Statistical Mechanical Approach to CDMA Multiuser Detection Algorithm

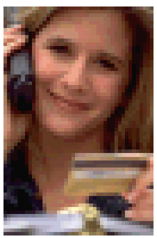
Yoshiyuki Kabashima
Dept. of Compt. Intel. & Syst. Sci.
Tokyo Institute of Technology
Email: kaba [at mark] dis.titech.ac.jp
http://www.sp.dis.titech.ac.jp/~kaba/index_e.html
http://www.smapip.eei.metro-u.ac.jp/e/index.html

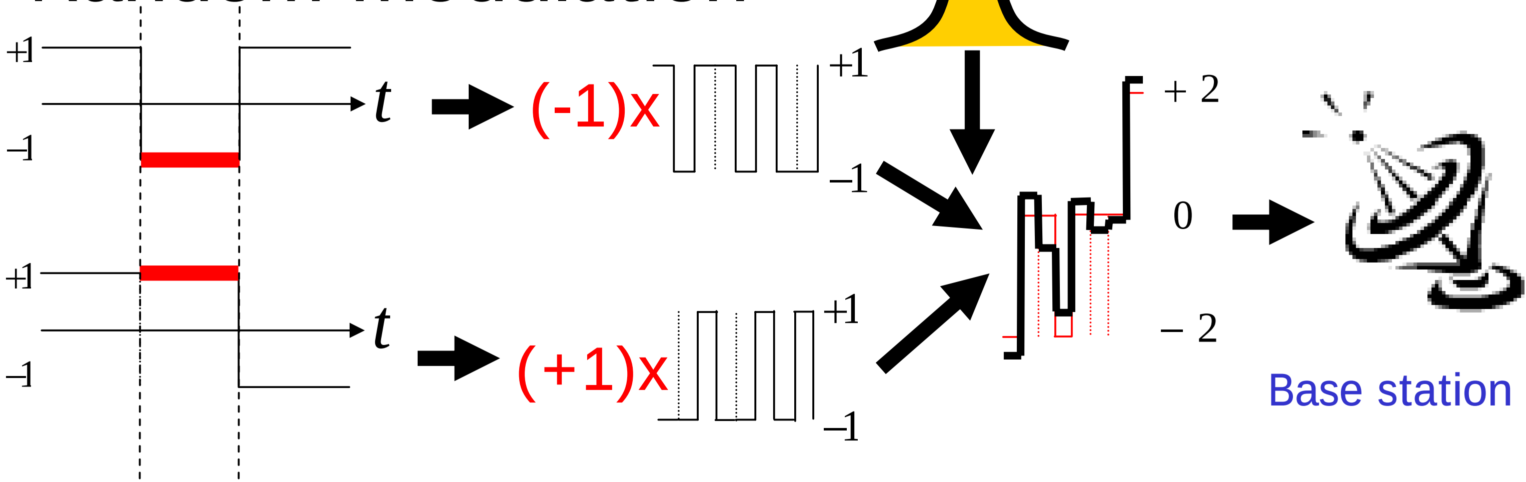
Outline

- CDMA system and multi-user detection
- Graphical expression & belief propagation (BP)
- Statistical-mechanical approach
- Extension to survey propagation (SP)
- Summary

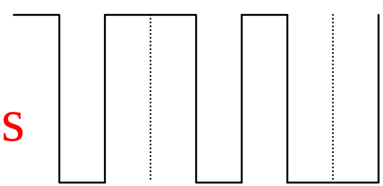
DS/BPSK CDMA system

- Random modulation





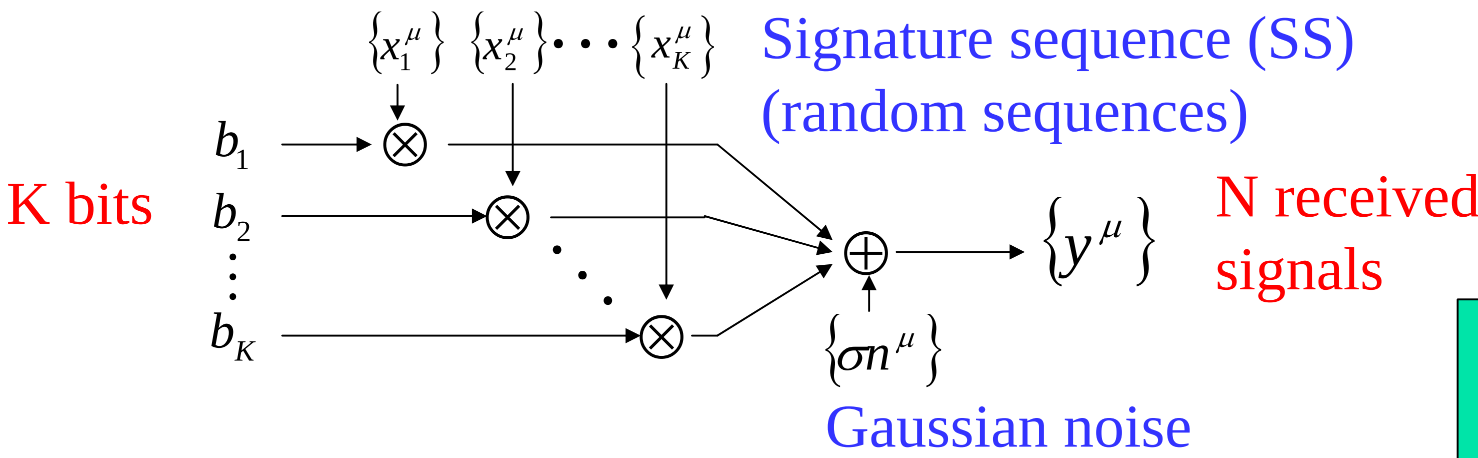
Modulate each signal with
N random binary sequences
(W-CDMA: up to N=256)



