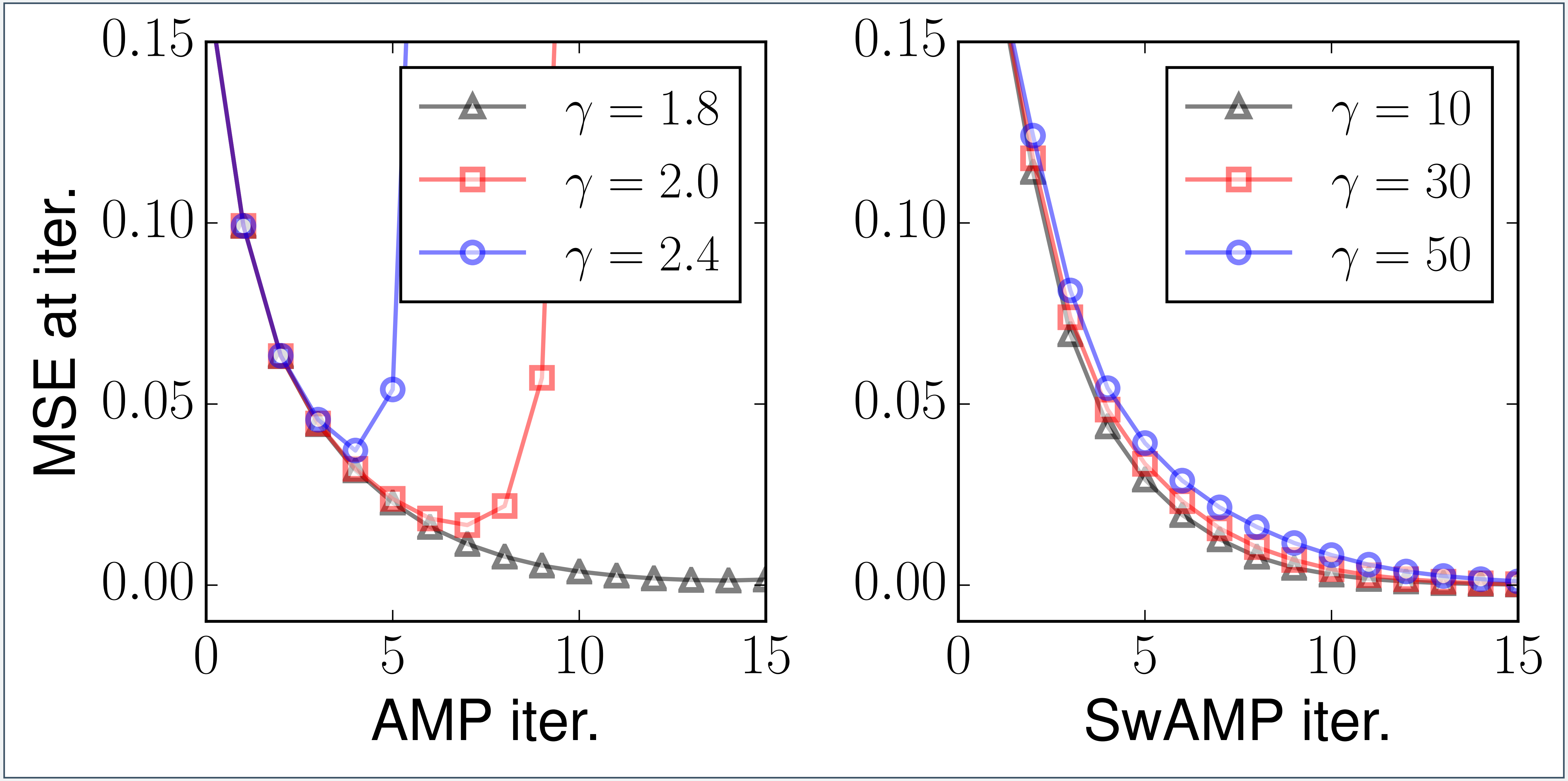
$\rho = 0.2$

$\alpha = 0.5$

$\Delta = 10^{-8}$



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



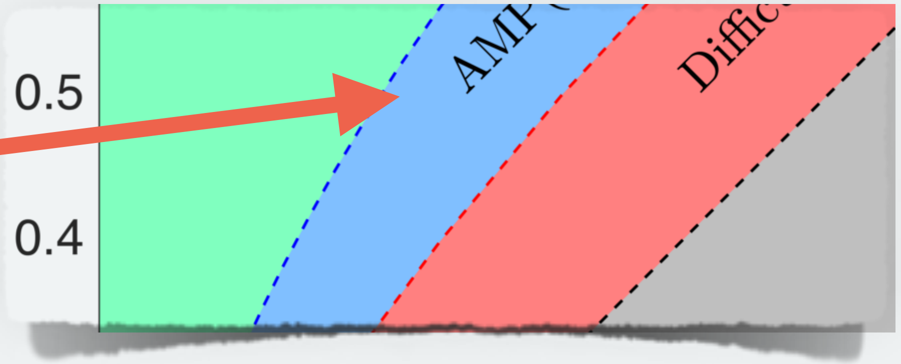
Case I: Non-zero Mean

$N = 10^4$

