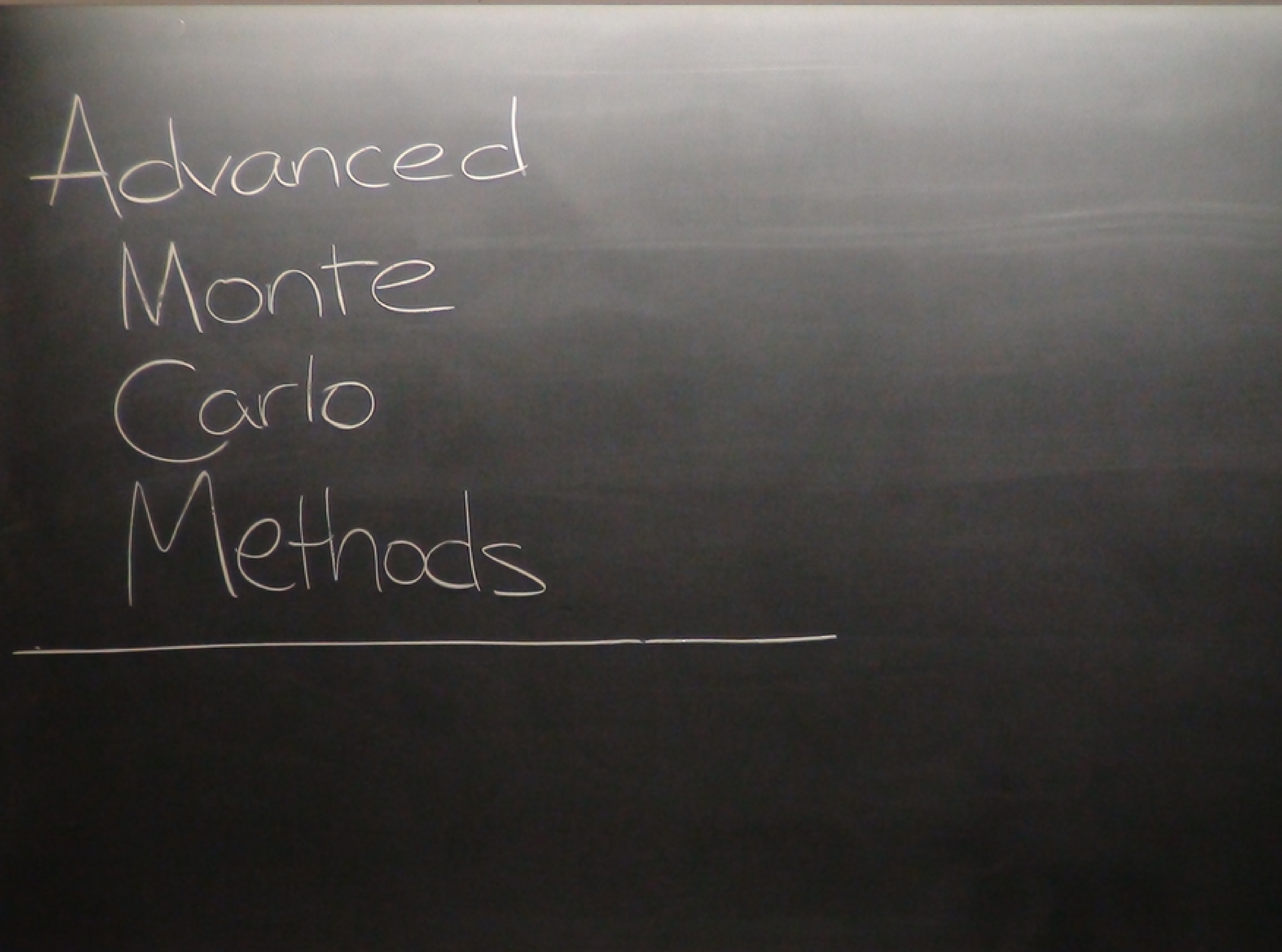
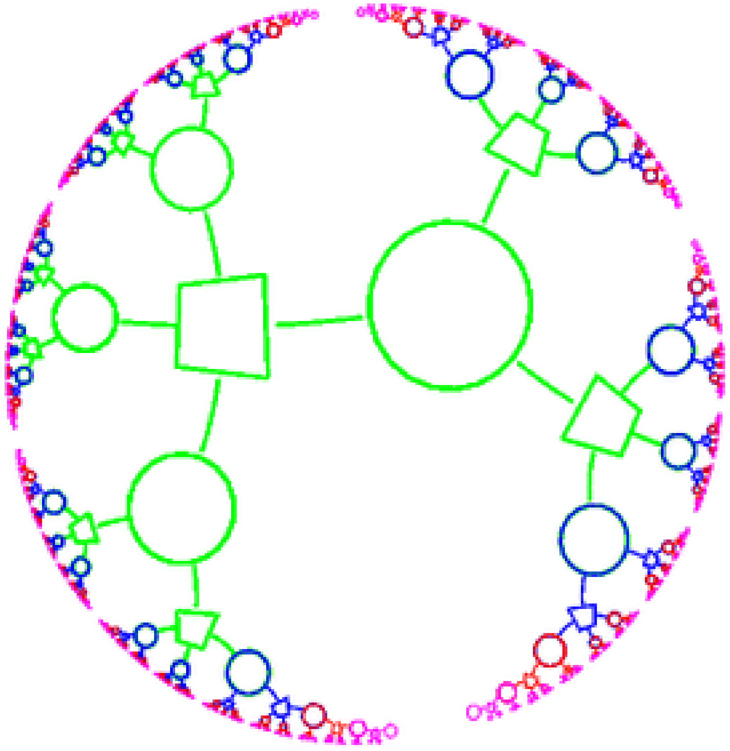