$$\mathbf{x} = (x_1, x_2, \cdots, x_8)$$

Mathematical model

- Multi-user (many-to-one) communication



$$
y^{\mu} = \frac{1}{\sqrt{N}} \sum_{i=1}^{K} x_i^{\mu} b_i + \sigma n^{\mu}
$$

Problem: Detect $\boldsymbol{b}$ from $\boldsymbol{y}$.

What is required for the user detection

- Real-time communication
  - Detection must be quickly performable
  - Low error rate is preferred
- As b_i are binary variables, development of detection schemes satisfying these requirements is non-trivial

$$\boxed{y^\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^{K} x_i^\mu b_i + \sigma n^\mu}$$

Linear equation: difficult to solve for discrete variables

A useful property

- Random signature sequences

$$x_i^\mu \text{ : i.i.d. from } P(x) = \frac{1}{2}\delta(x+1) + \frac{1}{2}\delta(x-1)$$

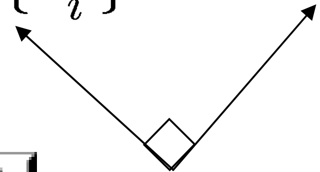
$$\Downarrow$$

$$\frac{1}{N}\boldsymbol{x}_i \cdot \boldsymbol{x}_j = \frac{1}{N}\sum_{\mu=1}^{N} x_i^\mu x_j^\mu = \begin{cases} 1, & i = j \\ O(N^{-1/2}) & i \neq j \end{cases}$$

  - Random SSs are nearly orthogonal to each other

User i's SS $\boldsymbol{x}_i = \{x_i^\mu\}$

User j's SS $\boldsymbol{x}_j = \{x_j^\mu\}$



Single-user detection

- Convenient method to detect b_i (in use)
  - Operate $\boldsymbol{x}_i$ to the received signals $\{y^\mu\}$

$$h_i = \frac{1}{\sqrt{N}}\sum_{\mu=1}^{N} x_i^\mu y^\mu = \underbrace{b_i}_{\text{Signal}} + \underbrace{\sum_{j\neq k}\frac{1}{N}\sum_{\mu=1}^{N} x_i^\mu x_j^\mu b_j}_{\text{Cross-talk noise}} + \underbrace{\frac{1}{\sqrt{N}}\sum_{\mu=1}^{n} x_i^\mu \sigma n^\mu}_{\text{Channel noise}}$$

$$\boxed{\hat{b}_i^{\mathrm{SD}} = \mathrm{sign}(h_i)}$$

$\boxed{\text{small:}\, O(N^{-1/2})}$

- Good news
  - Quickly performable using only the focused user's SS
  - In use in standard systems
- Bad news
  - High bit error rate (BER) when # of users is large.

#Better scheme is demanded for high-perform. comm.

Optimal detection

- Bayesian approach

Uniform Prior

$$P(\boldsymbol{b}) = 2^{-K}$$

Gaussian channel

$$P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(y^\mu - \frac{\boldsymbol{x}^\mu \cdot \boldsymbol{b}}{\sqrt{N}}\right)^2}{2\sigma^2}}$$

$$P(\boldsymbol{b}|D^N) = \frac{P(\boldsymbol{b})\prod_{\mu=1}^{N} P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b})}{\sum_{\boldsymbol{w}} P(\boldsymbol{b})\prod_{\mu=1}^{p} P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b})} \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{\mu=1}^{N}\left(y^\mu - \frac{\boldsymbol{x}^\mu\cdot\boldsymbol{b}}{\sqrt{N}}\right)^2\right]$$

- Optimal detection (Max. Posterior Marginals: MPM)

$$\hat{b}_i = \underset{b_i}{\text{argmax}}\left\{\sum_{\boldsymbol{b}\backslash b_i} P(\boldsymbol{b}|D^N)\right\}$$

