Rare Event Sampling using Multicanonical Monte Carlo

Yukito IBA

The Institute of Statistical Mathematics

This is my 3rd oversea trip;
two of the three is to Australia.

Now I (almost) overcome airplane phobia

This is my 3rd oversea trip;
two of the three is to Australia.

Now I (almost) overcome airplane phobia

History of Dynamic Monte Carlo

1953: Metropolis algorithm

1950s~: thermal averages in physics

sampling from canonical distribution

1990s~: Bayesian data analysis

sampling from posterior distribution

Not the subject of this talk

It is called “MCMC”: Markov Chain Monte Carlo
a must-study for business school students

Dynamic Monte Carlo is a general methodology.
There should be many other potential applications.

This Talk: Rare Event Sampling

sampling from tails from distributions
sampling "large deviations"
→ [quenched random averages in stat. phys.]

↔ conventional usage

(sampling from canonical distribution
thermal averages in physics)

Contents

First, we explain an example:

"large dev. in the largest eigenvalue of random matrices"

A direct application of

multicanonical/ Wang-Landau algorithm

Berg(1991,1992), Wang and Landau(2001)

energy E → target quantity ξ

densiy of states $D(E)$ → probability $P(\xi)$

Second, we show a few other examples of

"rare event sampling" in physics.

joint works with..

- Nen Saito (Osaka Univ.)
- Koji Hukushima (Univ. of Tokyo)
- Akimasa Kitajima (Osaka Univ.)

- Tatsuo Yanagita (Hokkaido Univ.)
- Toshio Aoyagi (Kyoto Univ.)

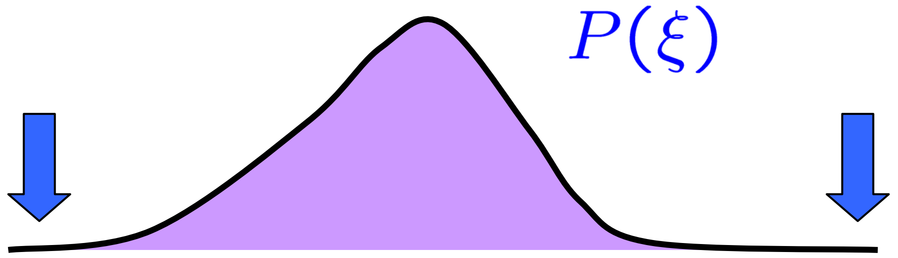
Rare Event Sampling

Generic

Assume that a variable x is sampled from $Q(x)$

Calculate the distribution $P(\xi)$ of a statistics $\xi(x)$

Tails of the distribution are difficult to estimate
by the naive sampling from $Q(x)$



Example.1 $\xi = \lambda_{\max}$

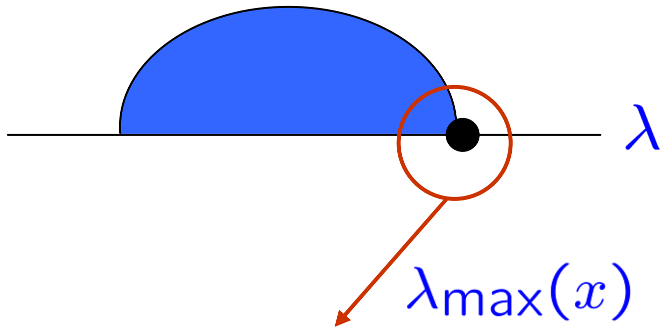
Rare Events in Random Matrices

Random Matrix x

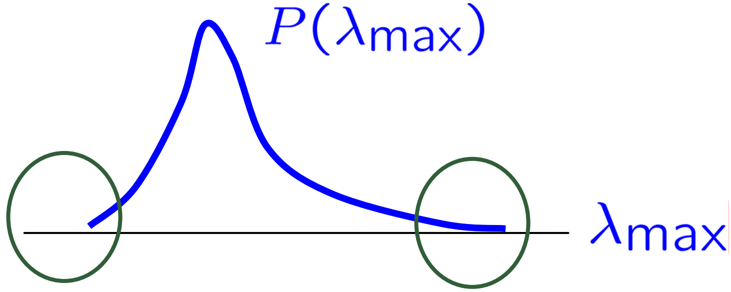
- i.i.d. Gaussian, symmetric
- sparse $Q(x)$
- zero-one (random graph)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots \\ a_{12} & a_{22} & a_{23} & a_{24} & \cdots \\ a_{13} & a_{23} & a_{33} & a_{34} & \cdots \\ a_{14} & a_{24} & a_{34} & a_{44} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

$\longleftarrow \quad N \quad \longrightarrow$



many samples from $Q(x)$



tails of distribution of max. eigenvalue

N Saito, Y Iba and K Hukushima
arXiv:1002.4499 (2010)

Large deviation / Rare event

- Naive Method : Sample directly from $Q(x)$

  Inefficient in the extreme tail region

- Introduction of bias:
  Sample from "fictitious" Gibbs distributions

$$W(x|\beta) \propto Q(x)\, \exp\left(-\beta \underline{\lambda_{\max}(x)}\right)$$

+ Parallell Tempering (Replica Exchange Monte Carlo) can be a solution

Large deviation / Rare events

- Naive Method : Sample directly from $Q(x)$

  Inefficient in the extreme tail region

- Introduction of bias:
  Sample from "fictitious" Gibbs distributions

$$W(x|\beta) \propto Q(x)\,\exp\left(-\beta\lambda_{\mathrm{max}}(x)\right)$$

- Muticanonical / Wang-Landau Sampling

$$W(x) \propto Q(x) P(\lambda_{\mathrm{max}}(x))^{-1}$$

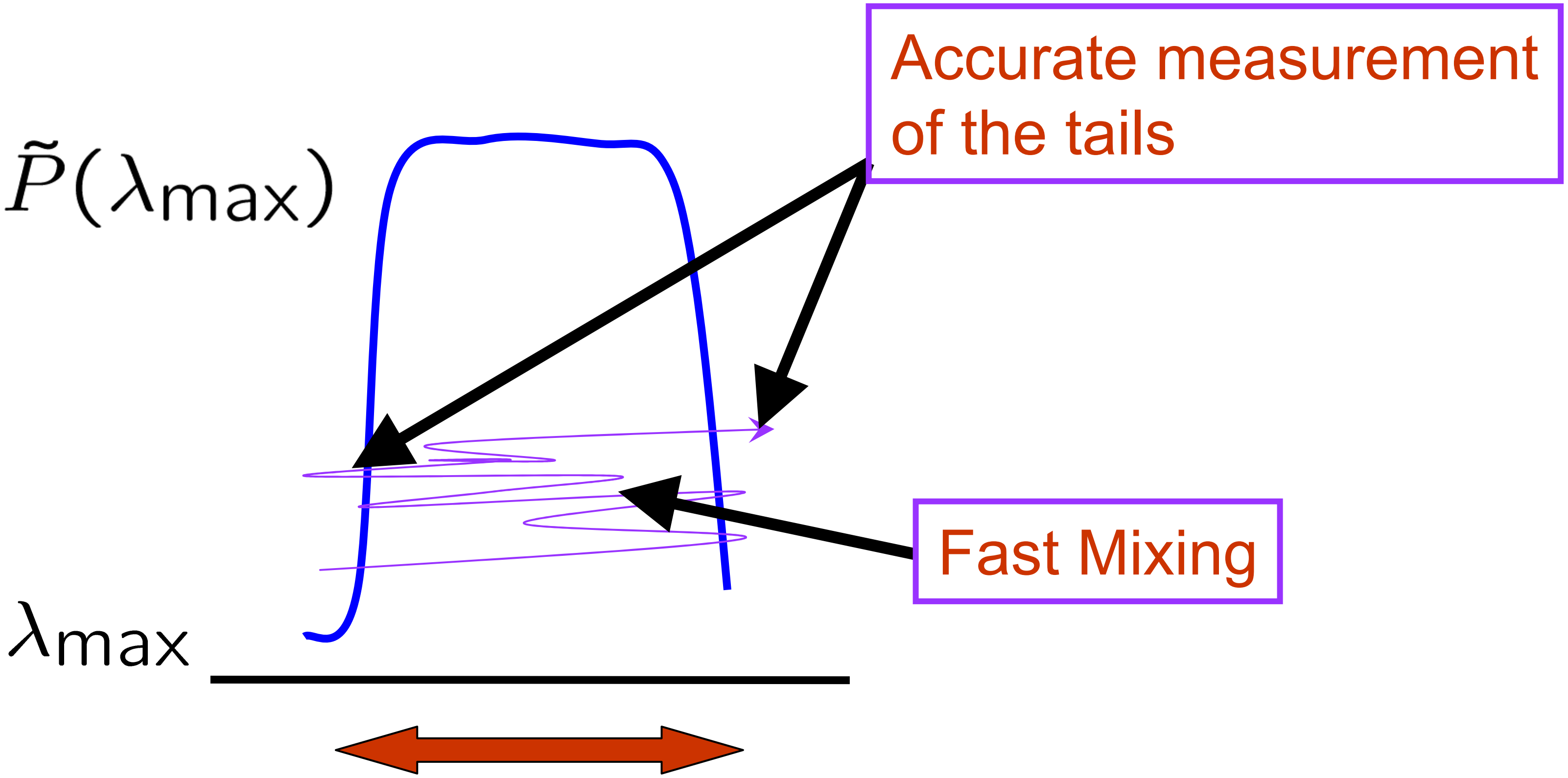
Multicanonical / WL Sampling

$$w(x) \propto \underbrace{P^*(\lambda_{\mathrm{max}}(x))^{-1}}_{\text{biasing factor}} \underbrace{Q(x)}_{\text{original}}$$