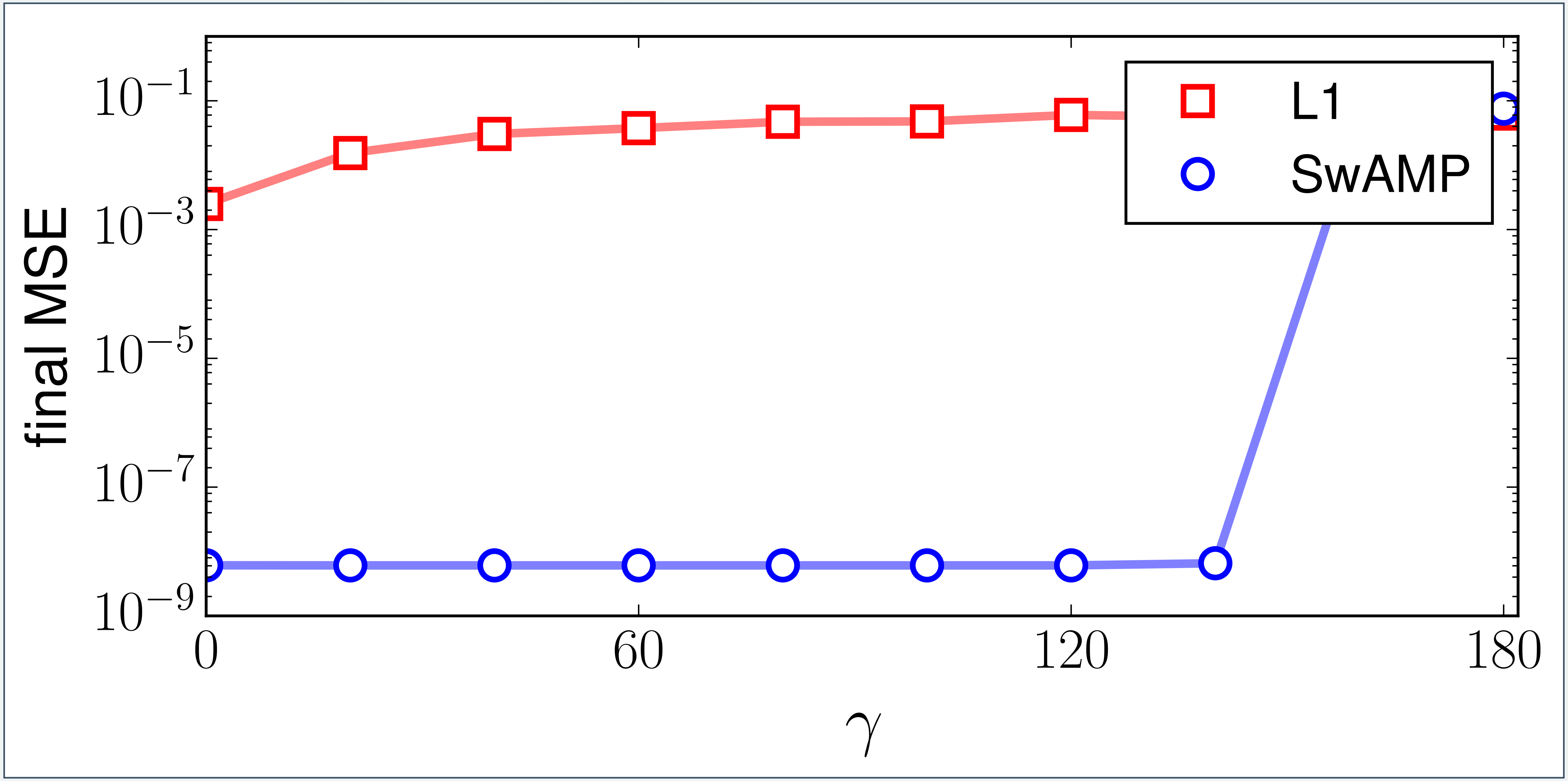
$\rho = 0.2$

$\alpha = 0.5$

$\Delta = 10^{-8}$



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



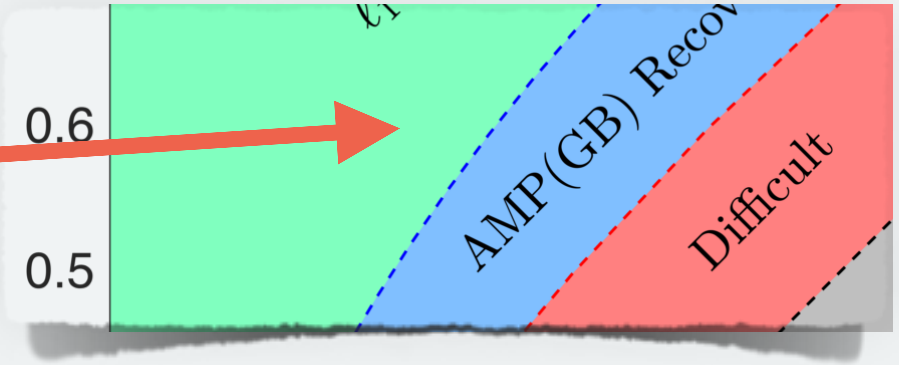
Case II: Correlations & Low Rank

$N = 10^3$

$\rho = 0.2$