#Minimizes BER $\text{Prob}(\hat{b}_i \neq b_i)$

Statistical-Mechanical Approach to Probabilistic Information Processing

Computational difficulty

- **Unfortunately, performing the optimal detection is computationally difficult!**

$$P(\boldsymbol{b}|D^N) = \frac{P(\boldsymbol{b})\prod_{\mu=1}^{N} P(y^\mu|\boldsymbol{x}^\mu, \boldsymbol{b})}{\sum_{\boldsymbol{b}} P(\boldsymbol{b})\prod_{\mu=1}^{N} P(y^\mu|\boldsymbol{x}^\mu, \boldsymbol{b})}$$

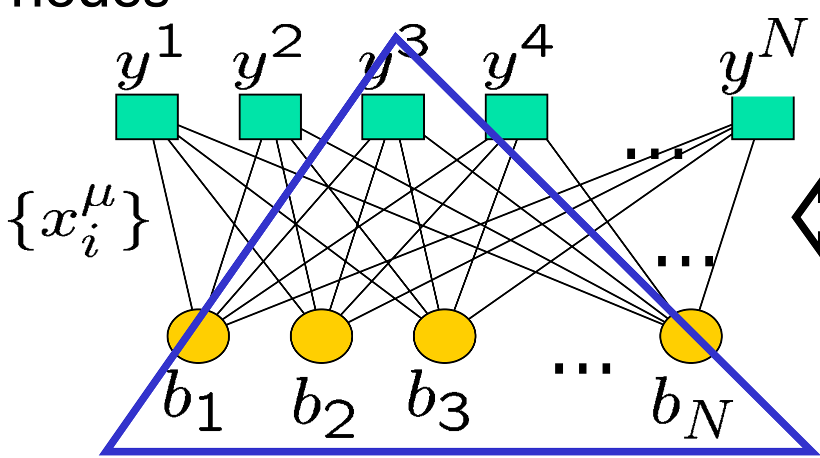
$O(2^N)$ summations!

- Development of good ***approx. algorithms*** is necessary.

Graphical representation

- In order to answer this, we introduce a factor (bipartite) graph expression of the posterior dist.

Data nodes



Variable nodes

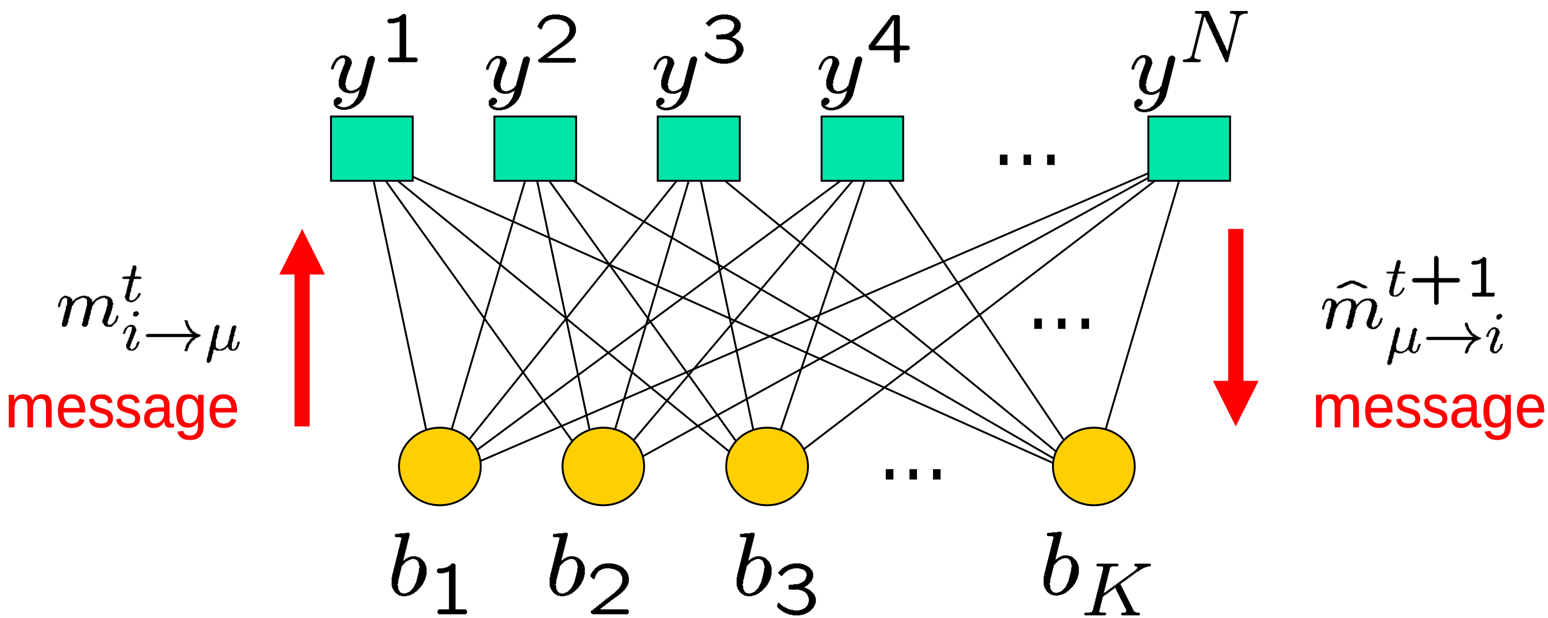
Posterior dist.

