Statistical mechanical approach to massive Bayesian inference

Yoshiyuki Kabashima
Dept. of Compt. Intell. & Syst. Sci.
Tokyo Institute of Technology

Outline

- What is Bayesian statistics?
- Statistical mechanics: Bayesian statistics of material objects
- Statistical mechanical approach to Bayesian inference
- Conclusion

Bayesian statistics

- A framework to infer unobserved variables from observed data
- ***Bayes' theorem***

Cond. prob.

Prior prob.

Posterior Prob.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Unobserved variables

Observed data

  - Systematic unification of novel information and prior knowledge
  - Mathematically natural (compatible with Kolmogorov's postulate)
  - Wide applicability

Generic difficulties (I)

- However, the Bayesian framework was not widely used for a long time
- ***Lack of objectivity*** : old (fundamental) difficulty

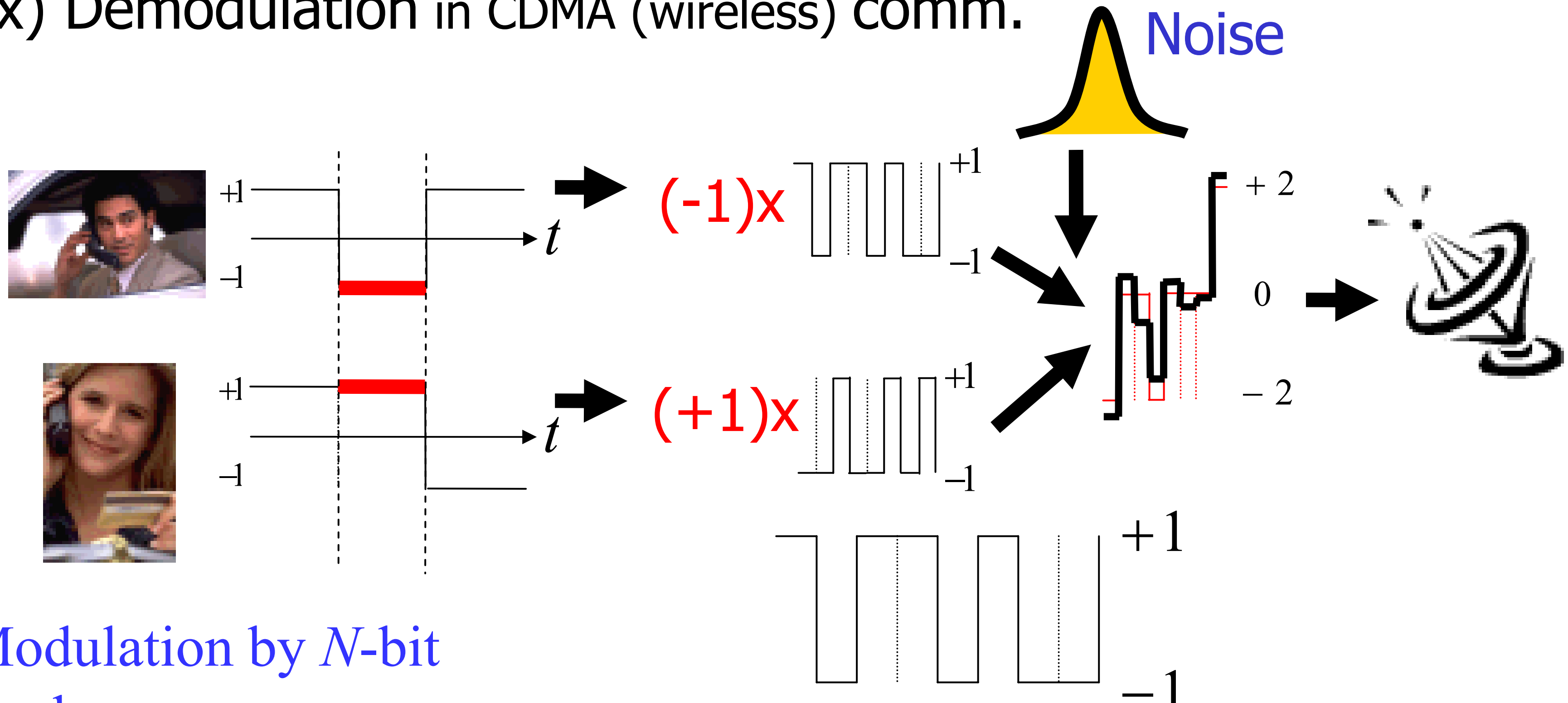
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

  - Choice of prior probability depends on individual's preference
  - Some people hesitate to use such a *subjective probability* for scientific purpose
  - However, recent IT innovation sometimes makes it possible to provide the prior prob. *objectively* from huge preliminary data

→ ***Therefore, we here skip this problem***

Generic difficulties (II)

- ***Computational complexity*** : new (technological) difficulty
- Ex) Demodulation in CDMA (wireless) comm.