$\alpha = 0.6$

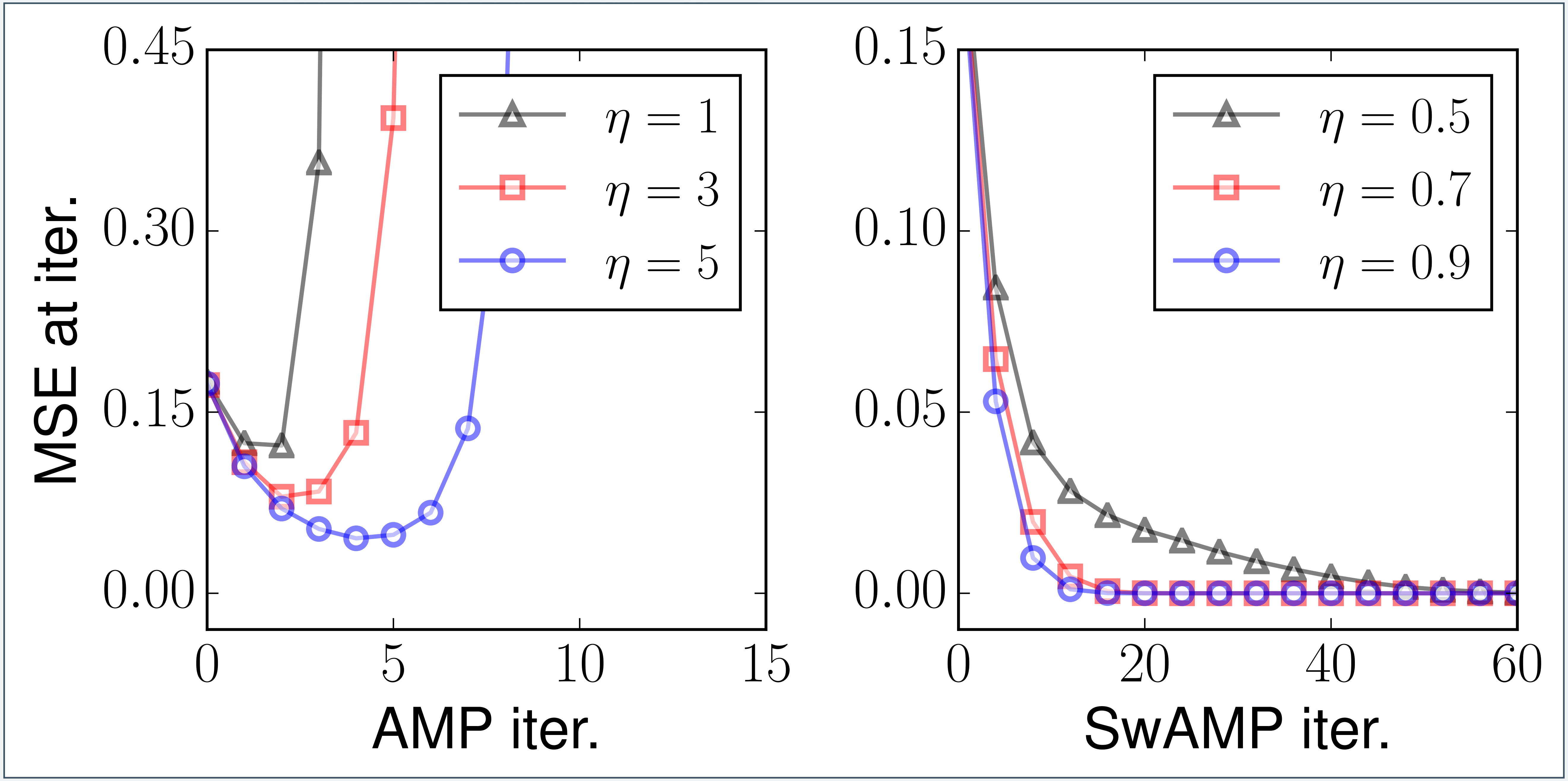
$\Delta = 10^{-8}$



$$F = \frac{1}{N} PQ$$

$$P_{\mu k}, Q_{ki} \sim \mathcal{N}(0,1)$$

$$P : M \times R, \quad Q : R \times N, \quad R \triangleq \eta N$$