Advanced Monte Carlo Methods



Information theory, pattern recognition, and neural networks

- Source coding (Data compression)
- Noisy-channel coding
- Inference + probabilistic methods
  - 9-10 Inference
  - 11 Clustering
  - 12 Monte Carlo methods
  - 13 Advanced Monte Carlo methods
  - 14 Variational methods

www.inference.phy.cam.ac.uk/itprnn/

www.inference.phy.cam.ac.uk/itila/

Overview

- Data compression
- Noisy-channel coding
    - Chs 1-6, 8-10, 14
- Inference, data modelling
    - clustering, pattern recognition
        - Chs 20, 22
- Probability toolbox
    - Monte Carlo methods
        - Ch 29, 30, 32
    - Variational methods
        - Ch 33
- Neural networks
    - Chs 38, 39, (& perhaps 41, 44), 42
- State-of-the-art error-correcting codes

Recommended exercises

- 29.14, 33.5, 33.7, 27.1, 22.11, 39.4, 39.5
- Handouts 2, 3 (on website)
- The 5 cards magic trick (15.6)

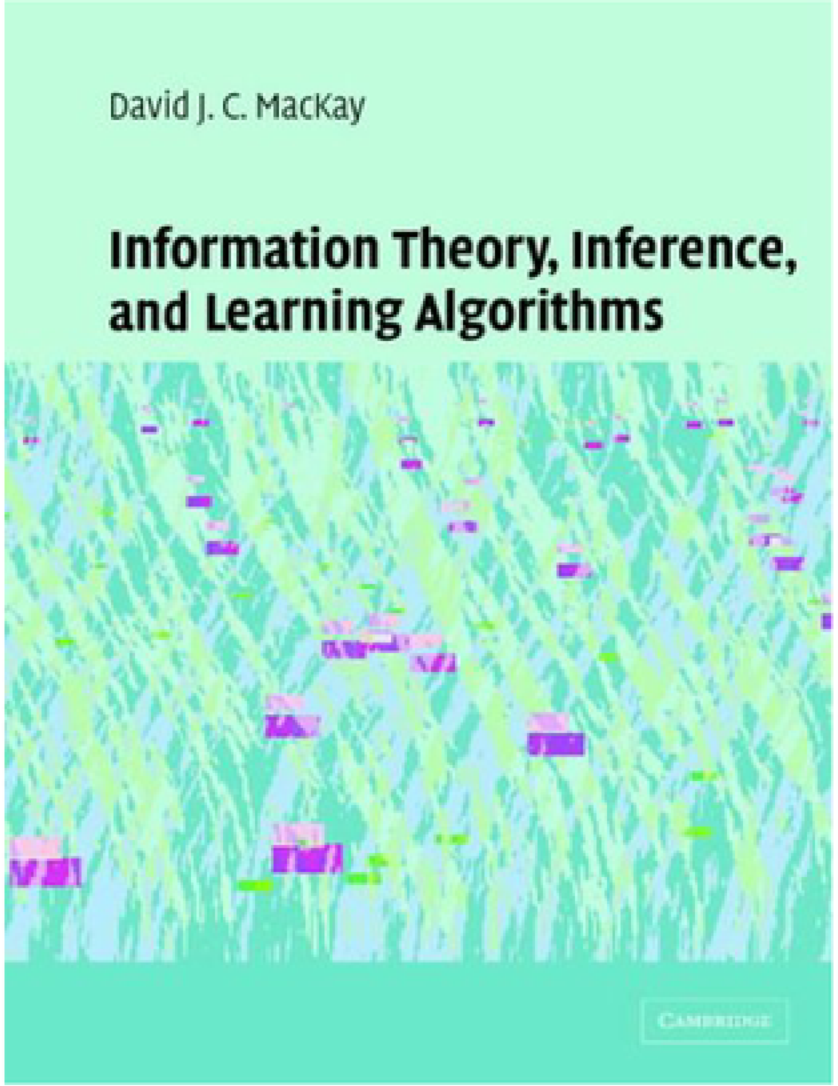
$8\diamondsuit$, $2\heartsuit$, $10\clubsuit$, $9\diamondsuit$

Additional reading

- Laplace's method (Ch 27)
- Ising models (Ch 31)