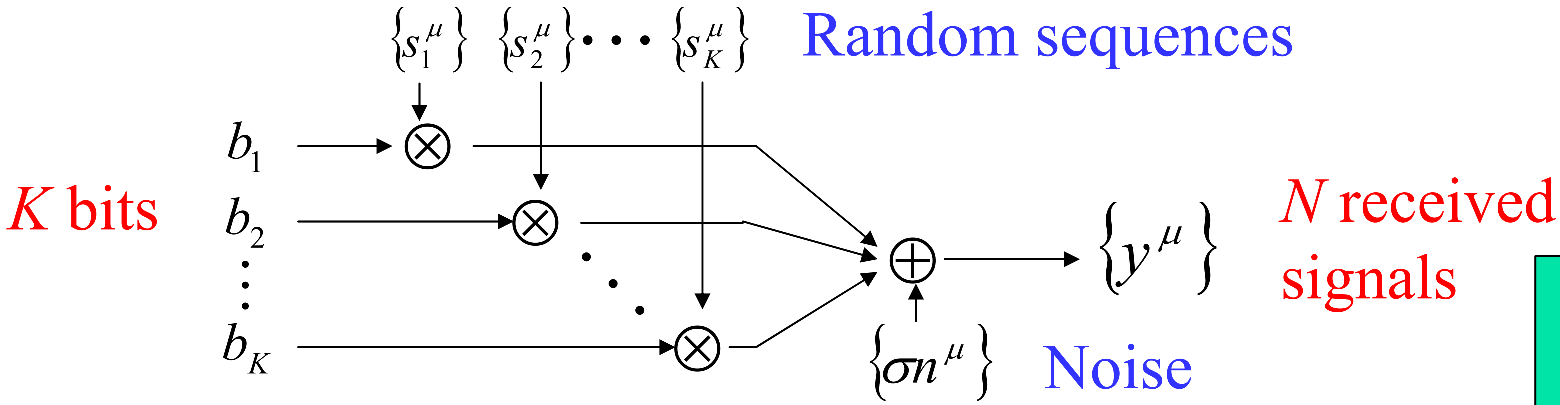


Modulation by N-bit
random sequences

$$\boldsymbol{s} = (s^1, s^2, \ldots, s^8)$$

CDMA model

- K-user model



$$y^{\mu} = \frac{1}{\sqrt{N}} \sum_{k=1}^{K} s_k^{\mu} b_k + \sigma n^{\mu}$$

Demodulation : estimate $\boldsymbol{b}$ from $\boldsymbol{y}$

Bayesian formulation

- Posterior probability given received signals

$$P(\boldsymbol{b}|\boldsymbol{y},\boldsymbol{s}) = \frac{P(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{s})P(\boldsymbol{b})}{P(\boldsymbol{y},\boldsymbol{s})} = \frac{P(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{s})P(\boldsymbol{b})}{\sum_{\boldsymbol{b}} P(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{s})P(\boldsymbol{b})}$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{\mu=1}^{N}(y^\mu - N^{-1/2}\sum_{k=1}^{K} s_k^\mu b_k)^2\right]$$

✓ Distribution of *K* (many) binary variables

✓ Variables are interdependent

Computational difficulty

- Bit error rate (BER: performance measure)