Case II: Correlations & Low Rank

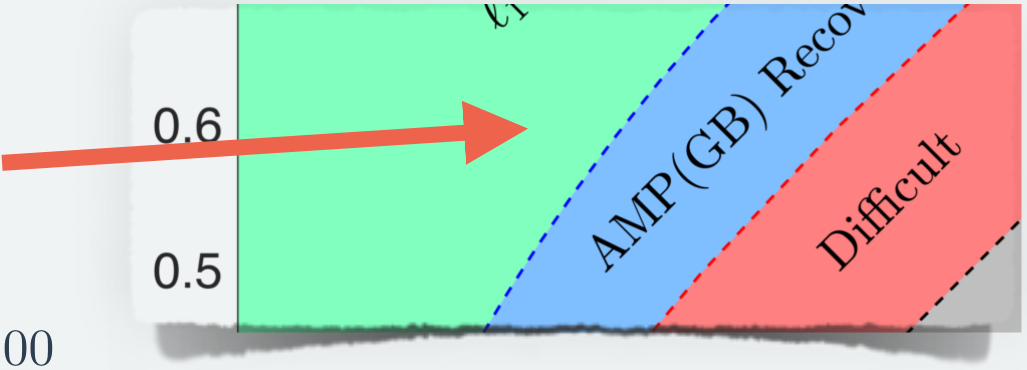
$N = 10^3$

$\rho = 0.2$

$\alpha = 0.6$

$\Delta = 10^{-8}$

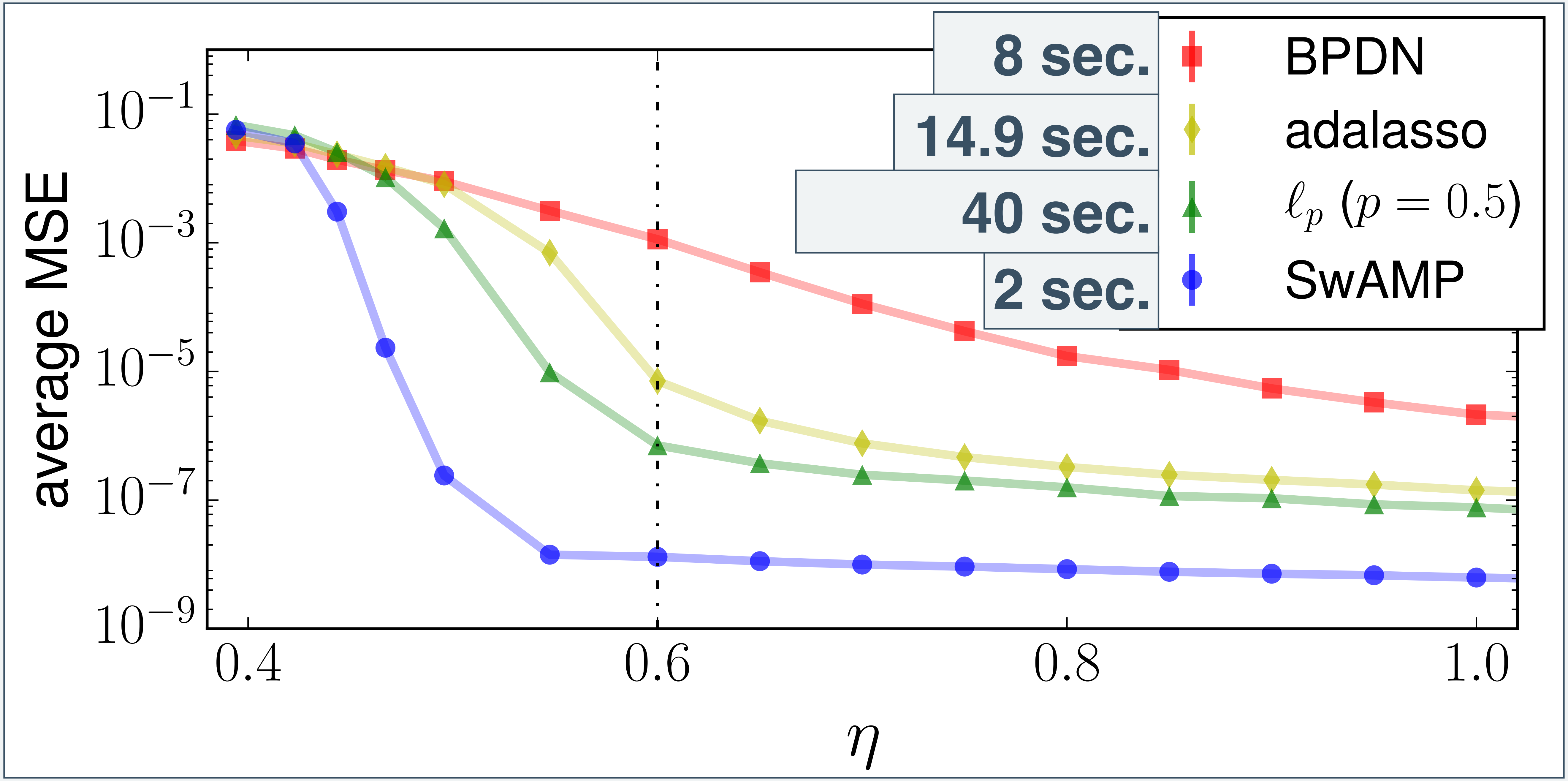
Trials $= 500$



$$F = \frac{1}{N} PQ$$

$$P_{\mu k}, Q_{ki} \sim \mathcal{N}(0,1)$$