$$P(\boldsymbol{b}|D^N) = \frac{P(\boldsymbol{b})\prod_{\mu=1}^{N} P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b})}{\sum_{\boldsymbol{b}} P(\boldsymbol{b})\prod_{\mu=1}^{N} P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b})}$$

$$P(y^3|\boldsymbol{x}^3,\boldsymbol{b})$$

Belief Propagation (BP)

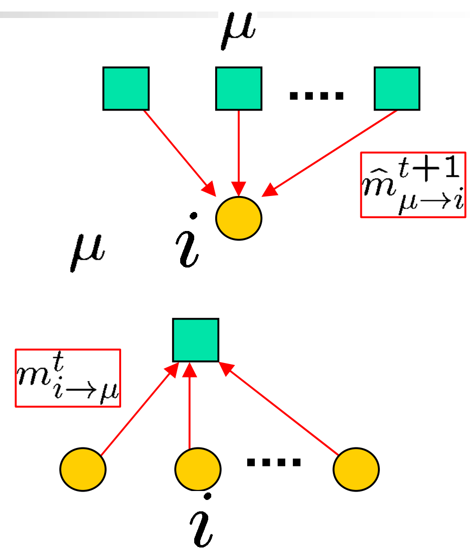
- Pearl (1987), MacKay (1995)
- Iteratively passing ***messages*** between the two types of nodes.



BP-based detection

- More specifically,

$$\begin{cases} \hat{m}_{\mu\to i}^{t+1} = \dfrac{\sum_{\boldsymbol{b}} b_i P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b}) \prod_{j\neq i}\left(\frac{1+m_{j\to\mu}^t b_j}{2}\right)}{\sum_{\boldsymbol{b}} P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b}) \prod_{j\neq i}\left(\frac{1+m_{j\to\mu}^t b_j}{2}\right)} \\ m_{i\to\mu}^t = \tanh\left[\sum_{\nu\neq\mu}\tanh^{-1}(\hat{m}_{\nu\to i}^t)\right] \end{cases}$$



- Bayes detection

$$m_i^t = \langle b_i \rangle^t = \tanh\left[\sum_{\mu=1}^{N}\tanh^{-1}(\hat{m}_{\mu\to i}^t)\right]$$:Posterior mean

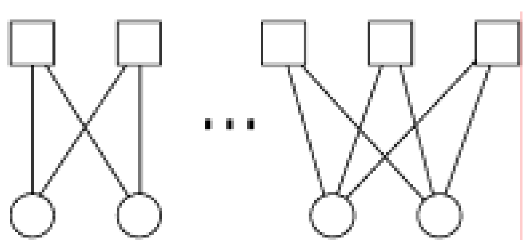
$$\hat{b}_i^t = \underset{b_i}{\mathrm{argmax}}\left\{\sum_{\boldsymbol{b}\backslash b_i} P^t(\boldsymbol{b}|D^N)\right\} = \mathrm{sign}(m_i^t)$$

- As information of all users is required, this is among multi-user detection algorithms.

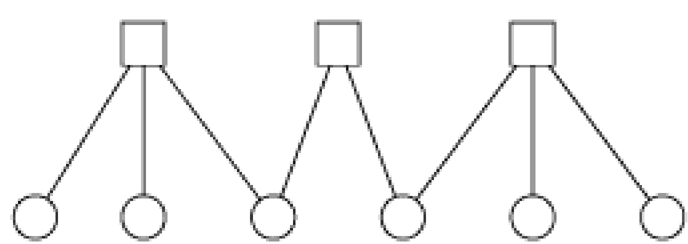
Theorem (Pearl 1987)

- BP provides the exact average for loop-free graphs after messages propagate once over the graph.
  - BP can be also employed in loopy graphs as an approximation algorithm (loopy BP)