$$P_b = \mathrm{Prob}[\widehat{b}_k \neq b_k]$$

is minimized by demodulation based on the posterior

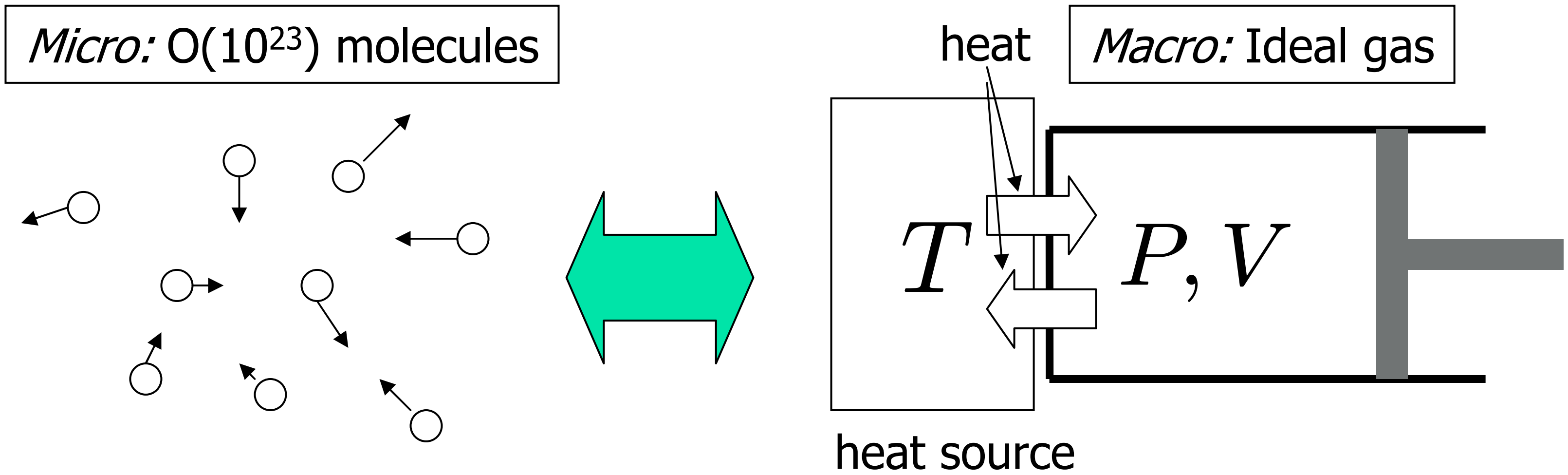
- Unfortunately, computational cost for evaluating this probability makes this scheme *practically unfeasible for large systems*

$$P(\boldsymbol{b}|\boldsymbol{y},\boldsymbol{s}) = \frac{P(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{s})P(\boldsymbol{b})}{P(\boldsymbol{y},\boldsymbol{s})} = \frac{P(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{s})P(\boldsymbol{b})}{\sum_{\boldsymbol{b}} P(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{s})P(\boldsymbol{b})}$$

$O(2^K)$ summations

Statistical mechanics: Bayesian inference of material objects

- ***My claim:*** Statistical mechanics can offer practical solutions to the intrinsic computational difficulty of Bayesian inference in large systems
- **What is *statistical mechanics?***
  - A branch of physics which relates the microscopic properties of many elements to the macroscopic behavior of a system that the elements constitute



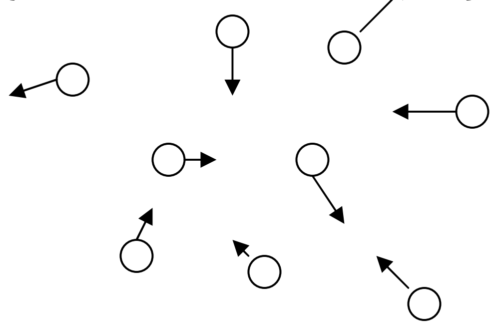
From *microscopic* constituents to *macroscopic* behavior

- Microscopic description
  - O(10^{23}) variables
- Macroscopic description
  - a few variables

Hamiltonian

$$H = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2m}$$

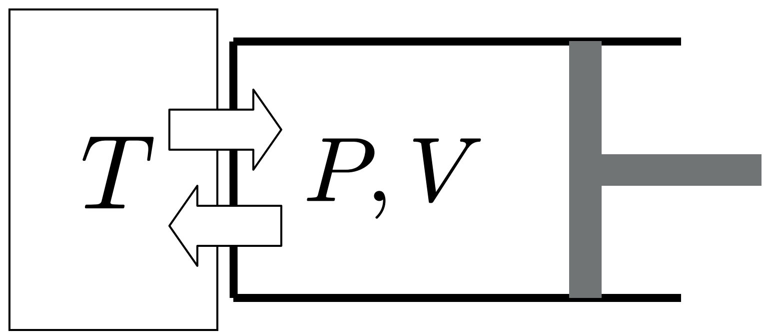
$$\left(\Longleftrightarrow \quad \boldsymbol{F}_i = m\boldsymbol{a}_i \right)$$



Stat. Mech.

Equation of state

$$PV = Nk_{\mathrm{B}}T$$



Analogy between Bayesian inference and stat. mech.