$$P : M \times R, \quad Q : R \times N, \quad R \triangleq \eta N$$



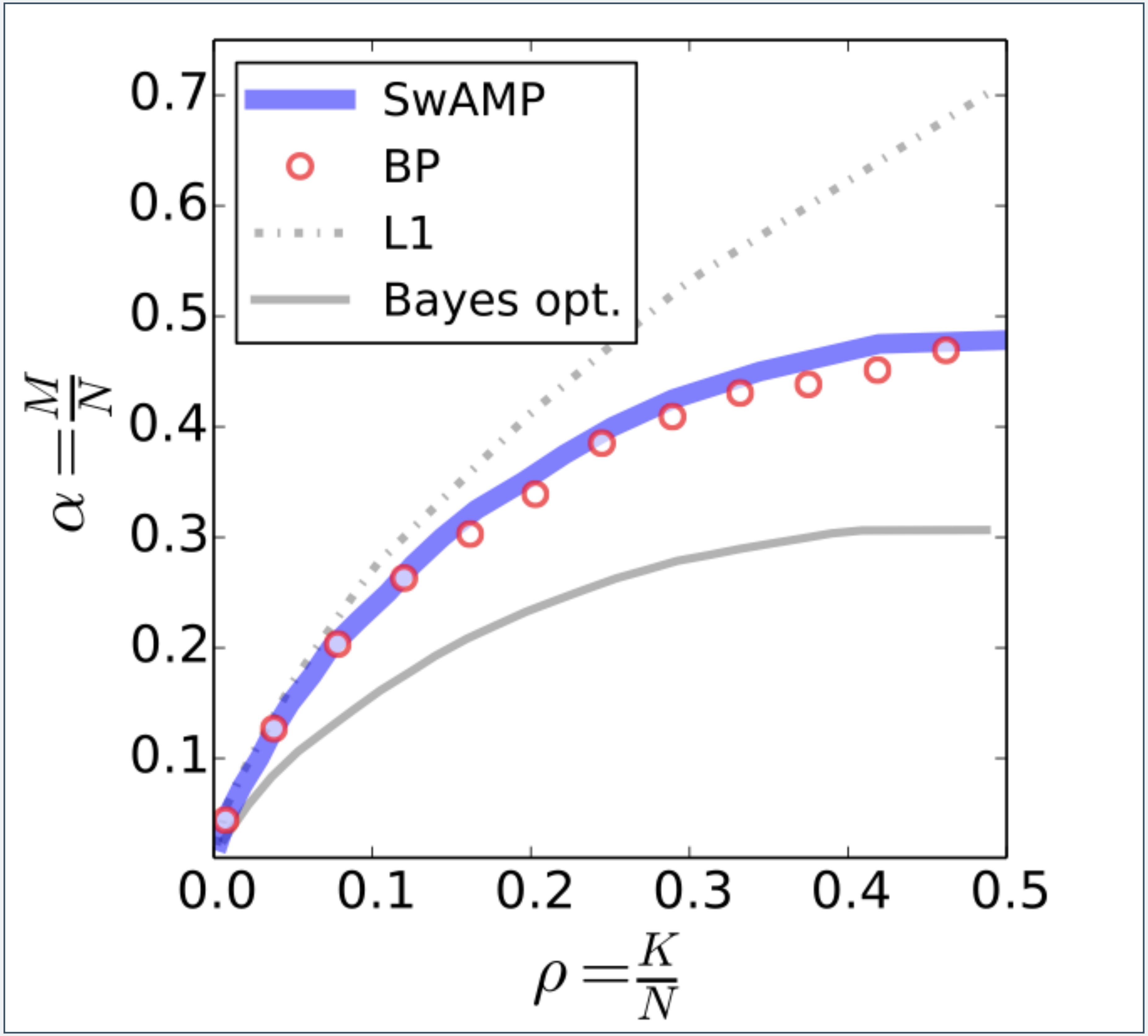
8 sec.

14.9 sec.

40 sec.

2 sec.

Case III: Group Testing



Sparse Matrices

Works for group testing, too!