Loops in graphs

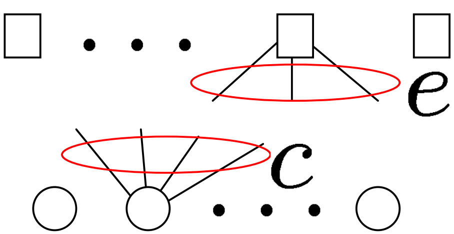


Loop-free graph



Intuitive speculation about loopy BP

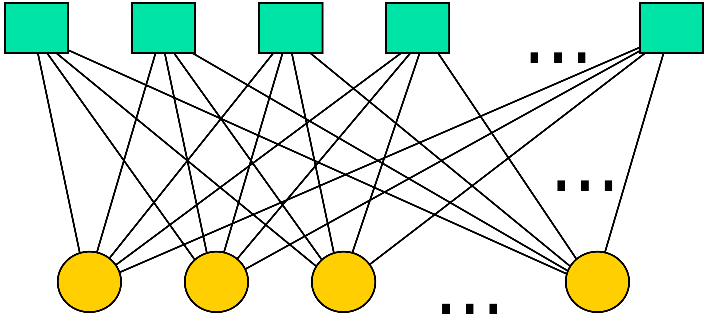
- As BP provides the exact result for loop free graphs, the lower the density of short loops in the graph is, the better performance will be gained.
- Cf) Random sparse graphs:
  - Loop lengths $O(\ln N)$.
  - Clique sizes $O(1)$.



- This speculation is empirically confirmed in several applications such as LDPC codes.

Pessimistic perspective

- CDMA system = Complete bipartite graph



Two bad news

- ***Many Short loops***
- ***Highly dense***

- Seeing these, you may feel that the current BP-based approach is not promising...
- However, techniques from S.M. can develop a nearly optimal algorithm based on BP, making use of the denseness of the graph appropriately when the network size is large.

Gaussian approximation

- Key property

$$P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(y^\mu - \frac{\boldsymbol{b}\cdot\boldsymbol{x}^\mu}{\sqrt{N}}\right)^2}{2\sigma^2}}$$ depends on $\boldsymbol{b}$ only through $\frac{\boldsymbol{b}\cdot\boldsymbol{x}^\mu}{\sqrt{N}}$.

- When $\boldsymbol{b} \sim \prod_j \left(\frac{1+m^t_{j\to\mu}b_j}{2}\right)$ and N is large, $\frac{\boldsymbol{b}\cdot\boldsymbol{x}^\mu}{\sqrt{N}}$ can be handled as a Gaussian variable (Mezard 1989, Opper and Winther 1996).
- This provides the following Gaussian approximation

$$\sum_{\boldsymbol{b}} P(y^\mu|\boldsymbol{x}^\mu,\boldsymbol{b}) \prod_{j\neq i}\left(\frac{1+m^t_{j\to\mu}b_j}{2}\right) \simeq \int \frac{d\Delta^\mu e^{-\frac{(\Delta^\mu - \langle\Delta^\mu\rangle^t_\mu)^2}{2V^t_\mu}}}{\sqrt{2\pi V^t_\mu}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y^\mu-\Delta^\mu)^2}{2\sigma^2}}$$

$O(2^N)$ compt → $O(1)$ compt.

Drastic reduction!

Statistical-Mechanical Approach to Probabilistic Information Processing

Stat. mech. algorithm

- Algorithm of $O(K^2N)$ computations/update

$$\begin{cases} a_{\mu\to i}^{t+1} = \dfrac{1}{\sigma^2+\beta(1-Q^t)}\left(y^\mu - \displaystyle\sum_{j\neq i}\frac{x_j^\mu m_{j\to\mu}^t}{\sqrt{N}}\right) \\ m_{i\to\mu}^t = \tanh\left[\displaystyle\sum_{\nu\neq\mu}\frac{x_i^\nu}{\sqrt{N}}a_{\nu\to i}^t\right] \end{cases}$$

$$m_i^t = \tanh\left[\sum_{\mu=1}^{N}\frac{x_i^\mu}{\sqrt{N}}a_{\mu\to i}^t\right]$$

$$\left(\begin{array}{l} \beta = \dfrac{K}{N} \sim O(1) \\ Q^t \equiv \frac{1}{N}\sum_{i=1}^N (m_i^t)^2 \end{array}\right)$$

Further reduction of compt. cost

- Taylor expansion

$$m_{i\to\mu}^t \simeq m_i^t - \left(1 - (m_i^t)^2\right) \underbrace{\hat{m}_{\mu\to i}^t}_{\text{small}}$$

- This makes it possible to express the BP update using only the singly-indexed variables

Further reduction to O(NK) computations/update

Faster BP-based algorithm

- Algorithm of $O(KN)$ computations/update

$$
\begin{cases}
a_\mu^{t+1} = \dfrac{1}{\sigma^2 + \beta(1-Q^t)} \left(y^\mu - \displaystyle\sum_{i=1}^{K} \frac{x_i^\mu}{\sqrt{N}} m_i^t + \beta(1-Q^t) a_\mu^t \right) \\
m_i^t = \tanh \left(\displaystyle\sum_{\mu=1}^{N} \frac{x_i^\mu}{\sqrt{N}} a_\mu^t + \frac{m_i^{t-1}}{\sigma^2 + \beta(1-Q^{t-1})} \right)
\end{cases}
$$