The course

www.inference.phy.cam.ac.uk/itprnn/



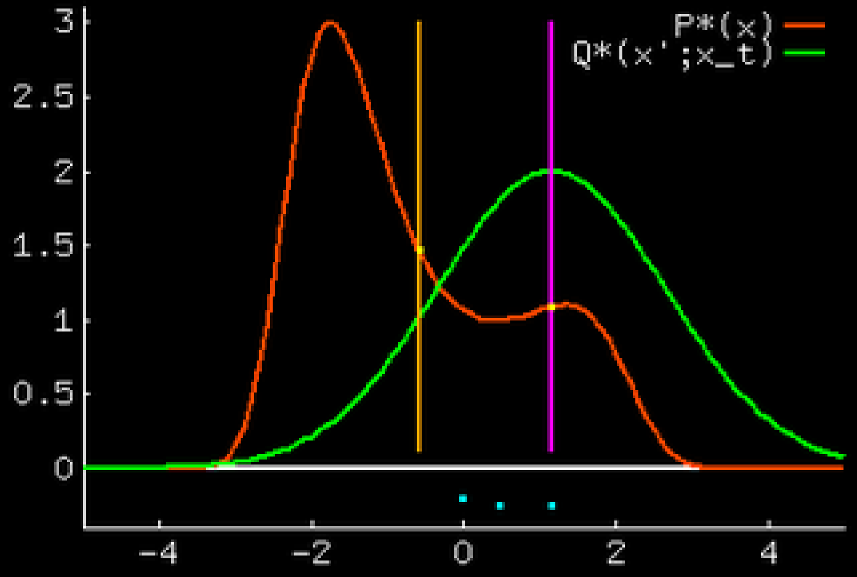
The book

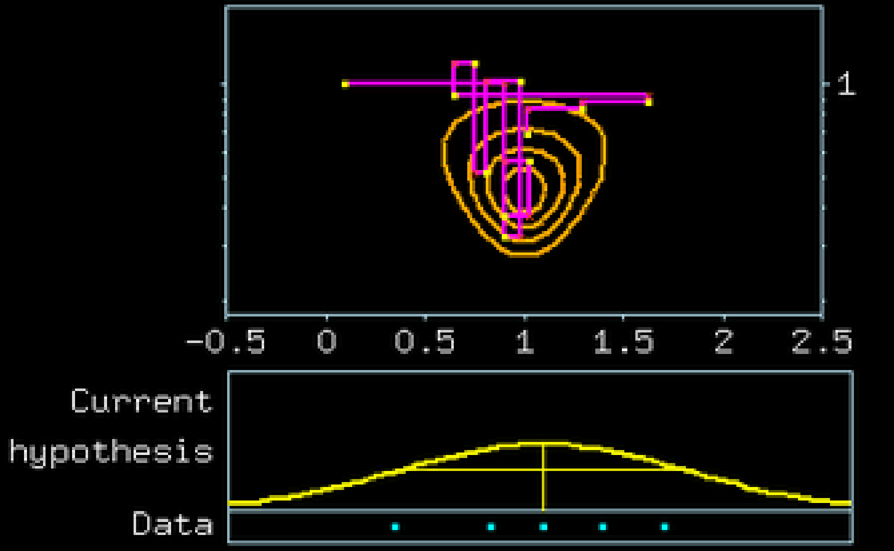
www.inference.phy.cam.ac.uk/itila/

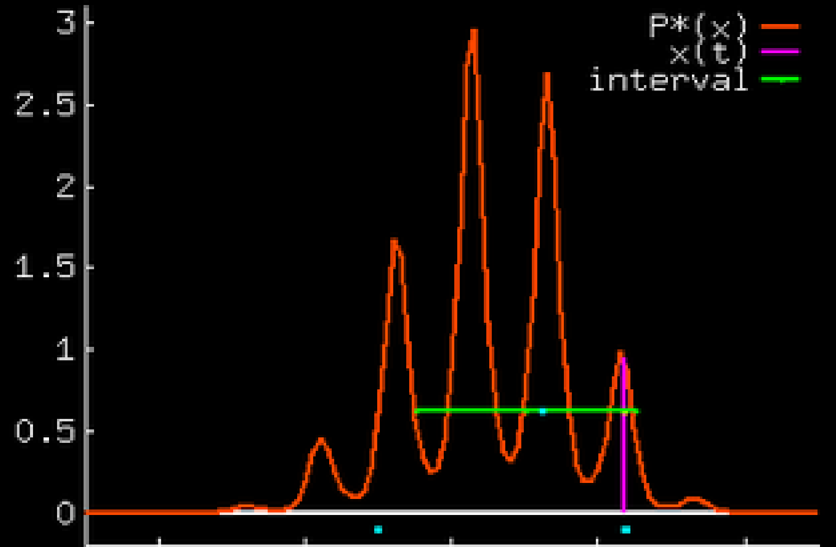
Monte Carlo methods

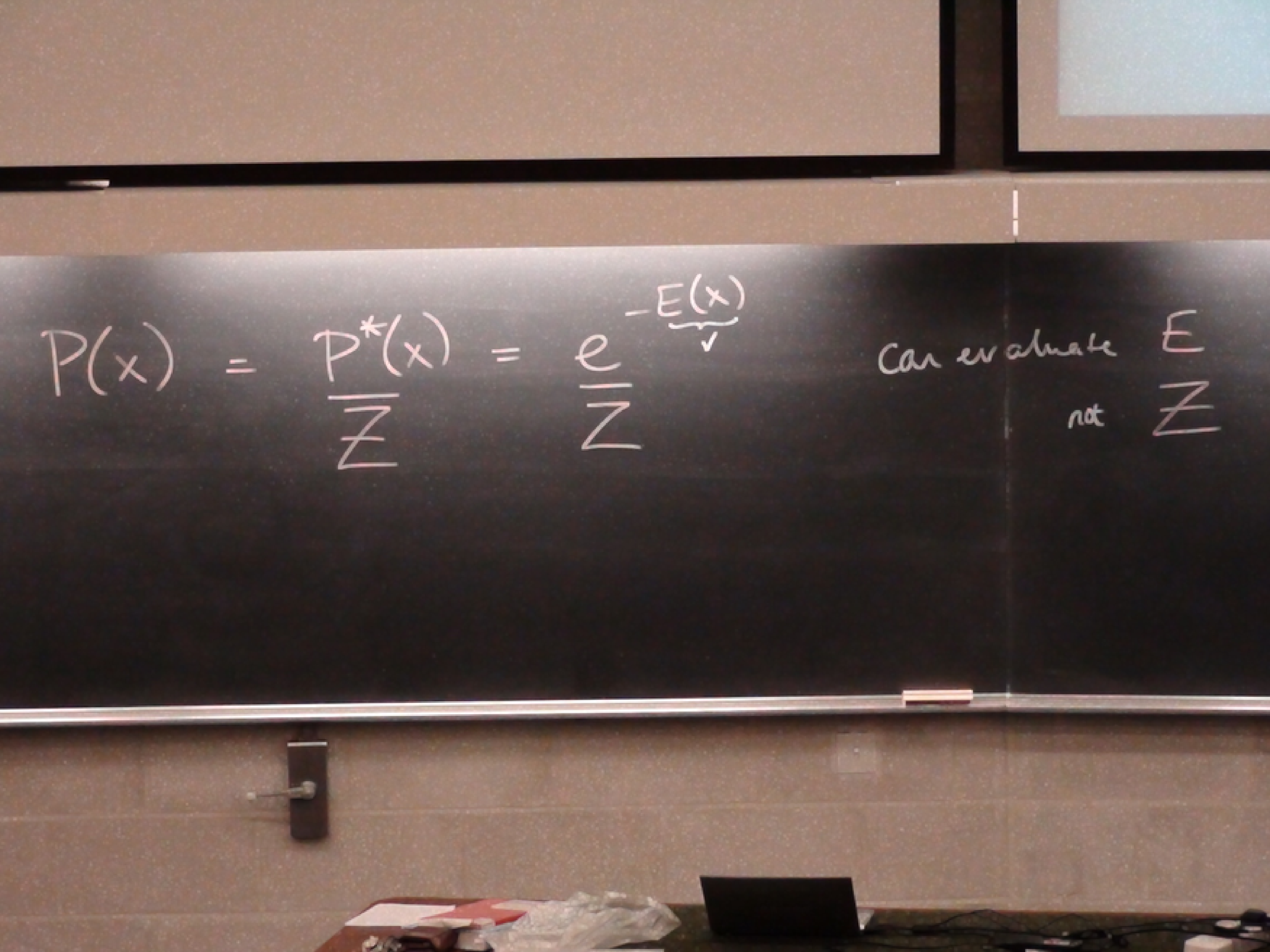
- Simple Monte Carlo methods
  - Importance sampling
  - Rejection sampling
- Markov-chain Monte Carlo methods
  - Metropolis method
  - Gibbs sampling
  - Slice sampling
- Reducing random-walk behaviour
  - Hamiltonian Monte Carlo
  - Overrelaxation
- Exact sampling