$$F_{\mu i} \in \{0,1\}$$

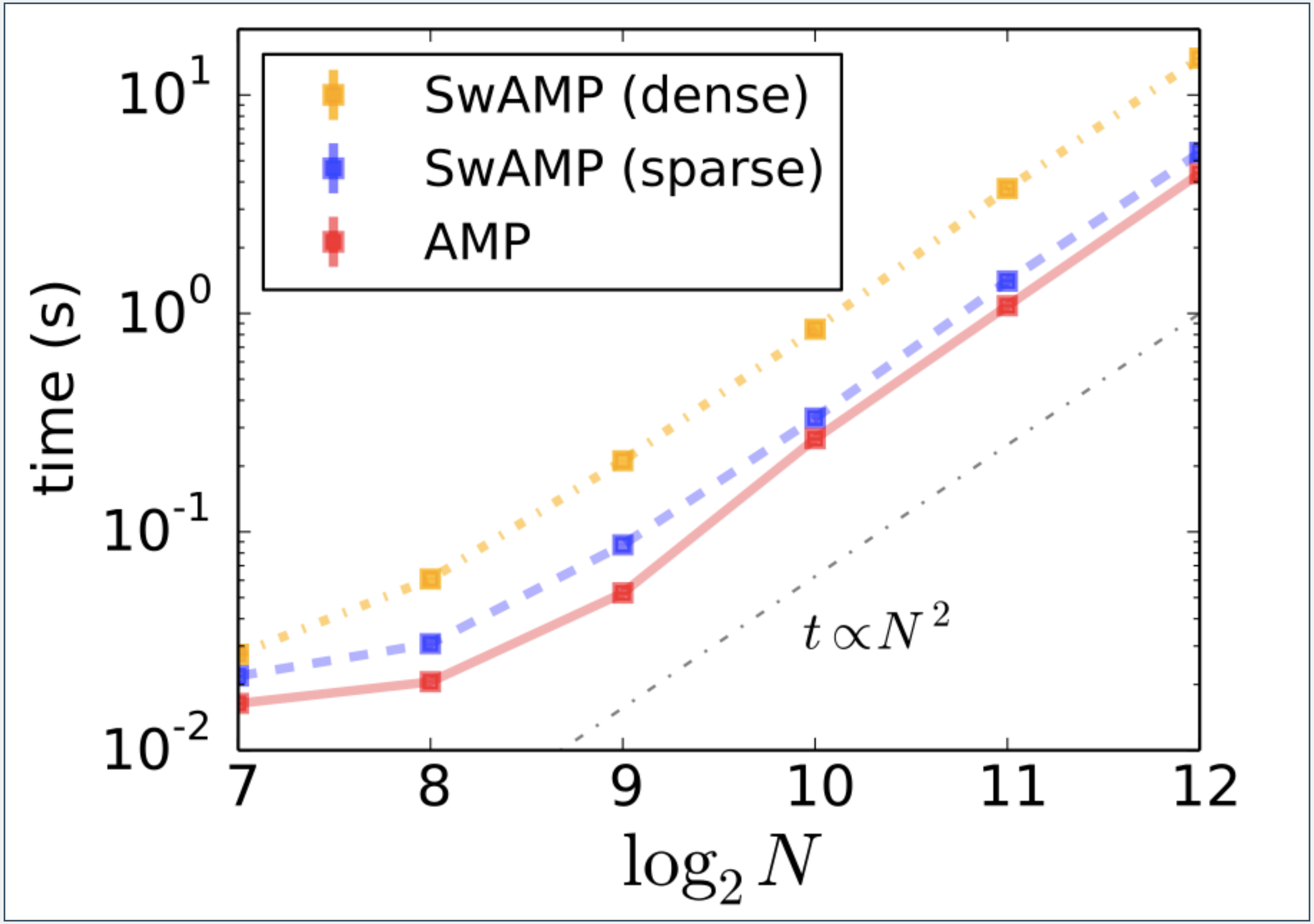
$$\sum_i F_{\mu i} = 7$$

$$x_i \in \{0,1\}$$

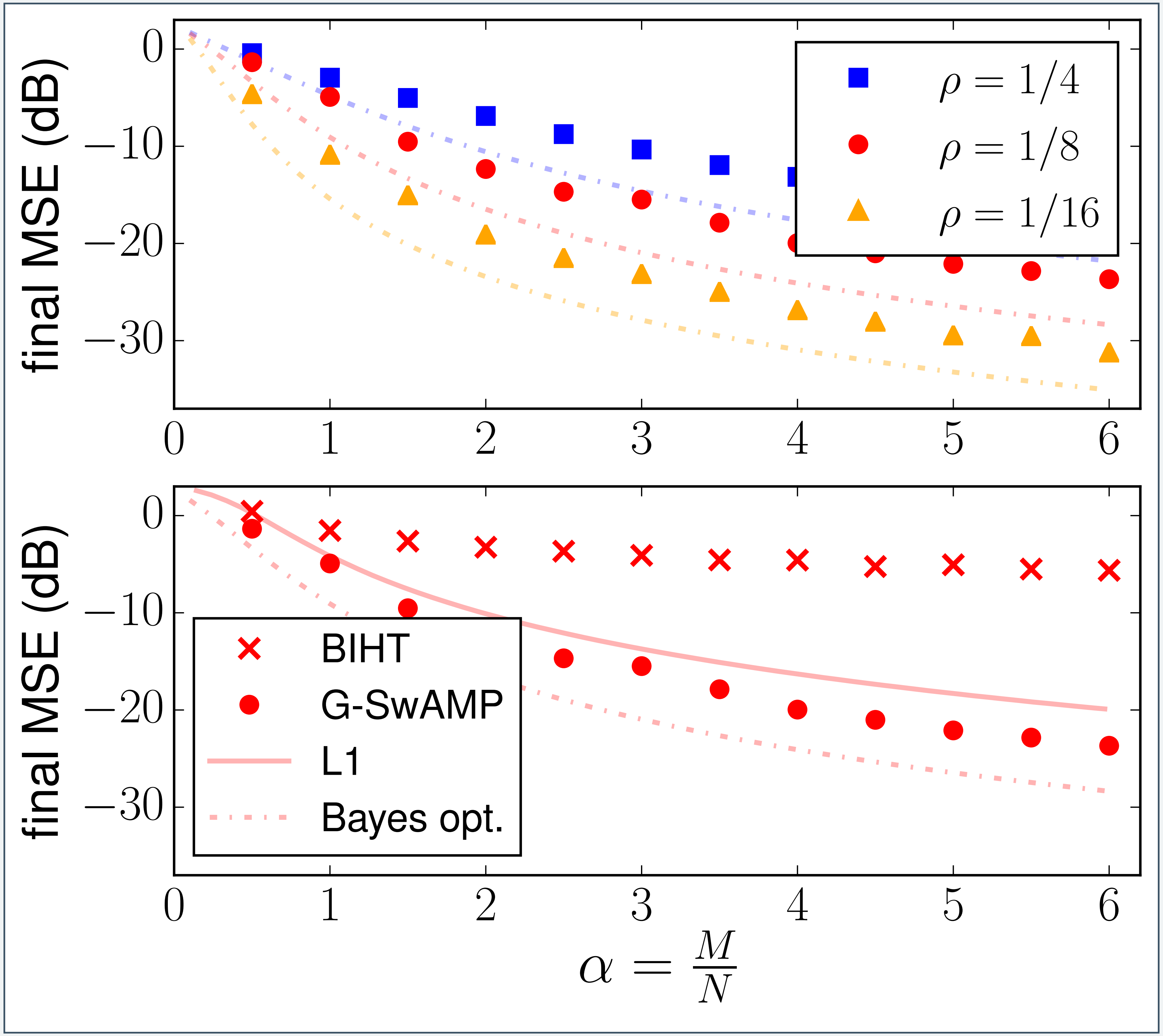
$$\phi \sim \delta(x-1)$$

Case III: Group Testing

Performance Impact: *not so bad!*



Case IV: One-bit CS



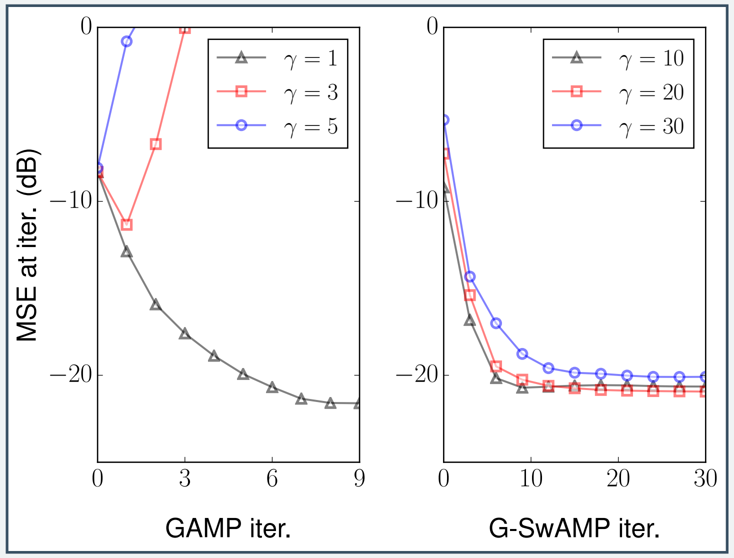
$$\boxed{\mathbf{y} = \mathrm{sign}(F\mathbf{x})}$$

$$F \sim \mathcal{N}\left(\frac{\gamma}{N}, \frac{1}{N}\right)$$

$$N = 512$$

$$\gamma = 20$$

$$\text{Trials} = 200$$



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**?
- ◉ Is parallel FP-iteration doomed for wide class of problems w/o fundamental changes (i.e. S-AMP) ?

SPHINX @ENS

Statistical **PH**ysics of **IN**formaiton e**X**traction

«OU»

Statistical **PH**ysics of **IN**verse comple**X** sysems

ENS

1794

Questions?

Thanks!

Available Online ! Try it out !