#Expressed by only singly-indexed variables

Remark (I)

- Computational cost is similar to that of a conventional algorithm
- Multi-stage detection (Varanasi and Aazhang 1991)

$$\begin{cases} a_\mu^{t+1} = y^\mu - \sum_{i=1}^{K} \frac{x_i^\mu}{\sqrt{N}} m_i^t \\ m_i^t = \mathrm{sign}\left(\sum_{\mu=1}^{N} \frac{x_i^\mu}{\sqrt{N}} a_\mu + m_i^{t-1} \right) \end{cases}$$

Remark (II)

- The fixed point of the BP-based algorithm is characterized by the TAP equation of this system.

$$m_i = \tanh\left[\frac{1}{\sigma^2}\left(h_i - \sum_{j\neq i} J_{ij} m_j\right) - \frac{\beta(1-Q)m_i}{\sigma^2(\sigma^2+\beta(1-Q))}\right]$$

$$\left(h_i = \frac{1}{\sqrt{N}}\sum_{\mu=1}^{N} x_i^\mu y^\mu, \quad J_{ij} = \frac{1}{N}\sum_{\mu=1}^{N} x_i^\mu x_j^\mu, \quad Q = \frac{1}{K}\sum_{i=1}^{K} m_i^2\right)$$

Remark (III)

- Macroscopic dynamics (Density Evolution) can be well captured by the iteration of SP eq. of the replica symmetric (RS) analysis (Tanaka 2002).

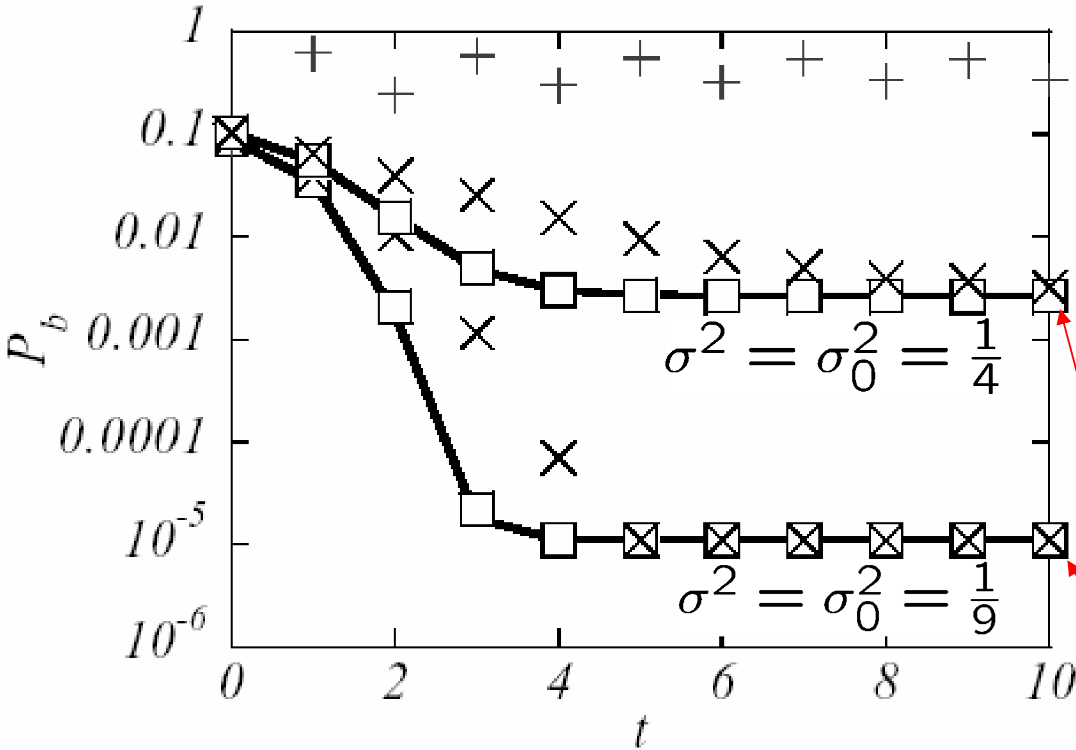
$$
\begin{cases}
E^{t+1} = \dfrac{1}{\sigma^2 + \beta(1-Q^t)}, \quad F^{t+1} = \dfrac{\beta\left(1-2M^t+Q^t\right)+\sigma_0^2}{\left[\sigma^2+\beta(1-Q^t)\right]^2} \\
M^t = \int Dz \tanh\left(\sqrt{F^t}z + E^t\right), \quad Q^t = \int Dz \tanh^2\left(\sqrt{F^t}z + E^t\right)
\end{cases}
$$

- This implies that the developed approx. algorithm practically converges to the 'exact' solution when Nishimori's condition $\sigma^2 = \sigma_0^2$ holds, for which no RSB is empirically expected.