itp/exact/rc RUNME





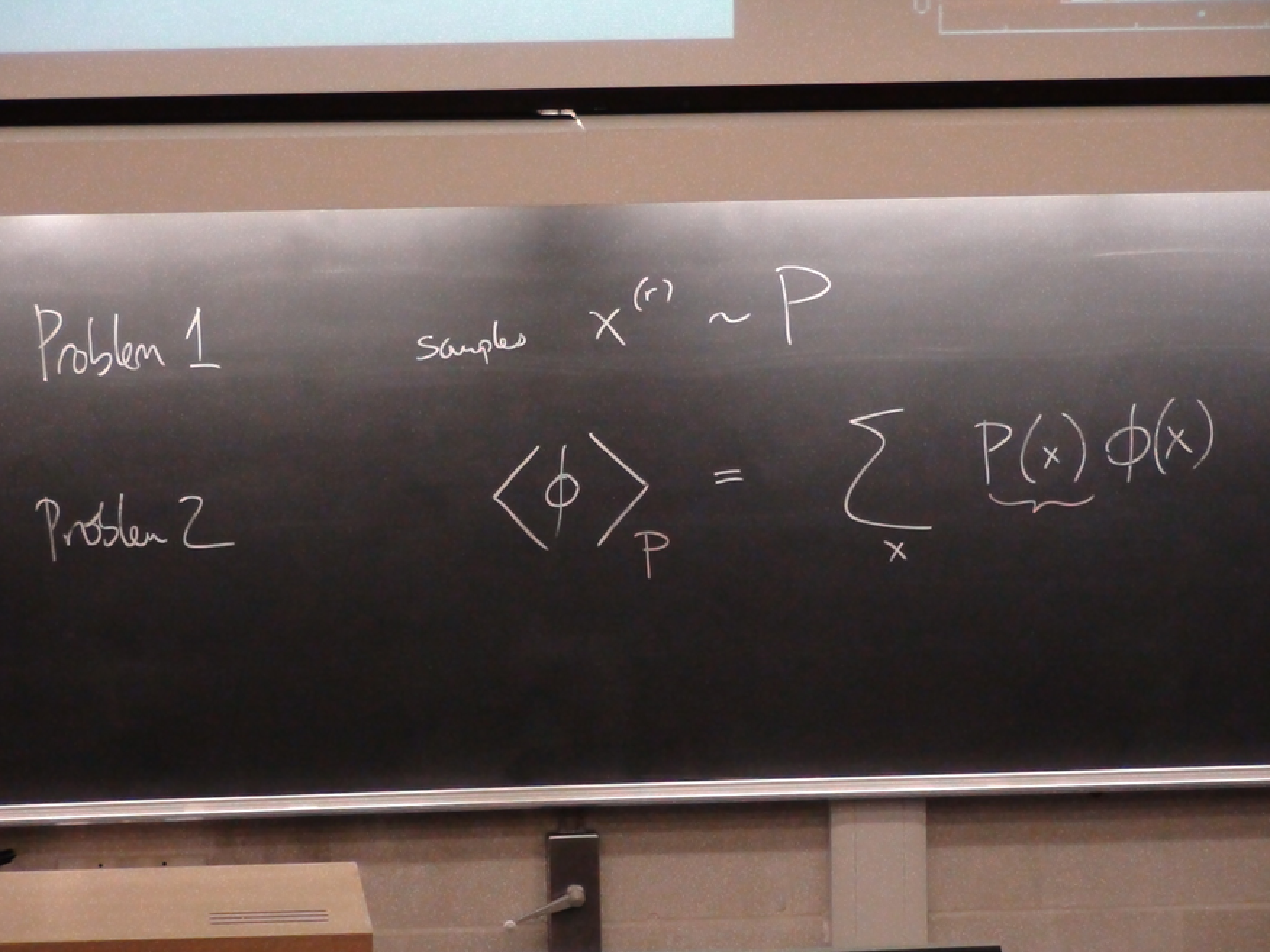




$$P(x) = \frac{P^*(x)}{Z} = \frac{e^{-\underbrace{E(x)}_{\checkmark}}}{Z}$$

can evaluate E

not Z



Problem 1

samples $x^{(r)} \sim P$

Problem 2

$$\langle \phi \rangle_P = \sum_x \underbrace{P(x)} \phi(x)$$

samples $x^{(r)} \sim P$

$$\left\langle \phi \right\rangle_P = \sum_x \underbrace{P(x)}_{} \phi(x)$$

$$\widehat{\langle \phi \rangle} = \frac{\sum \phi\left(x^{(r)}\right)}{R}$$

$$P(x) = \frac{e^{-E(x;\,\underline{\underline{J}})}}{Z(\underline{\underline{J}})}$$