	Bayesian Infer.	Stat. mech.
Micro	$P(B\|A) = \dfrac{P(A\|B)P(B)}{P(A)}$	$H = \sum_{i=1}^{N} \dfrac{\|\boldsymbol{p}_i\|^2}{2m}$
Macro	$\mathrm{Prob}[\widehat{b}_k \neq b_k]$	$PV = Nk_{\mathrm{B}}T$

Notions and techniques developed in stat. mech. may be useful for Bayesian inference as well

Statistical mechanical approach to Bayesian inference (I)

- ***Equation of state***
  - Typical performance of ***large*** CDMA systems can be characterized by coupled equations of a few macroscopic parameters

Micro

$$P(\boldsymbol{b}|\boldsymbol{y},\boldsymbol{s}) = \frac{P(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{s})P(\boldsymbol{b})}{\sum_{\boldsymbol{b}} P(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{s})P(\boldsymbol{b})}$$

Tanaka (2002)

$$(K, N \to \infty \ \beta = K/N \sim O(1))$$

Macro

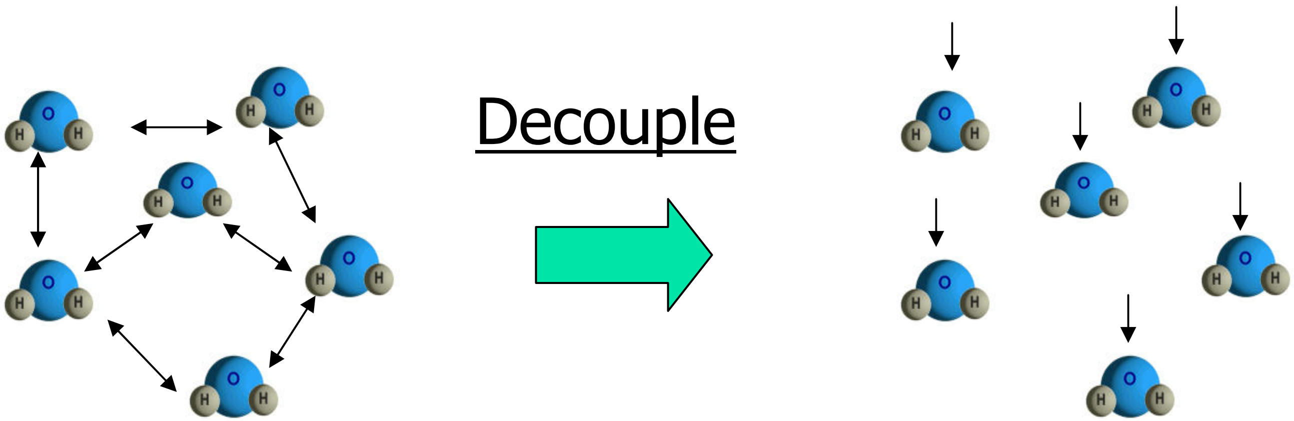
EOS

$$\begin{cases} m = \int \frac{dze^{-z^2/2}}{\sqrt{2\pi}} \tanh\left(\sqrt{\widehat{q}}z + \widehat{m}\right) & \widehat{m} = \frac{1}{\sigma^2 + \beta(1-q)} \\ q = \int \frac{dze^{-z^2/2}}{\sqrt{2\pi}} \tanh^2\left(\sqrt{\widehat{q}}z + \widehat{m}\right) & \widehat{q} = \frac{\beta(1-2m+q)+\sigma^2}{[\sigma^2 + \beta(1-q)]^2} \end{cases}$$

$$\mathrm{Prob}[\widehat{b}_k \neq b_k] = \int_{\widehat{m}/\sqrt{\widehat{q}}} \frac{dz}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right]$$

Statistical mechanical approach to Bayesian inference (II)

- ***Computationally feasible approximate inference***
  - Advanced *mean field approximation* methods offer a useful guideline for developing feasible approximate inference algs.
- **Mean field approximations**
  - Methods to approximate a many-body system with interactions by a bunch of single-body systems