Statistical-Mechanical Approach to Probabilistic Information Processing

Method comparison

- Condition: $K = 1000 \quad N = 2000$



□ : Proposed

X : MSD

+ : TAP eq.

Lines : Theory

'Exact values' predicted by the replica method.

Remark (IV)

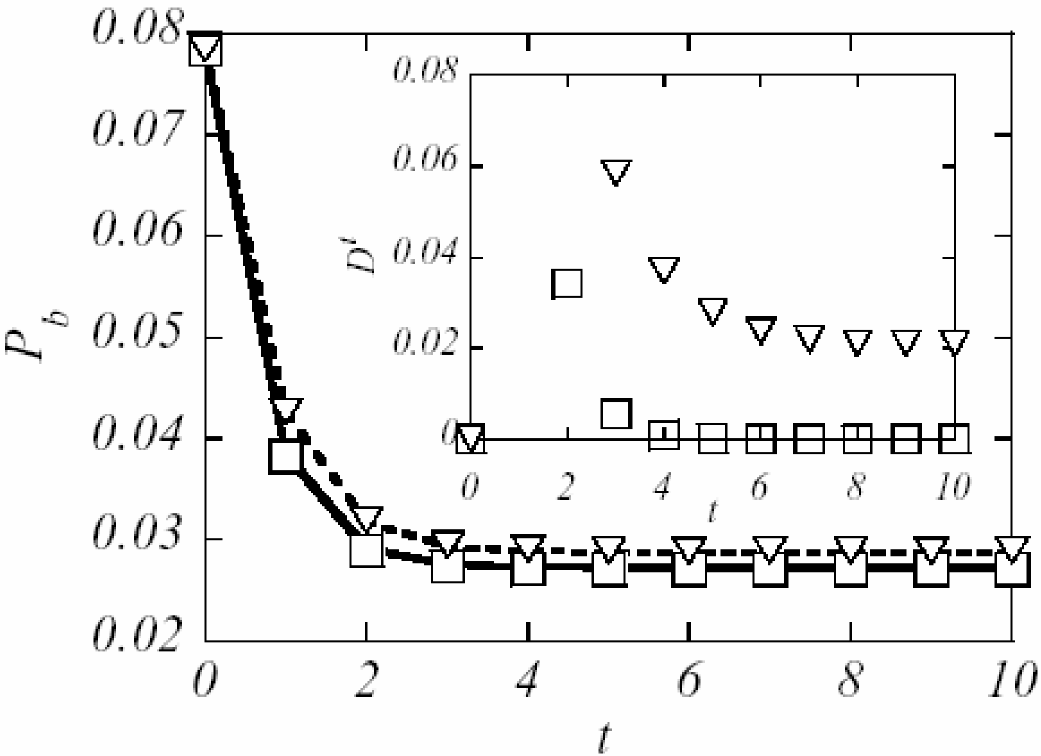
- Microscopic (dynamical) instability condition of the fixed point becomes identical to that of the AT instability of the (equilibrium) replica analysis
  - This can occur when the assumed noise parameter σ^2 is sufficiently smaller than the true one σ_0^2 .

Microscopic Instability Condition(=AT Condition)

$$\frac{1}{\left[\sigma^2+\beta(1-Q)\right]^2} \times \frac{1}{N} \sum_{i=1}^{K} \left(1-m_i^2\right)^2 > 1$$

Microscopic instability

$$K = 500,\ N = 2000,\ \sigma_0^2 = 0.25$$



∇: $\sigma^2 = 0.01$

$\square$: $\sigma^2 = \sigma_0^2 = 0.25$

$$D^t = \frac{1}{K}\sum_{i=1}^{K}(m_i^t - m_i^{t-1})^2$$

Measures Micro. Instability

Extension to Survey Prop.

- Link to the AT instability motivates us to develop an algorithm based on the survey propagation (SP), which can describe RSB phase.
- SP (Mezard-Parisi-Zecchina 2002)
  - Distributions of messages (surveys)
  - Introduction of RSB parameter x
  - Correspondence to the 1RSB solution
- In the current system, the central limit theorem makes it possible to express SP compact.