$$E(\underline{x}) = -\frac{1}{2} \sum_{m \neq n} J$$

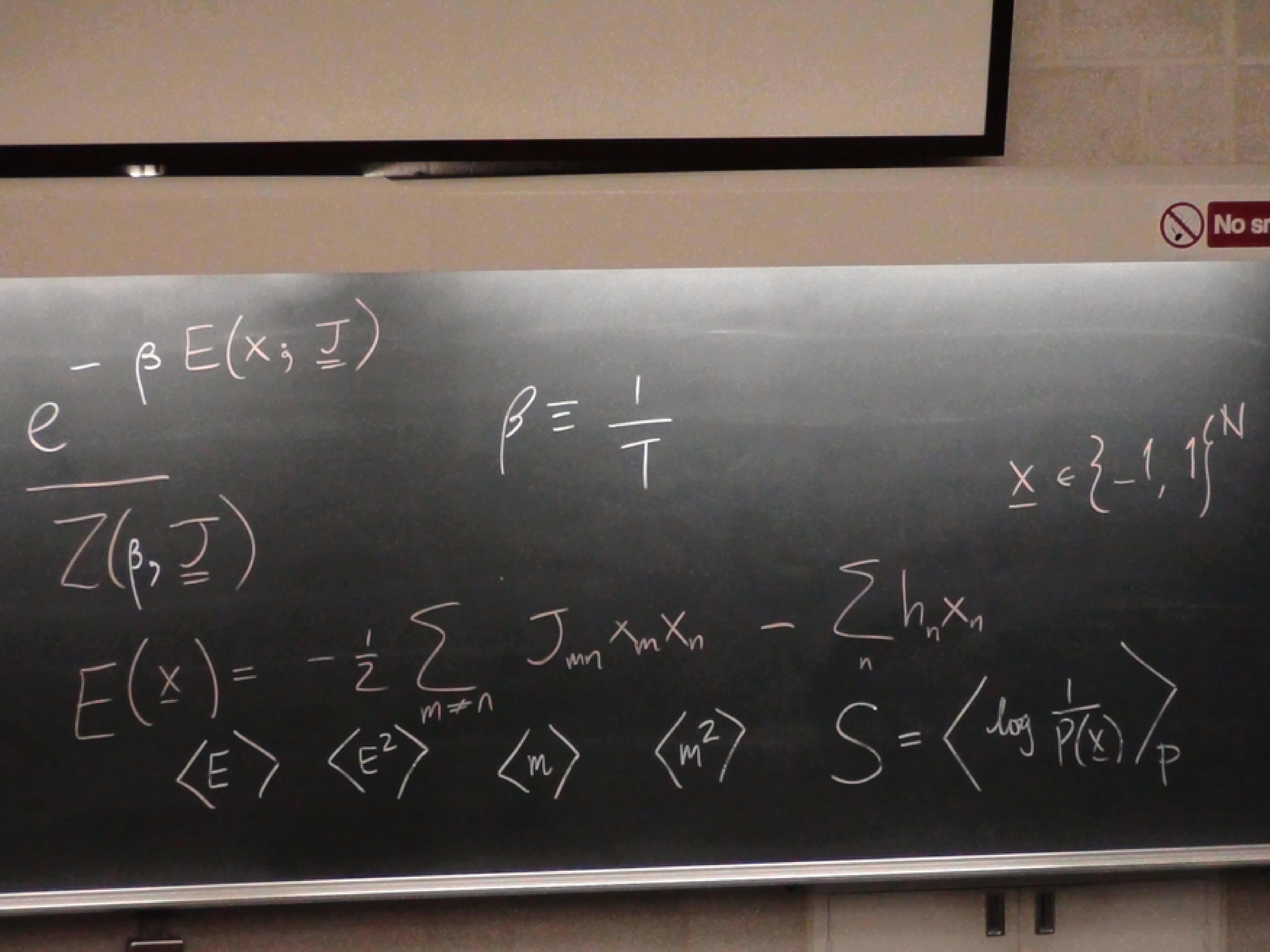
$$E(x ; \underline{\underline{J}})$$

$$Z(\underline{\underline{J}})$$

$$E(\underline{x}) = -\frac{1}{2}\sum_{m \neq n} J_{mn} x_m x_n - \sum_n h_n x_n$$

$$\underline{x} \in \{-1, 1\}^N$$

$$\sum_n h_n x_n$$



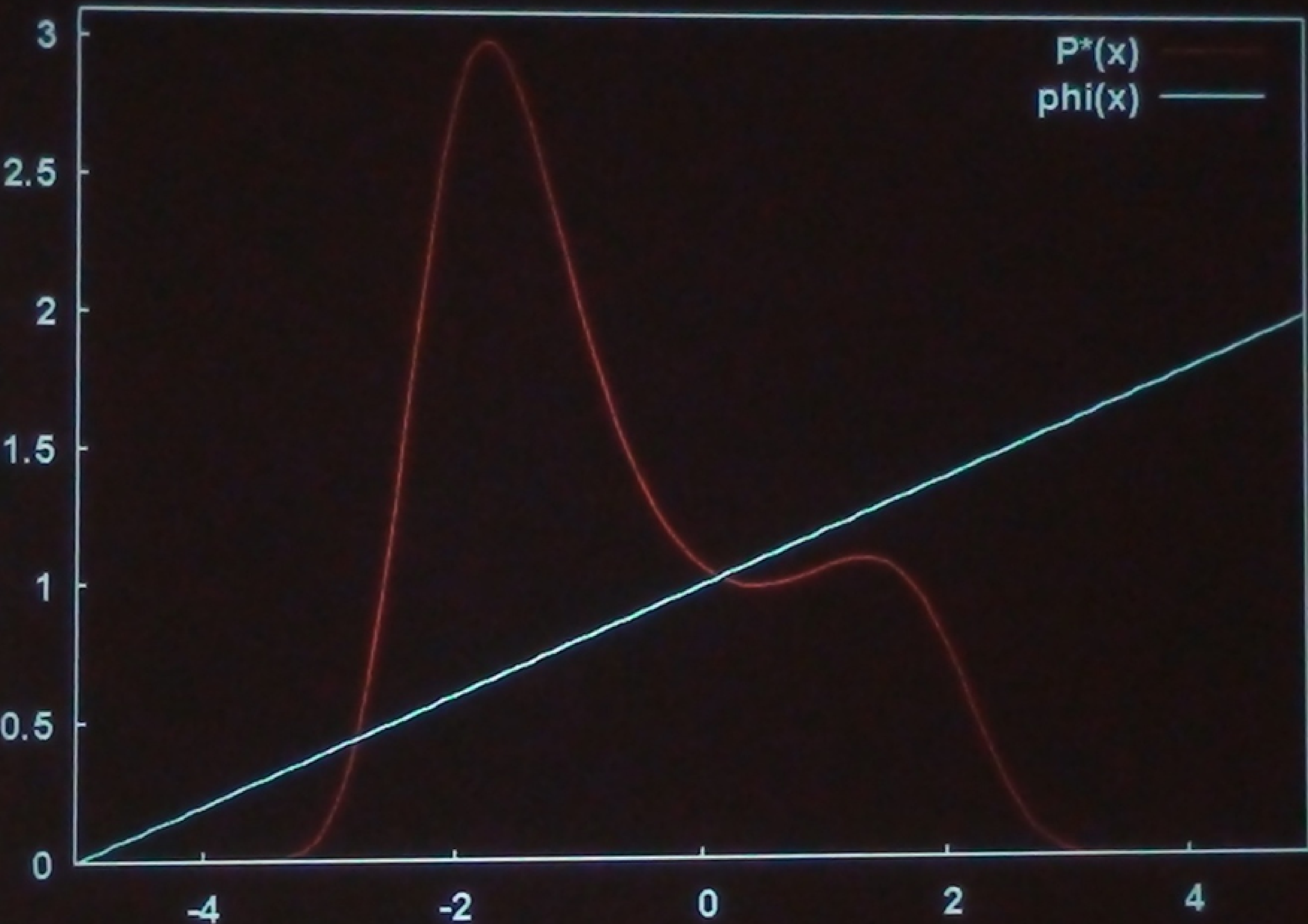
$$\frac{e^{-\beta E(x; \underline{\underline{J}})}}{Z(\beta, \underline{\underline{J}})}$$