$$P^*(\lambda_{\mathrm{max}}) \simeq const. \times \underbrace{P(\lambda_{\mathrm{max}})}_{\text{the one which we want to calculate}}$$

$\simeq$ holds in a range we are interesed in.

(1) Flat Distribution of Target Quantity



Scanning broad range of $\lambda_{\rm max}$

Proof of Flatness

Naive $w_N(x) \propto Q(x)$ $\tilde{P}(\lambda_{\mathrm{max}}) = P(\lambda_{\mathrm{max}})$

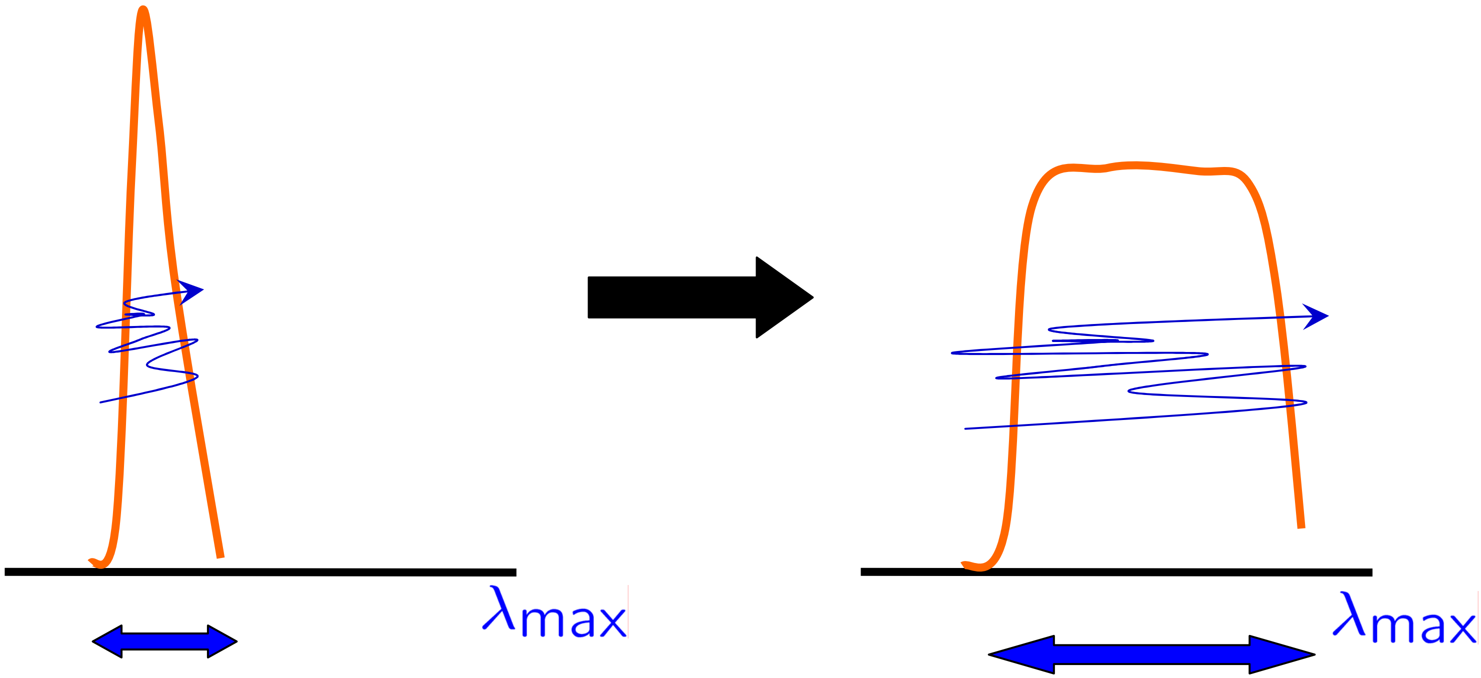
Multicanonical $w(x) \propto \underline{P^*(\lambda_{\mathrm{max}}(x))}^{-1} Q(x)$

→ $\tilde{P}(\lambda_{\mathrm{max}}) = \underline{P^*(\lambda_{\mathrm{max}})}^{-1} P(\lambda_{\mathrm{max}})$

$P^*(\lambda_{\mathrm{max}}) \simeq c P(\lambda_{\mathrm{max}})$ → $\tilde{P}(\lambda_{\mathrm{max}}) \simeq const.$

(2) Estimate $P^*(\lambda_{\max})$ by Iteration

$P^*(\lambda_{\max})$ Estimated by the iteration of preliminary runs



Wang-Landau algorithm

$\xi(x) \Leftrightarrow \lambda_{\text{max}}(x)$

1. Initialize Weights $w(\xi)$: Set $C = 1/e$
2. Set/Reset Histogram $H(\xi) = 0$
3. Metropolis update with the weight $w(\xi(x))Q(x)$
4. If the current state is x

   discount : $w(\xi(x)) \leftarrow w(\xi(x)) * C$

   increment: $H(\xi(x)) \leftarrow H(\xi(x)) + 1$
5. If $H(\xi)$ becomes "sufficiently flat"

   $C \leftarrow \sqrt{C}$ and Goto 2 else Goto 3.
6. If C become sufficiently small then quit.

(3) Reweighting

$\tilde{P}(\lambda_{\text{max}})$ output of the simulation

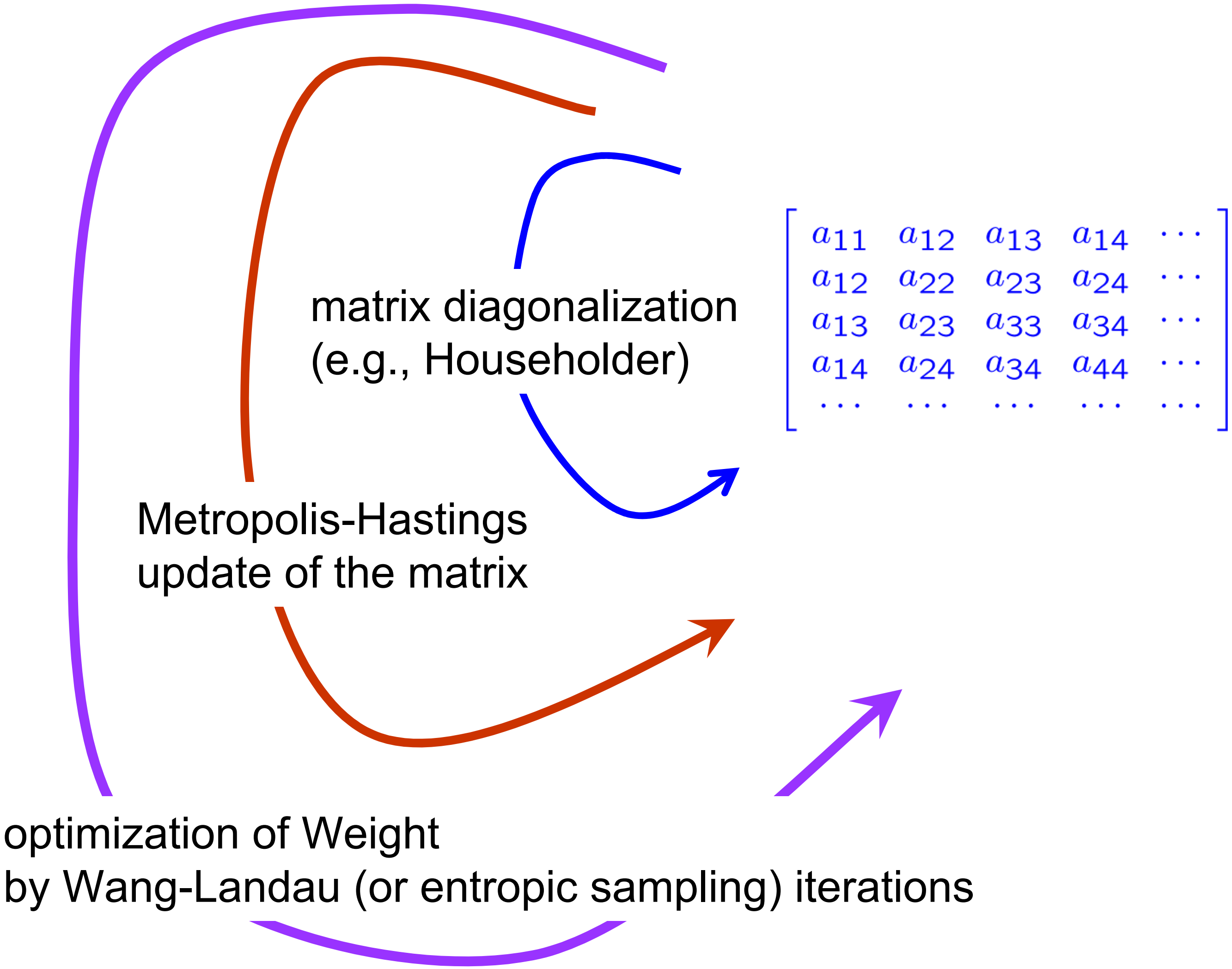
$Q(x)\underline{P^*(\lambda_{\text{max}})}^{-1}$ weight