Remarks on SP-based algorithm

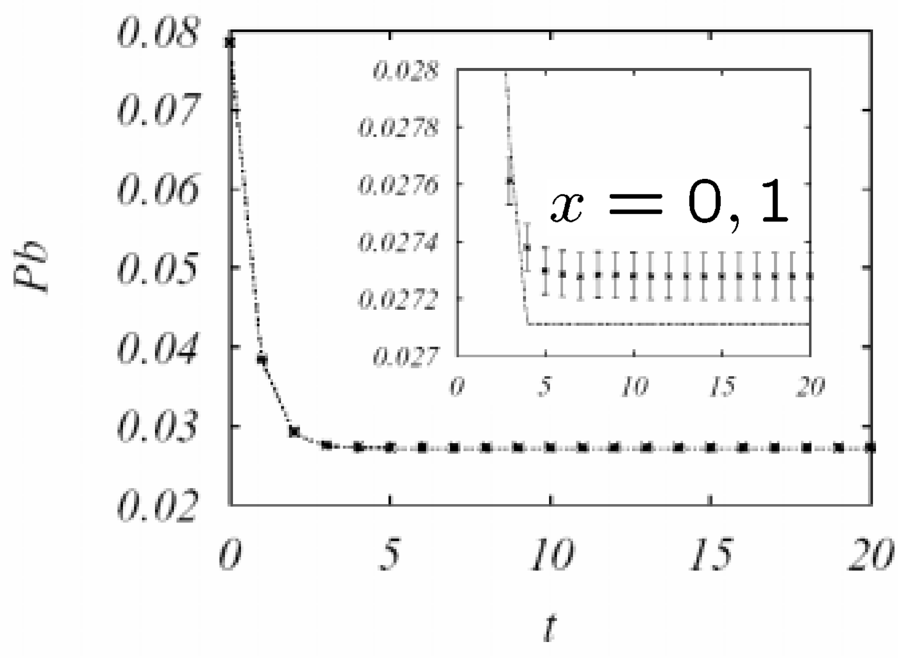
- Under tree approximation, the macroscopic dynamics is described by natural iteration of 1RSB SP eqs.
  - The microscopic instability does not correspond to the AT instability of the 1RSB solution.
- Robust for mismatch of noise parameter σ^2
  - AT stable: as good as BP independently of RSB param. x
  - AT unstable: can be better than BP
- Tuning x by the free energy maximization principle is not necessarily optimal for reducing BER.
- $x = 1$ and $x = 0$ provide different performance when BP is unstable although both of these param. choices are reduced to the RS solution in the replica analysis.

$x = 1$ vs $x = 0$

$$\beta = K/N = 500/2000$$

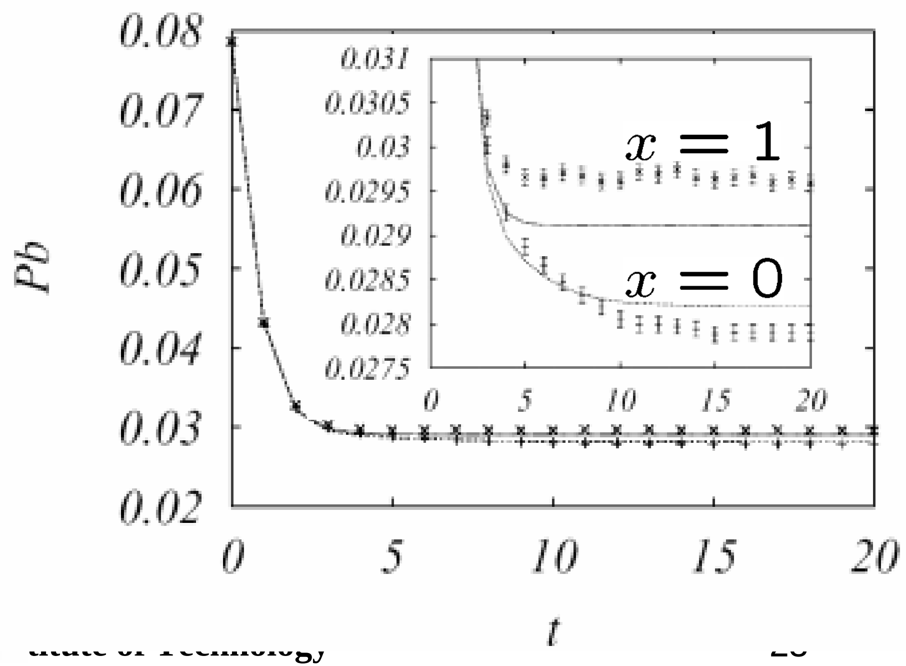
RS regime

$$(\sigma_0^2, \sigma^2) = (0.25, 0.25)$$



RSB regime

$$(\sigma_0^2, \sigma^2) = (0.25, 0.01)$$



Summary

- Approx. algorithm for the Bayes detection in CDMA system based on BP and SM.
  - Quicker convergence.
  - Compt. cost is not significantly increased.
  - Excellent consistency with the replica theory.
    - Converges to the optimal solution in the thermodynamic limit
      if the assumed hyper parameters are correct.
- Extension to SP
  - SP can serve as a robust algorithm for mismatch of hyper parameters.

References

- Y. Kabashima, A CDMA multiuser detection algorithm on the basis of belief propagation, J. Phys. A 36, 11111-11121 (2003)
- Y. Kabashima, Propagating beliefs in spin glass models, J. Phys. Soc. Jpn. 72, 1645-1649 (2003)
- Available from http://www.sp.dis.titech.ac.jp/~kaba/publish.html