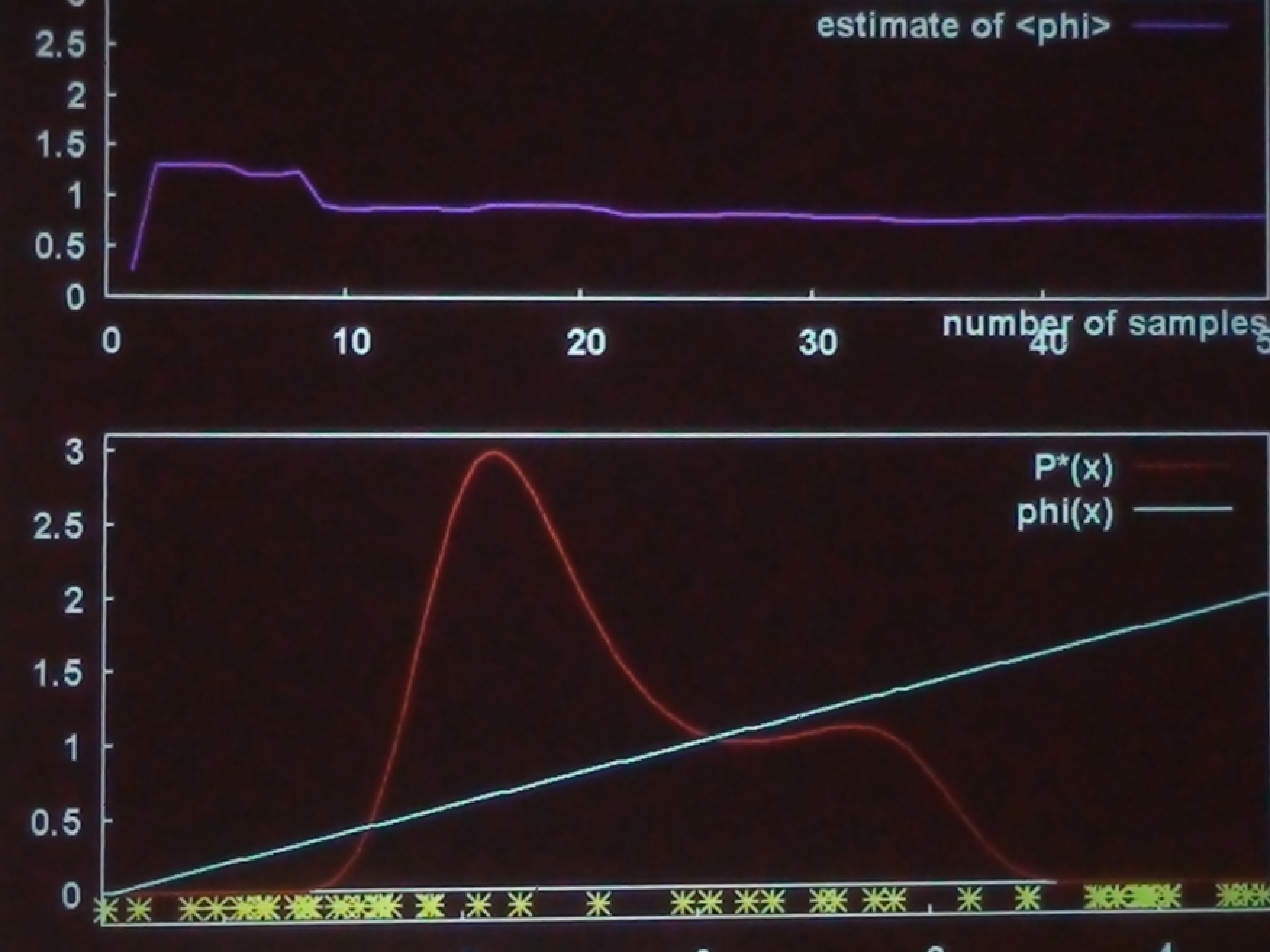
$$\beta \equiv \frac{1}{T}$$

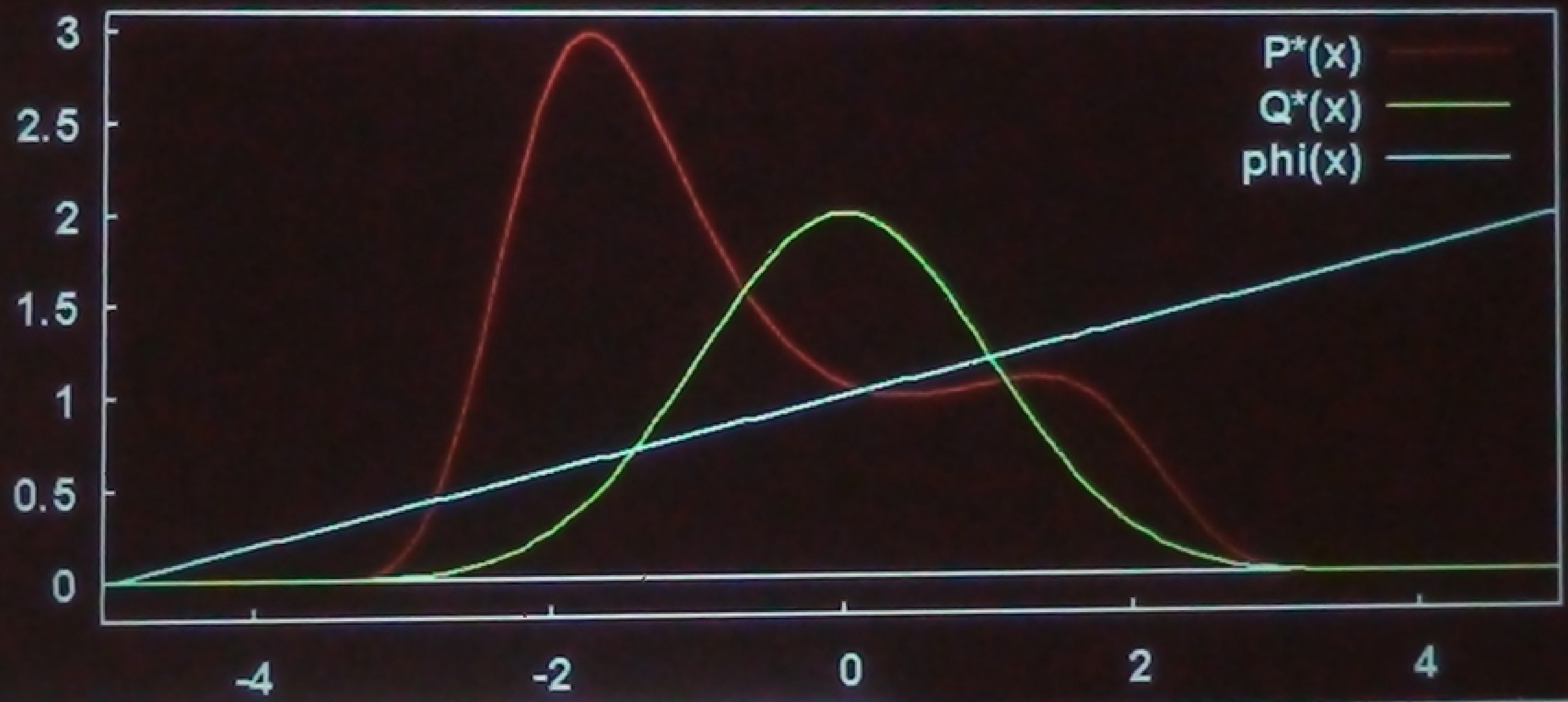
$$\underline{x} \in \{-1, 1\}^N$$

$$E(\underline{x}) = -\frac{1}{2}\sum_{m \neq n} J_{mn} x_m x_n - \sum_n h_n x_n$$