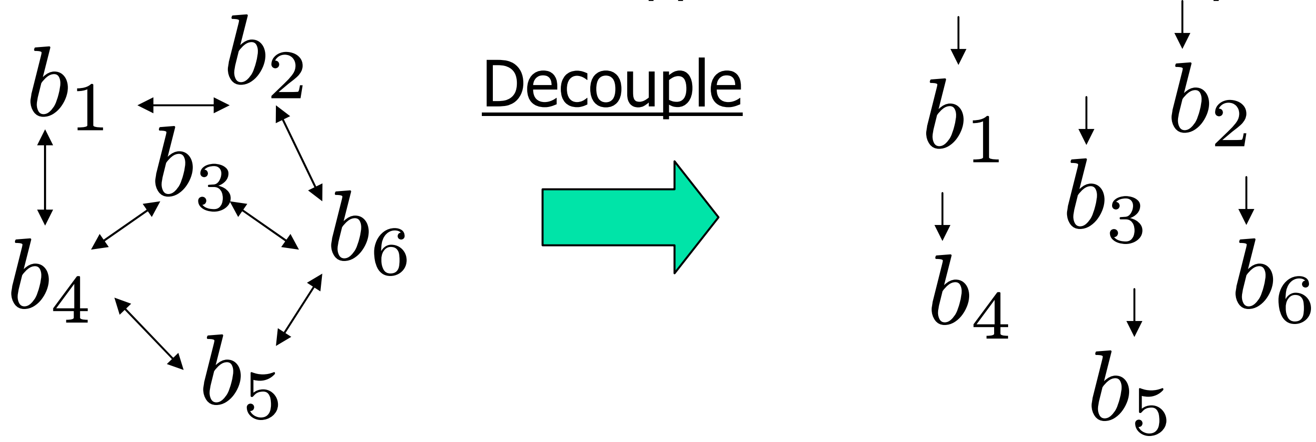


CDMA demodulation as a many body problem

- Posterior prob. of the CDMA demodulation problem

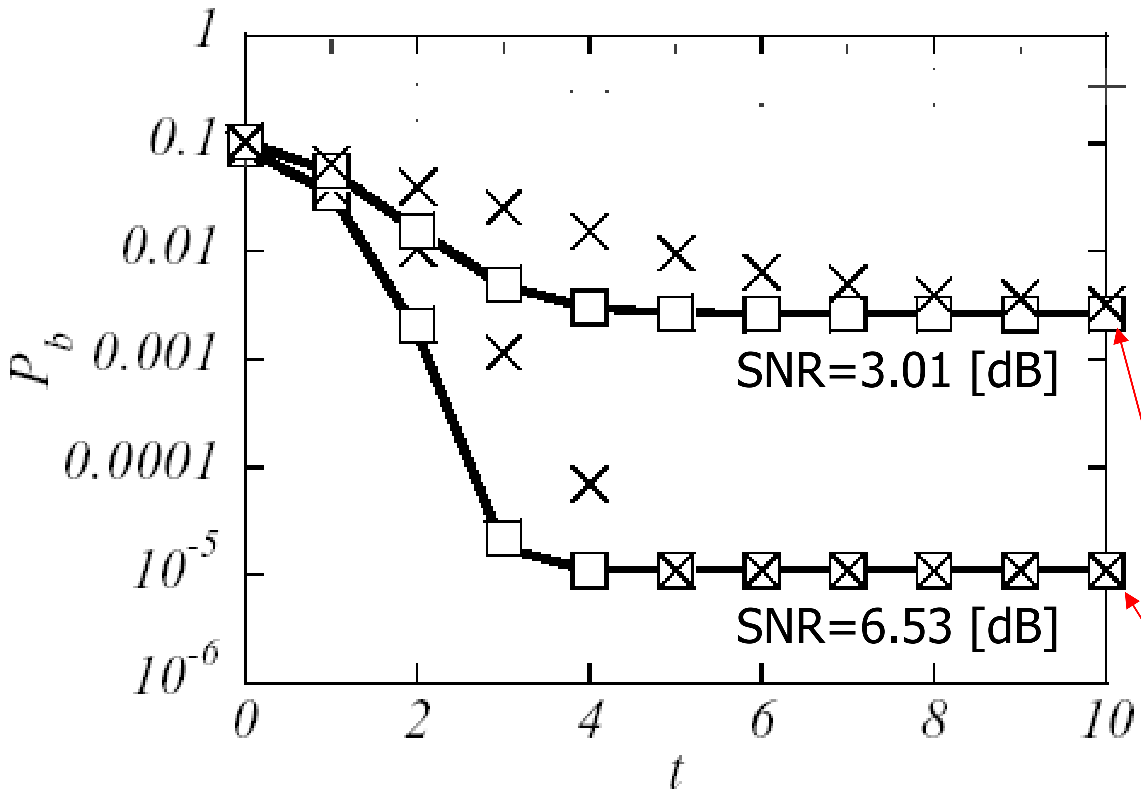
$$P(\boldsymbol{b}|\boldsymbol{y},\boldsymbol{s}) \quad \propto \quad \exp\left[-\frac{1}{2\sigma^2}\sum_{\mu=1}^{N}(y^{\mu} - N^{-1/2}\sum_{k=1}^{K} s_k^{\mu} b_k)^2\right]$$

- **Mean field demodulation**
  - The posterior can be regarded as a system of “many interacting bits”
  - Application of the mean field approximations to the Bayesian inference



Development of computationally feasible demodulation algs.