$$\tilde{P}(\lambda_{\text{max}}) \propto P(\lambda_{\text{max}})\underline{P^*(\lambda_{\text{max}})}^{-1}$$

$$\Downarrow$$

$$\boxed{P(\lambda_{\text{max}}) \propto P^*(\lambda_{\text{max}})\tilde{P}(\lambda_{\text{max}})}$$

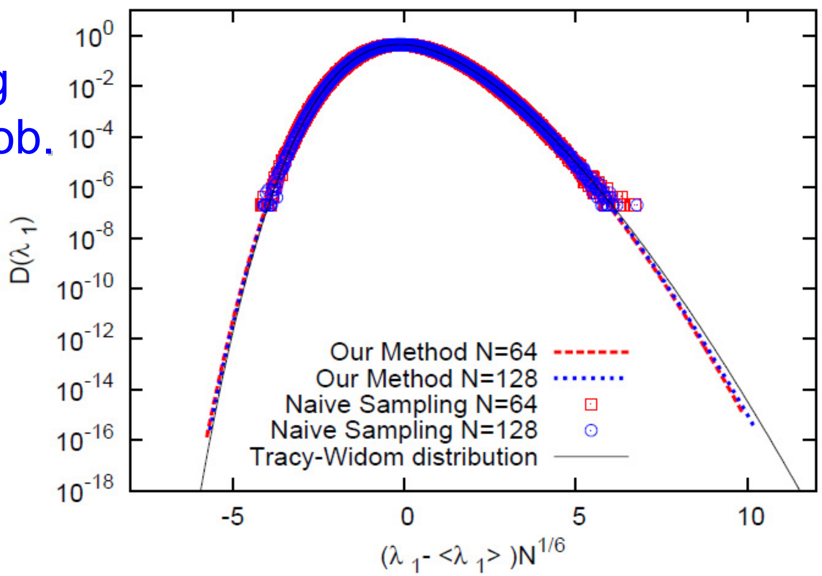
Random Matrices



Results (GOE)

Prob. $\sim 10^{-16}$ $\sim 10^{-120}$

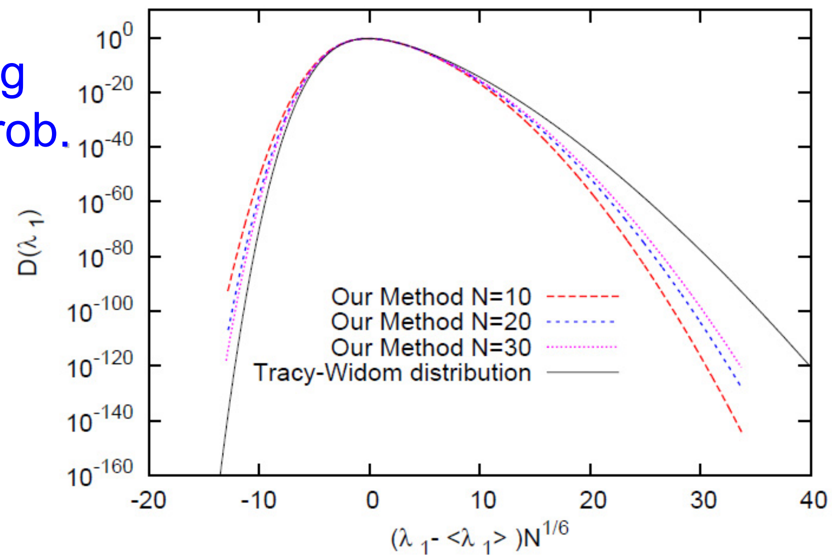
log prob.



max eigenvalue (scaled)

N=64, 128

log prob.

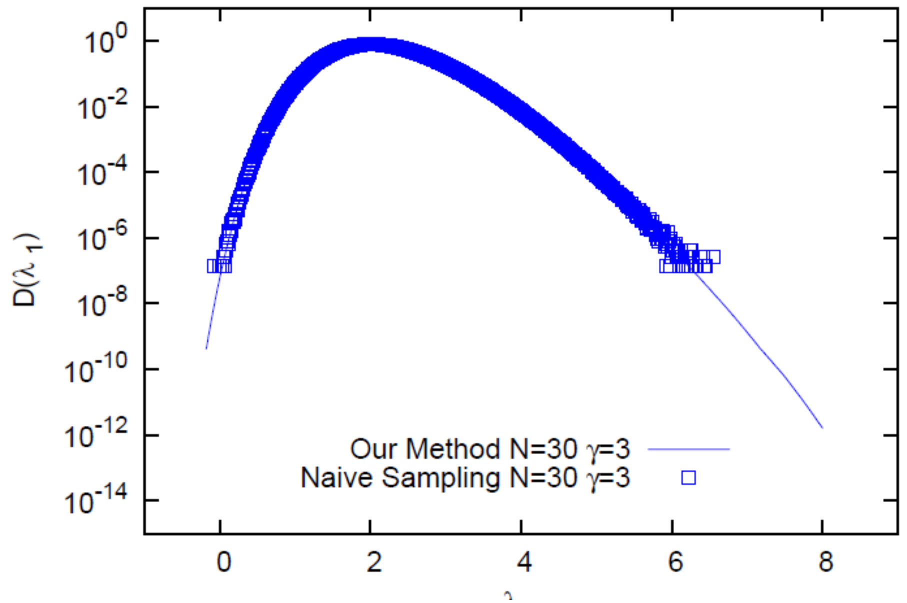


max eigenvalue (scaled)

N=10, 20,30

Sparse Random Matrices

log
prob.