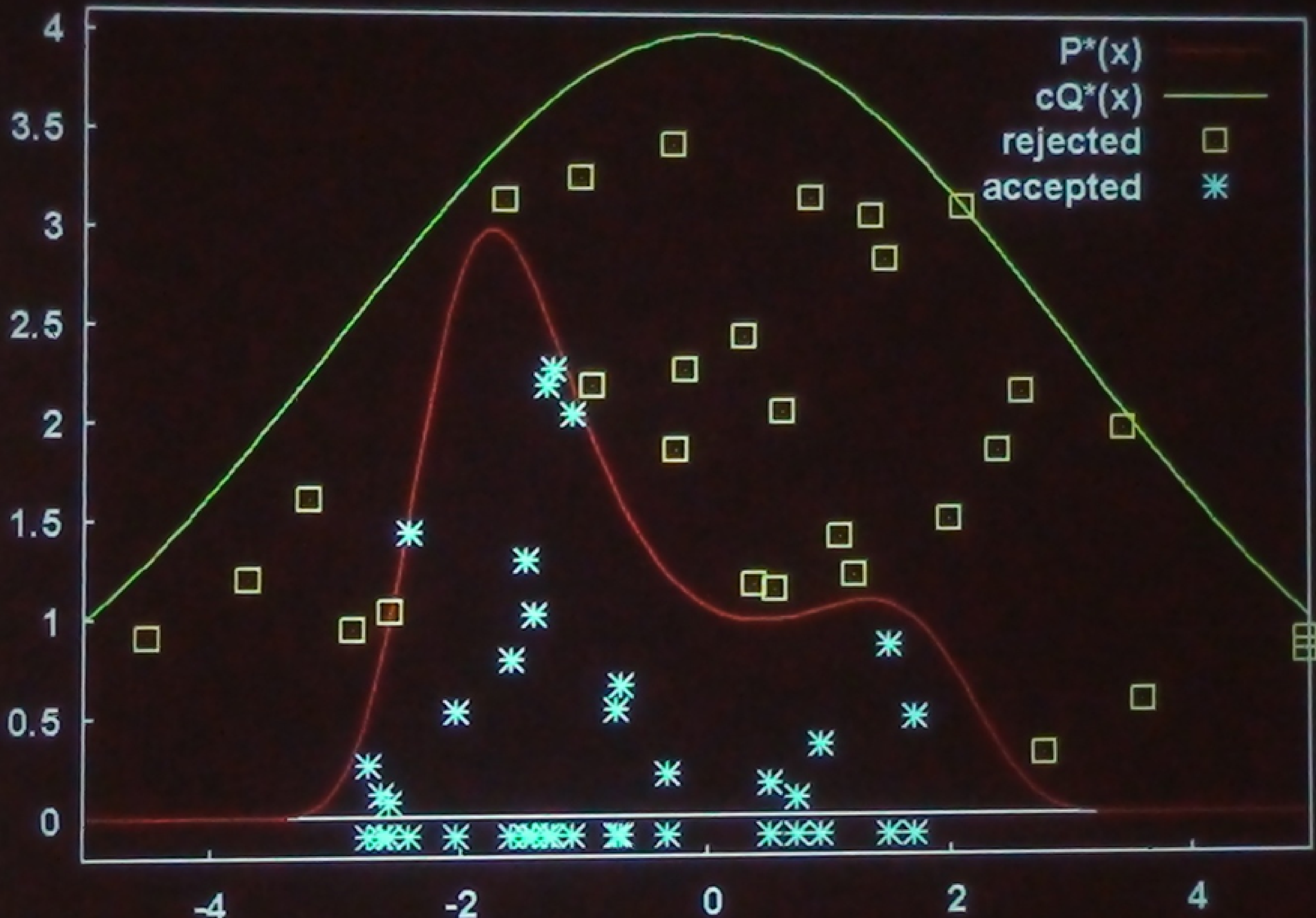
$$\langle E \rangle \qquad \langle E^2 \rangle \qquad \langle m \rangle \qquad \langle m^2 \rangle \qquad S = \left\langle \log \frac{1}{P(\underline{x})} \right\rangle_P$$