- Performable in O(NK) computational time



$$K = 1000 \quad N = 2000$$

□ : **Stat. Mech D.**
#YK (2003)

X : Multi-stage D.
#Baranasi-Aazhang (1990)

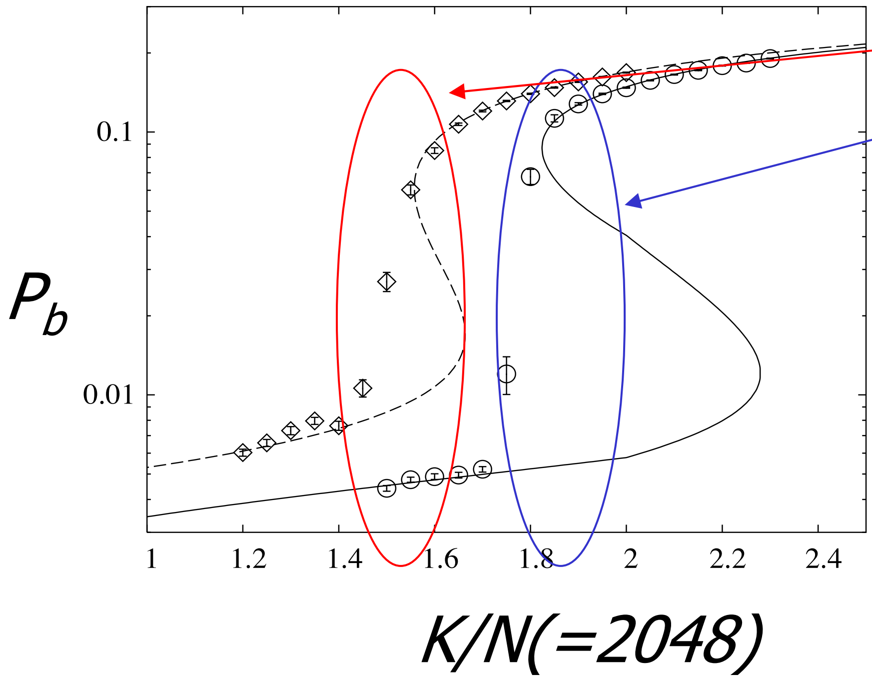
Lines : Theory

'Exact values' predicted by the *equation of state*

“Phase transition” in CDMA communication

- Peculiar behavior of the CDMA demodulation
  - Lines: predicted by the ***equation of state***
  - Markers: obtained by a ***mean field demodulation algorithm***

#Takeda, Uda and YK (2006)



Drastic change of performance

↕

“Phase transition”

Ex)
Water ↔ Ice
Water ↔ Vapor
Ice ↔ Vapor

Other applications

- **Error correcting codes**
  Sourlas (1989)
  YK and Saad (1998, 1999)
  YK, Murayama and Saad (2000)
  Montanari and Sourlas (2000)
- **Cryptography**
  YK, Murayama and Saad (2000)
- **Data compression**
  Hosaka, YK and Nishimori (2002)
  Murayama (2002)
  Ciliberti, Mezard and Zecchina (2005)
- **Combinatorial problems**
  Mezard and Parisi (1985)
  Fu and Anderson (1986)
  Monasson and Zecchina (1996)
  Mezard, Parisi and Zecchina (2002)
- **Pattern recognition**
  Gardner and Derrida (1988)
  Gyorgyi and Tishby (1990)
  Opper and Winther (1996, 2001)
  Uda and YK (2005)
  YK (2008)
  Shinzato and YK (2008)
- **Associative memory**
  Amit, Gutfreund and Sompolinsky (1984)
  Amari and Maginu (1988)
  Ozeki and Nishimori (1993)
  Shiino and Fukai (1993)
  Okada (1995)
- **Image restoration**
  Tanaka and Morita (1995)

Conclusion

- Bayesian inference in large systems has the intrinsic computational difficulty
- Statistical mechanics can offer practical solutions to this problem