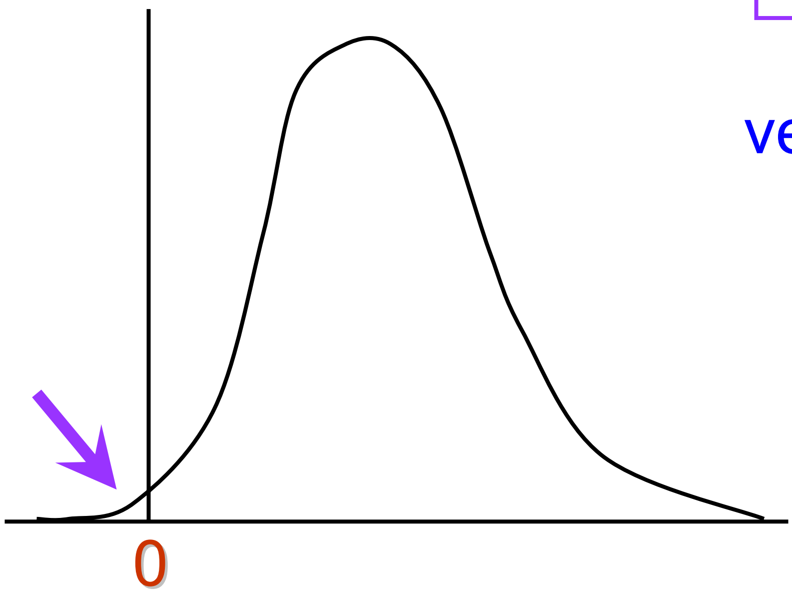


max eigenvalue

FIG. 5. Density $D(\lambda_1)$ in a case of sparse random matrices. The first definition is applied; results of the proposed method and the naive random sampling method are compared for $N = 30$ and $\gamma = 3$. The symbol □ appears only in the region where naive random sampling gives nonzero results.

$$\mathrm{Prob}(\lambda_{\mathrm{max}}(x) < 0)$$

= Prob. that
all eigenvalues are negative



very small when N becomes large

important in
cosmology, glass theory, ecology, statistical testing
…

GOE

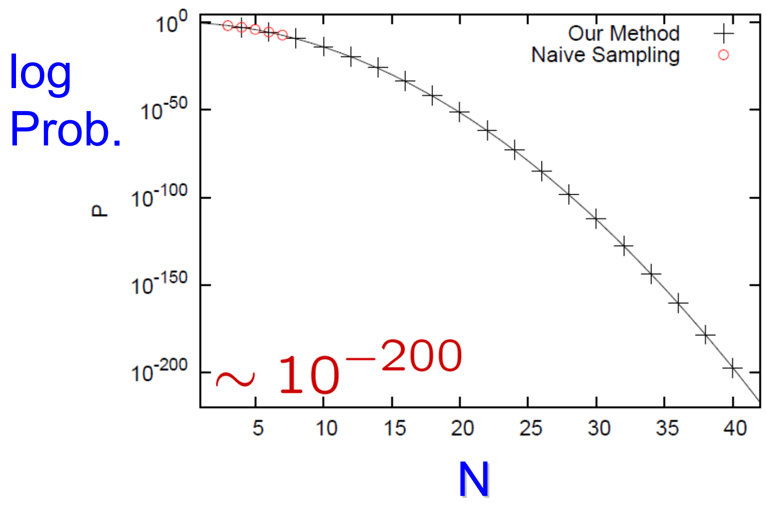


FIG. 3. Probability $P(\forall i, \lambda_i < 0)$ for GOE versus size N of the matrices. The results of the proposed method (+) and the naive random sampling method ($\odot$) are shown. The results of naive random sampling are available only in the region $N \leq 7$. The curve indicates a quadratic fit to the results with the Coulomb gas representation given in Dean and Majumdar[20].

UNIFORM

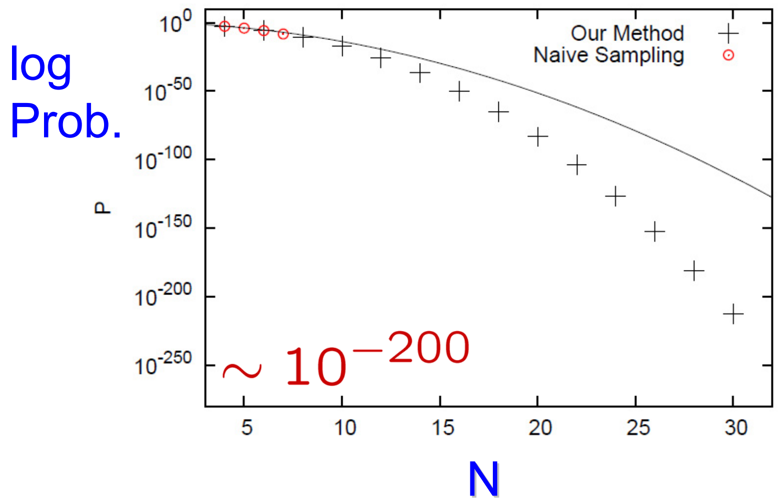


FIG. 4. Probability $P(\forall i, \lambda_i < 0)$ for an ensemble of matrices whose components are uniformly distributed. The horizontal axis corresponds to the size N of the matrices. The results of the proposed method (+) and the naive random sampling method ($\odot$) are shown. The results of naive random sampling are shown for $4 \leq N \leq 7$. The curve indicates the probability for the GOE with the same variance.

Curves: theoretical estimate (Coulomb gas) for GOE

Dean and Majumdar (2006, 2008)

Sparse

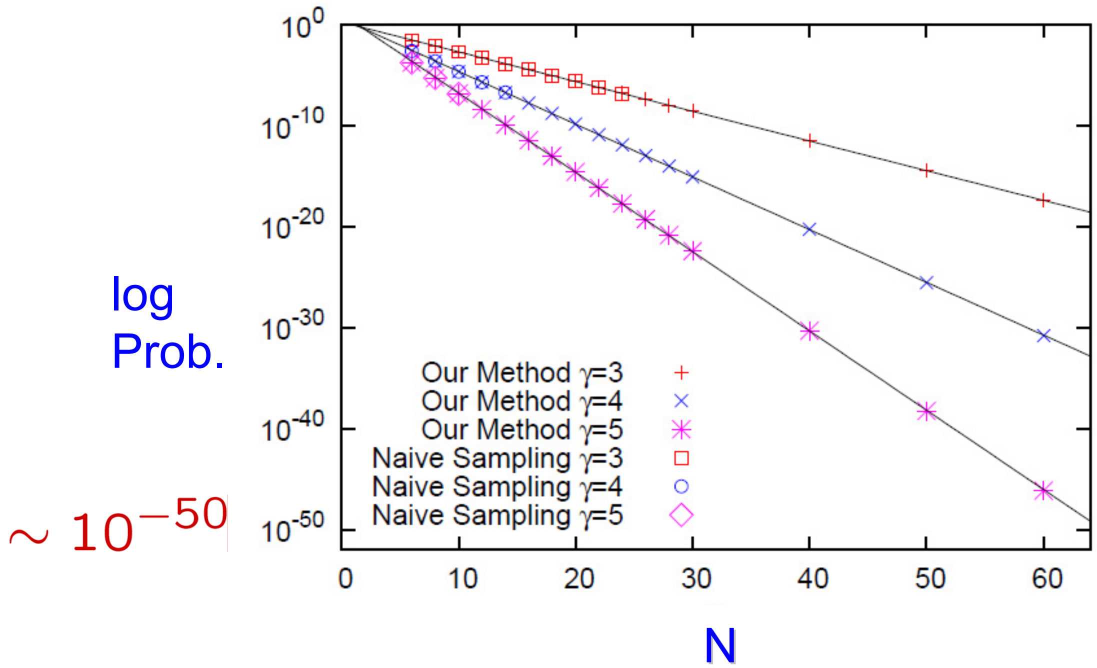


FIG. 6. Probabilities $P(\forall i, \lambda_i < 0)$ for an ensemble of sparse random matrices estimated by the proposed method. The first definition is applied; the results with $\gamma = 3$, 4, and 5 versus size N of the matrices are shown. The lines show linear fits of the data.

Example.2

Random Graphs

$\xi = \lambda_2$

adjacency matrix A_{ij}

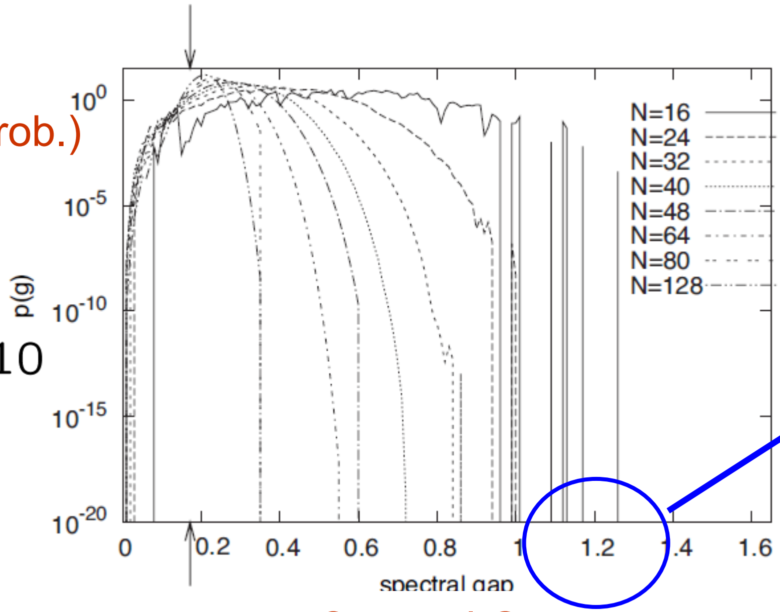
$$A_{ij} = \begin{cases} 1 & i \text{ and } j \text{ are connected,} \\ 0 & \text{otherwise.} \end{cases}$$

The second largest eig. λ_2

Spectral Gap

log (prob.)