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

Sum-Product AMP Algorithm

In Its Totality…

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

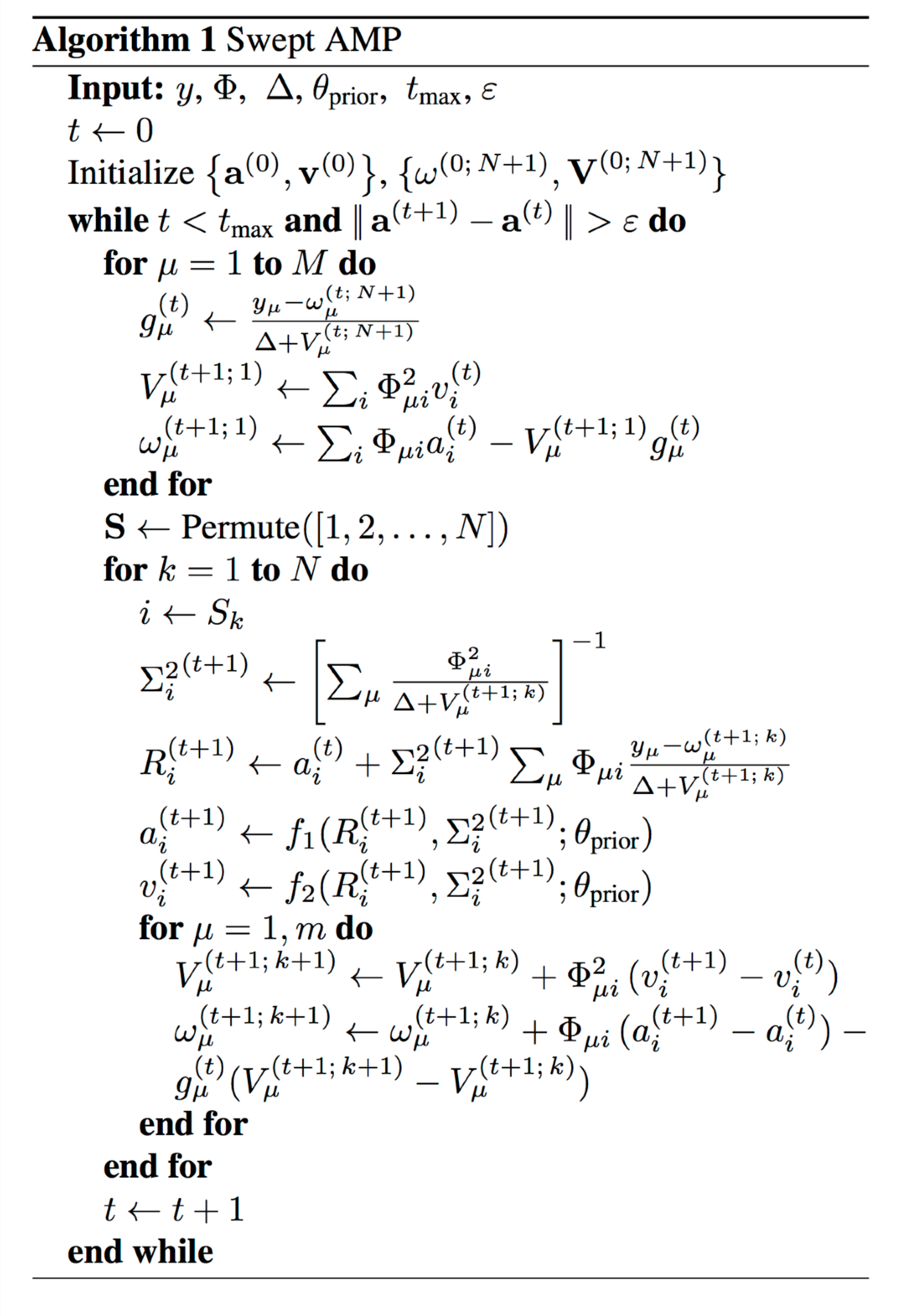
$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

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

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

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

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



Algorithm 1 Swept AMP

```
Input: $y, \Phi, \Delta, \theta_{\text{prior}}, t_{\max}, \varepsilon$
$t \leftarrow 0$
Initialize $\{\mathbf{a}^{(0)}, \mathbf{v}^{(0)}\}$, $\{\omega^{(0;\,N+1)}, \mathbf{V}^{(0;\,N+1)}\}$
while $t < t_{\max}$ and $\| \mathbf{a}^{(t+1)} - \mathbf{a}^{(t)} \| > \varepsilon$ do
    for $\mu = 1$ to $M$ do
        $g_\mu^{(t)} \leftarrow \frac{y_\mu - \omega_\mu^{(t;\,N+1)}}{\Delta + V_\mu^{(t;\,N+1)}}$
        $V_\mu^{(t+1;\,1)} \leftarrow \sum_i \Phi_{\mu i}^2 v_i^{(t)}$
        $\omega_\mu^{(t+1;\,1)} \leftarrow \sum_i \Phi_{\mu i} a_i^{(t)} - V_\mu^{(t+1;\,1)} g_\mu^{(t)}$
    end for
    $\mathbf{S} \leftarrow \text{Permute}([1, 2, \ldots, N])$
    for $k = 1$ to $N$ do
        $i \leftarrow S_k$
        $\Sigma_i^{2\,(t+1)} \leftarrow \left[ \sum_\mu \frac{\Phi_{\mu i}^2}{\Delta + V_\mu^{(t+1;\,k)}} \right]^{-1}$
        $R_i^{(t+1)} \leftarrow a_i^{(t)} + \Sigma_i^{2\,(t+1)} \sum_\mu \Phi_{\mu i} \frac{y_\mu - \omega_\mu^{(t+1;\,k)}}{\Delta + V_\mu^{(t+1;\,k)}}$
        $a_i^{(t+1)} \leftarrow f_1(R_i^{(t+1)}, \Sigma_i^{2\,(t+1)}; \theta_{\text{prior}})$
        $v_i^{(t+1)} \leftarrow f_2(R_i^{(t+1)}, \Sigma_i^{2\,(t+1)}; \theta_{\text{prior}})$
        for $\mu = 1, m$ do
            $V_\mu^{(t+1;\,k+1)} \leftarrow V_\mu^{(t+1;\,k)} + \Phi_{\mu i}^2 \left(v_i^{(t+1)} - v_i^{(t)}\right)$
            $\omega_\mu^{(t+1;\,k+1)} \leftarrow \omega_\mu^{(t+1;\,k)} + \Phi_{\mu i} \left(a_i^{(t+1)} - a_i^{(t)}\right) -$
            $g_\mu^{(t)} (V_\mu^{(t+1;\,k+1)} - V_\mu^{(t+1;\,k)})$
        end for
    end for
    $t \leftarrow t + 1$
end while
```