10^{-10}



Spectral Gap

Ramanujan Graphs

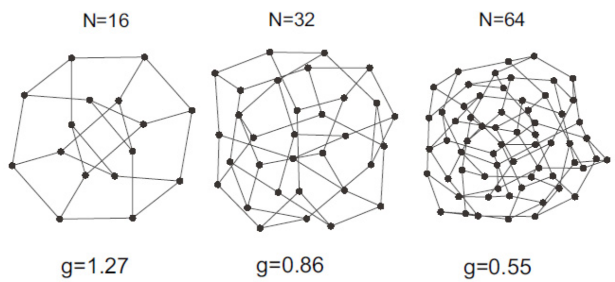


Figure 1: 3-regular graphs with the largest spectral gap found in the simulation.

N Saito and Y Iba, arXiv:1003.1023 (2010)

[optimization only] Donetti et al. (2005,2006)

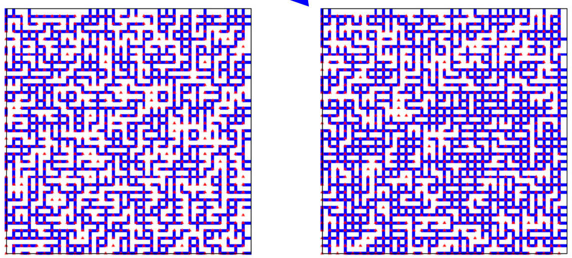
Example.3

$\xi = \chi^{-1}$

Griffiths Singularity in Random Magnets

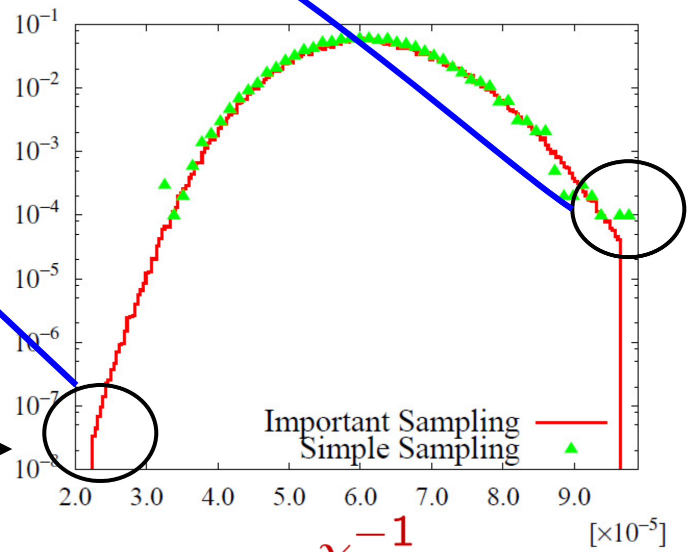
dynamic Monte Carlo sampling of quenced randomness

bond diluted Ising



sampling bond configrations

log (prob.)



10^{-8} →

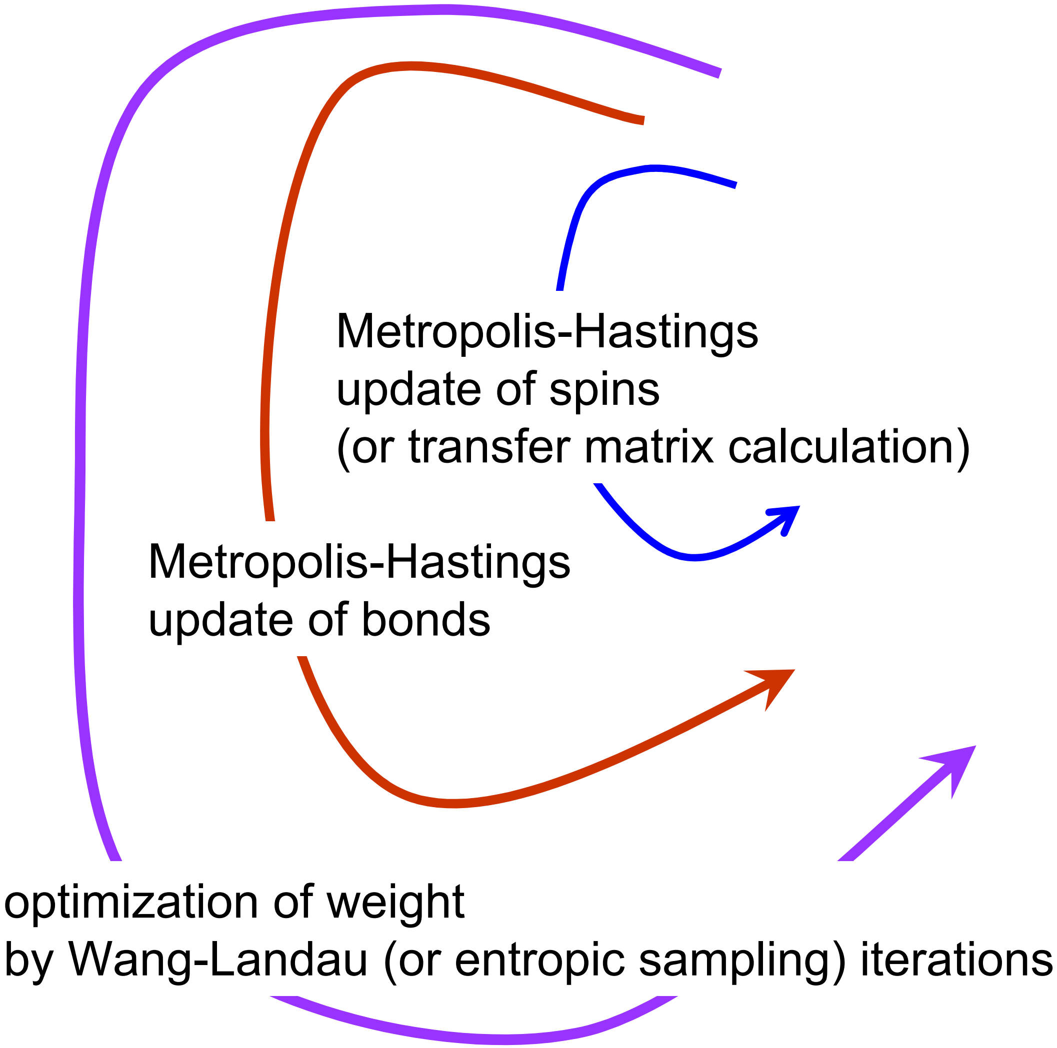
χ^{-1}

Figure 5. Distribution of the inverse susceptibility of the two-dimensional bond-diluted Ising model with $p = 0.6$ for $L = 32$ and $T/J = 1.5$. The solid line represents $P(\chi^{-1})$ obtained by the importance-sampling MC, and the triangles are that by the simple sampling.

[bond diluted]
K Hukushima and Y Iba
arXiv:0711.0870 (2007)

[spin glass+LeeYang zero]
Matsuda et al. (2008)

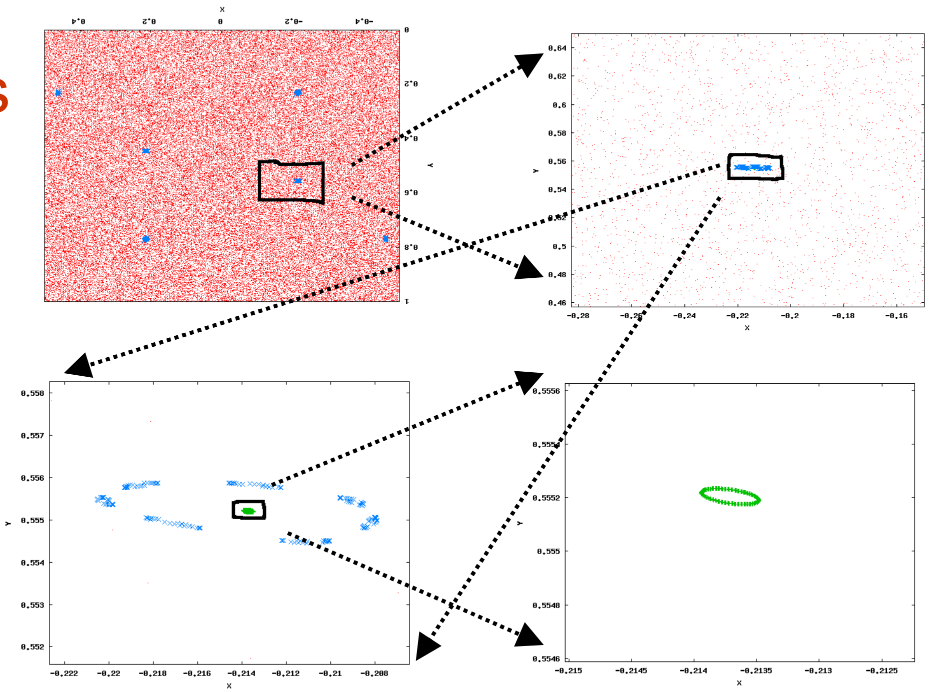
Griffiths Singularity in Random Magnets



Example 4 $\xi =$ "chaoticity"

Chaotic Dynamical Systems

Search for rare initial conditions
that gives rare trajectories



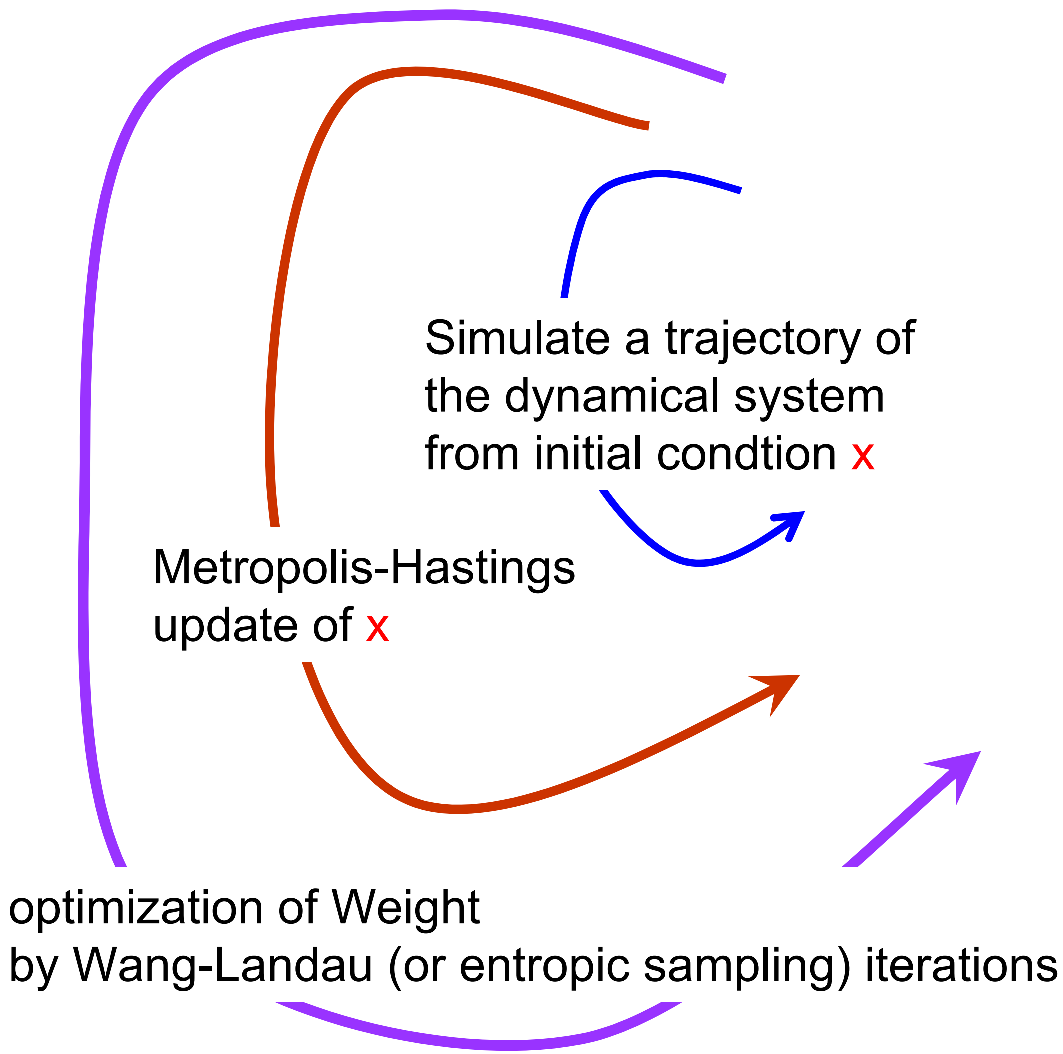
e.g., Coupled Standard Map

$$
\begin{aligned}
u_{n+1} &= u_n - \frac{K}{2\pi}\sin(2\pi v_n) + \frac{k}{2\pi}\sin(2\pi(v_n + y_n)) \\
v_{n+1} &= v_n + u_{n+1} \\
x_{n+1} &= x_n - \frac{K}{2\pi}\sin(2\pi y_n) + \frac{k}{2\pi}\sin(2\pi(v_n + y_n)) \\
y_{n+1} &= y_n + x_{n+1}
\end{aligned}
$$

A Kitajima and Y Iba,
arXiv:1003.2013 (2010)

Probability of regular trajectory (fragments)
embedded in chaotic sea

Dynamical System



related studies

*sampling quenched randomness (zero temperature)
canonical weight: Hartmann (2002)
other guiding function: Koerner et al.(2006) (and more)

*rare event sampling using multicanonical
optical communicatons: Holzloehner and Menyuk (2003)
“growth ratio” of matrices: Driscoll and Maki (2007)

*transition path sampling: Chandler’s group (around 1998~)

Summary & Conclusion

Multicanonical algorithm can be
a powerful tool for sampling rare events.

Applications to rare events in random matrices.
Dynamic sampling of quenched randomeness.

Algorithms provide bridges of
different fields of science and engineering

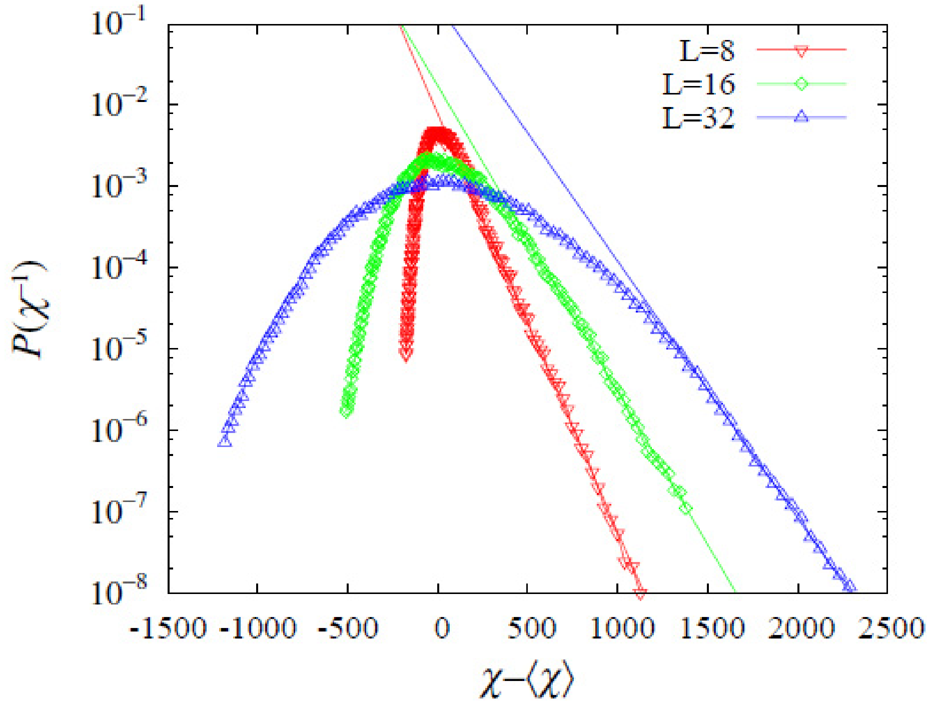


Figure 6. Distribution functions of the inverse susceptibility as a function of $1/\chi^{-1}$ of the two-dimensional bond-diluted Ising model with $p = 0.6$ and $T/J = 2.0$. The system sizes are $N = 8^2(\bigtriangledown)$, $16^2(\Diamond)$ and $32^2(\bigtriangleup)$. The straight lines are the fitting result of the exponential function $p(\chi^{-1}) = B\exp(-C/\chi^{-1})$ with B and C being fitting parameters.

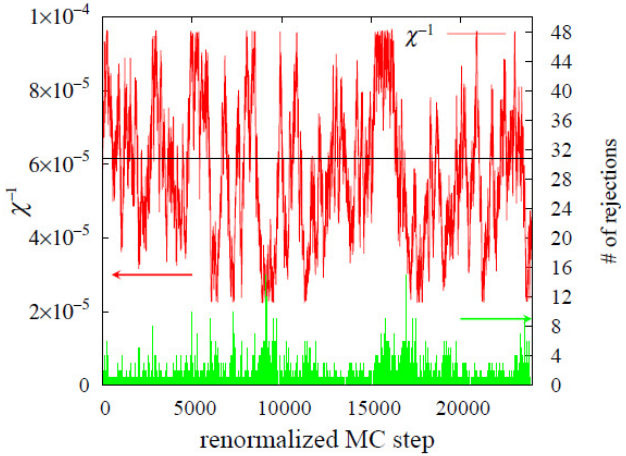


Figure 3. A Monte Carlo trajectory of the value of χ^{-1} of the two-dimensional bond-diluted Ising model for $L = 32$, $p = 0.6$ and $T/J = 1.5$. The horizontal axis means renormalized MC steps which are incremented by one when a new value of χ^{-1} is accepted. The value of χ^{-1} as a function of the MC step is represented by the straight line and the number of rejections at the MC step is given by bar chart.

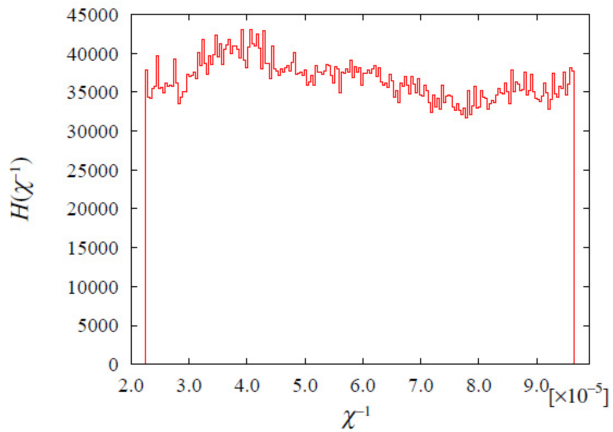


Figure 4. Histogram of the inverse susceptibility obtained by an importance sampling algorithm of the two-dimensional bond-diluted Ising model. The parameters used in the simulation is the same as those in figure 3.

Rare events in Dynamical Systems

Deterministic Chaos

Doll et al. (1994), Kurchan et al. (2005)

Sasa, Hayashi, Kawasaki .. (2005 ~)

Stagger and Step Method Sweet, Nusse, and Yorke (2001)

(Mostly) Stochastic Dynamics

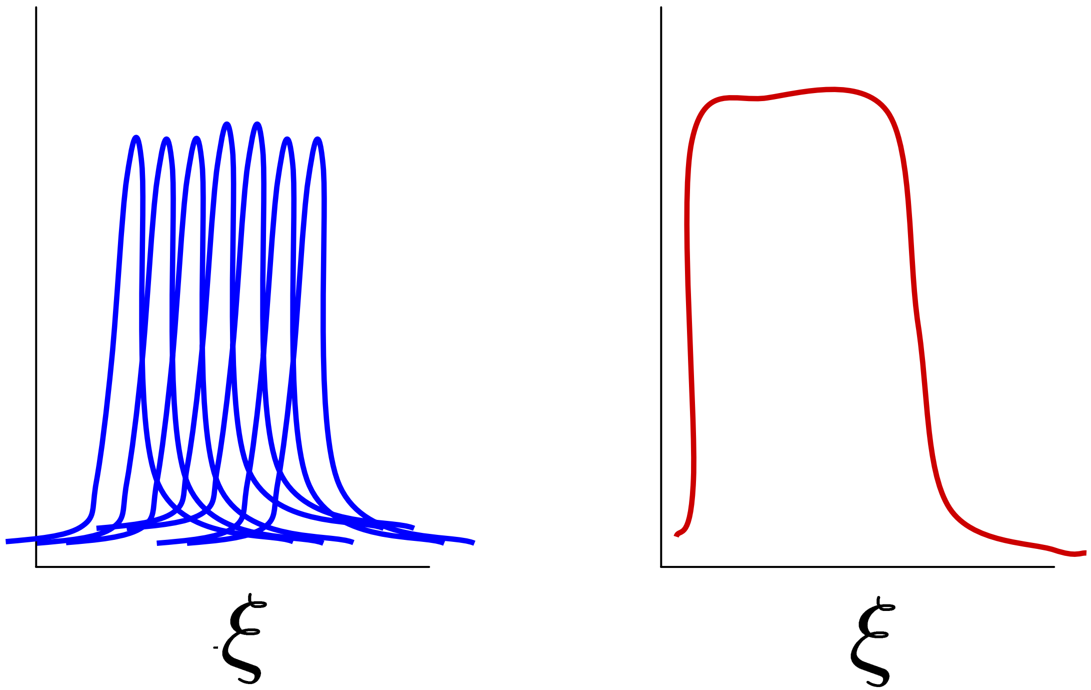
Chandler Group

Frenkel et al.

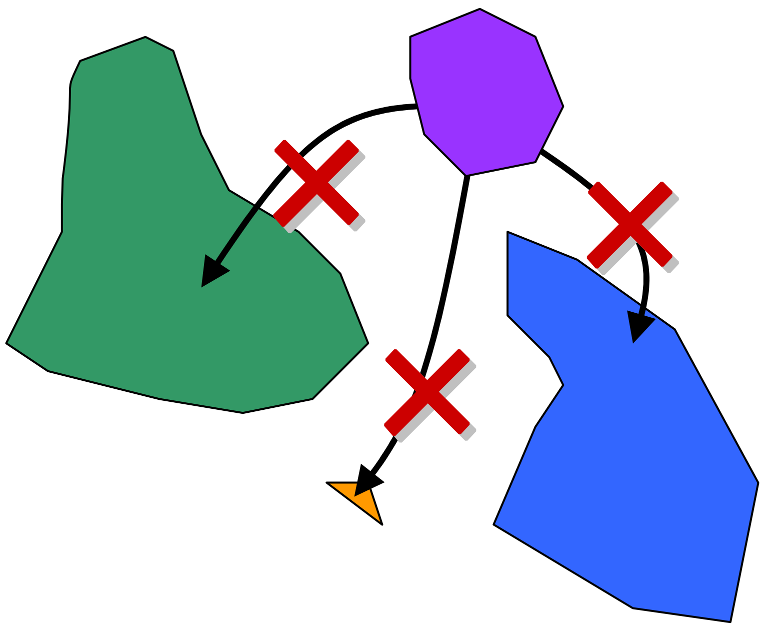
and more

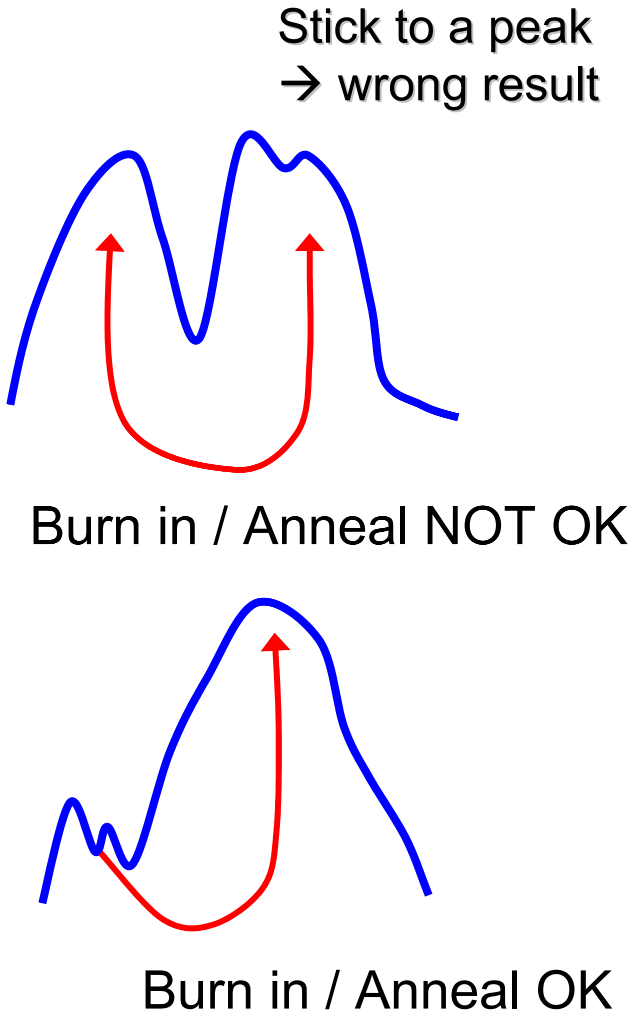
Transition Path Sampling

Exponential,Gibbs / Multicanonical

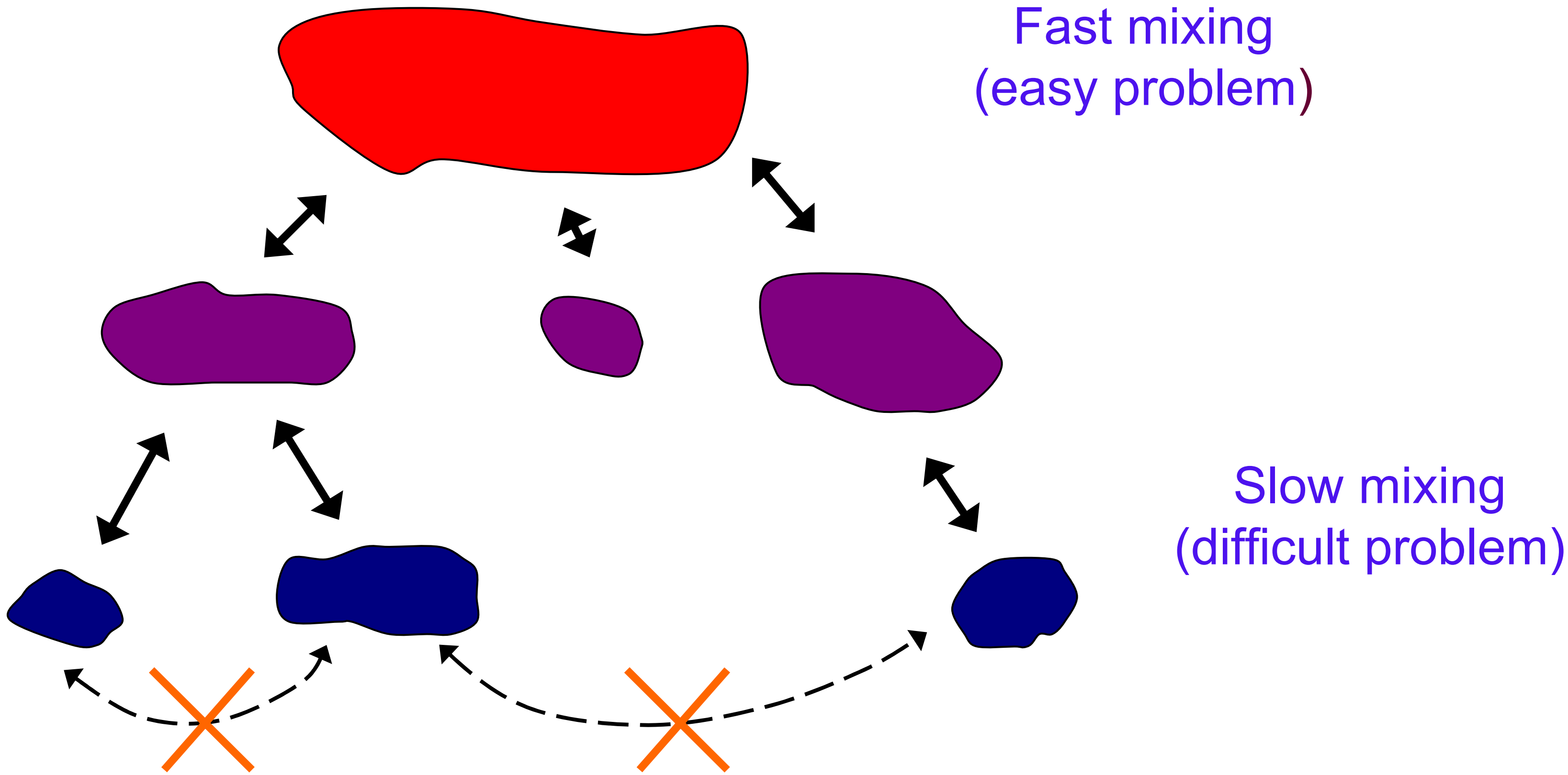


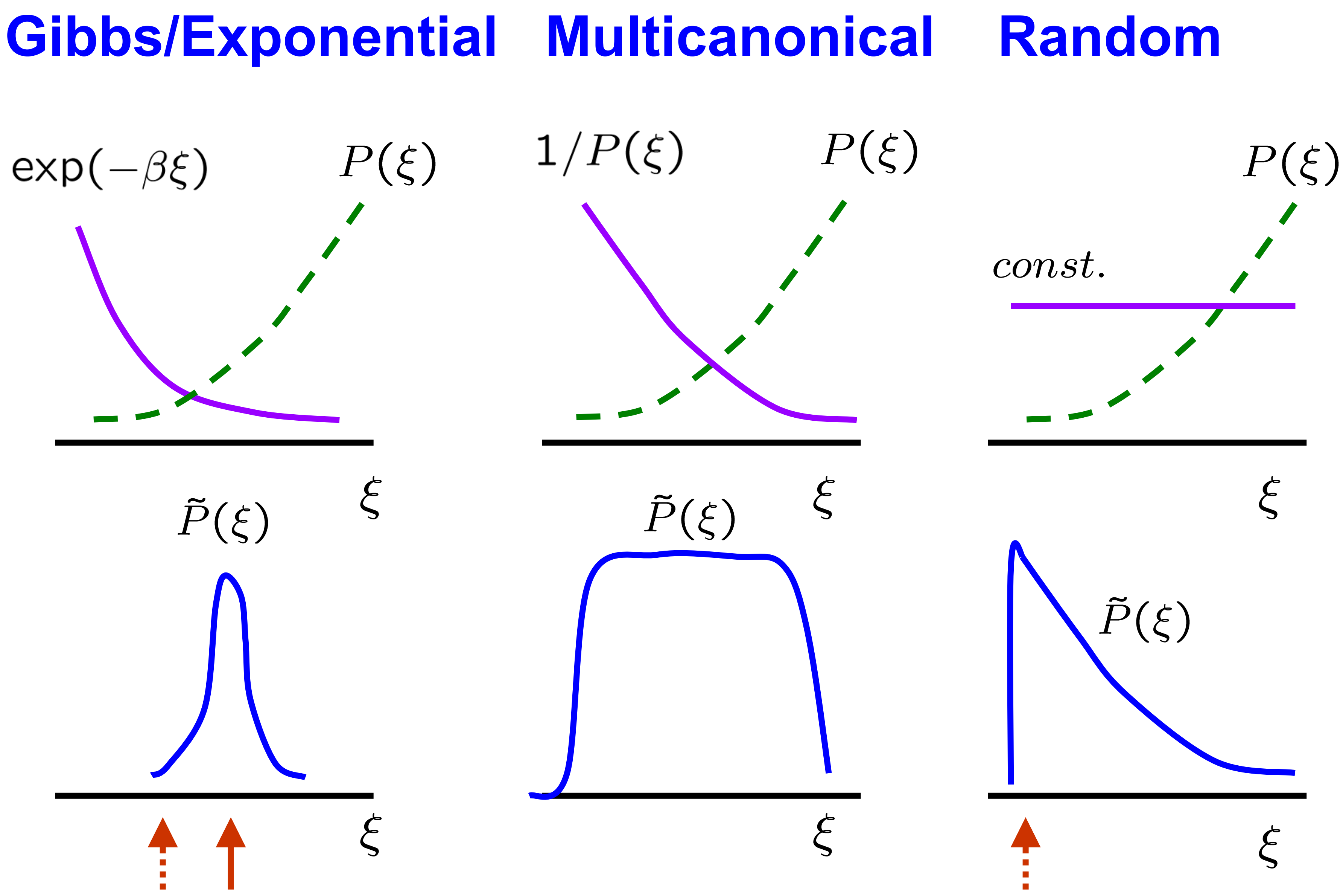
Slow mixing
Multimodality Issues





Bridge




Gibbs/Exponential
Multicanonical
Random
$\exp(-\beta\xi)$
$P(\xi)$
$1/P(\xi)$
$P(\xi)$
$P(\xi)$
$const.$
ξ
ξ
ξ
$\tilde{P}(\xi)$
$\tilde{P}(\xi)$
$\tilde{P}(\xi)$
ξ
ξ
ξ