P*(x)
phi(x)
3
2.5
2
1.5
1
0.5
0
-4
-2
0
2
4

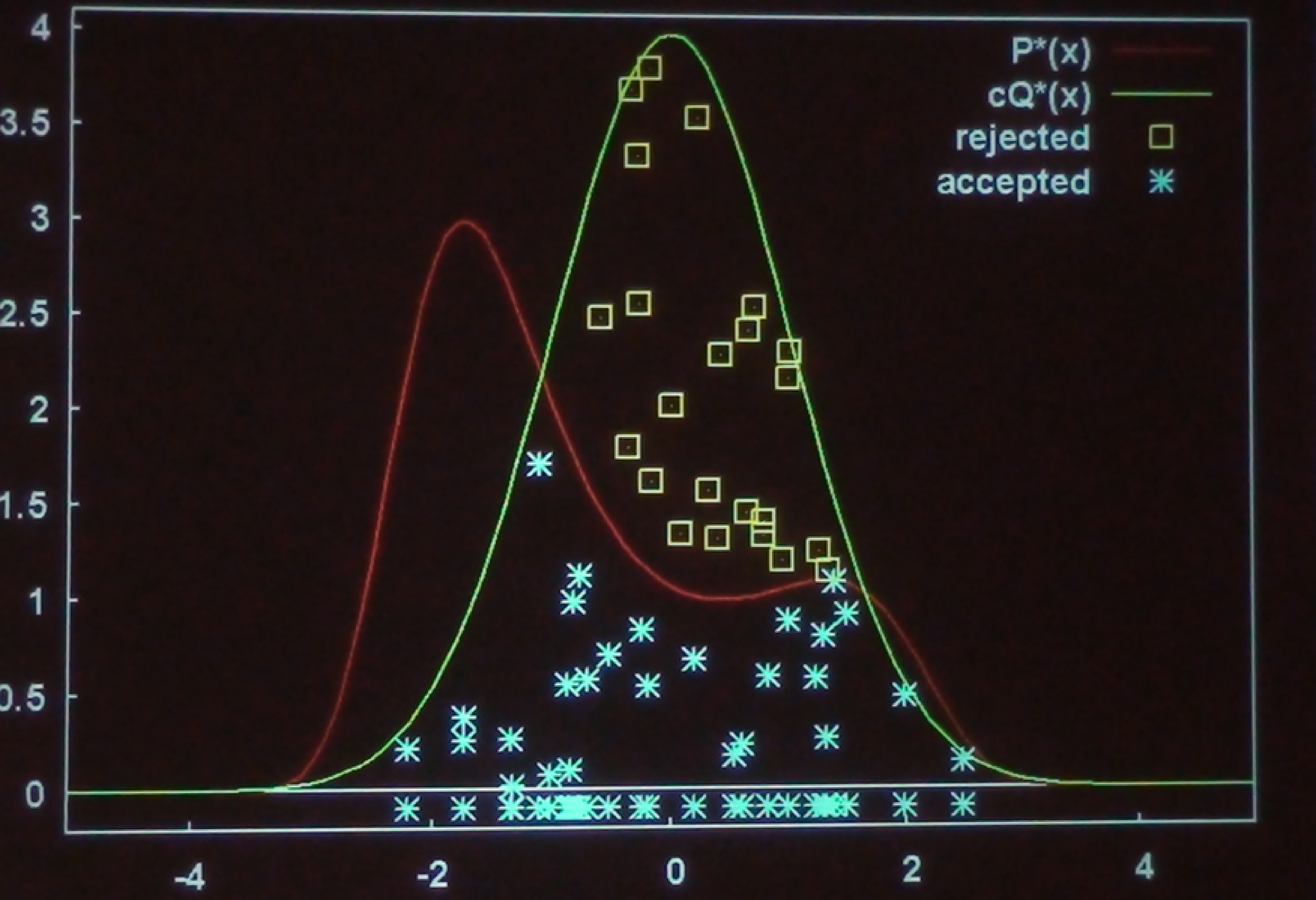

estimate of <phi>
2.5
2
1.5
1
0.5
0
0
10
20
30
40
number of samples
P*(x)
phi(x)
3
2.5
2
1.5
1
0.5
0

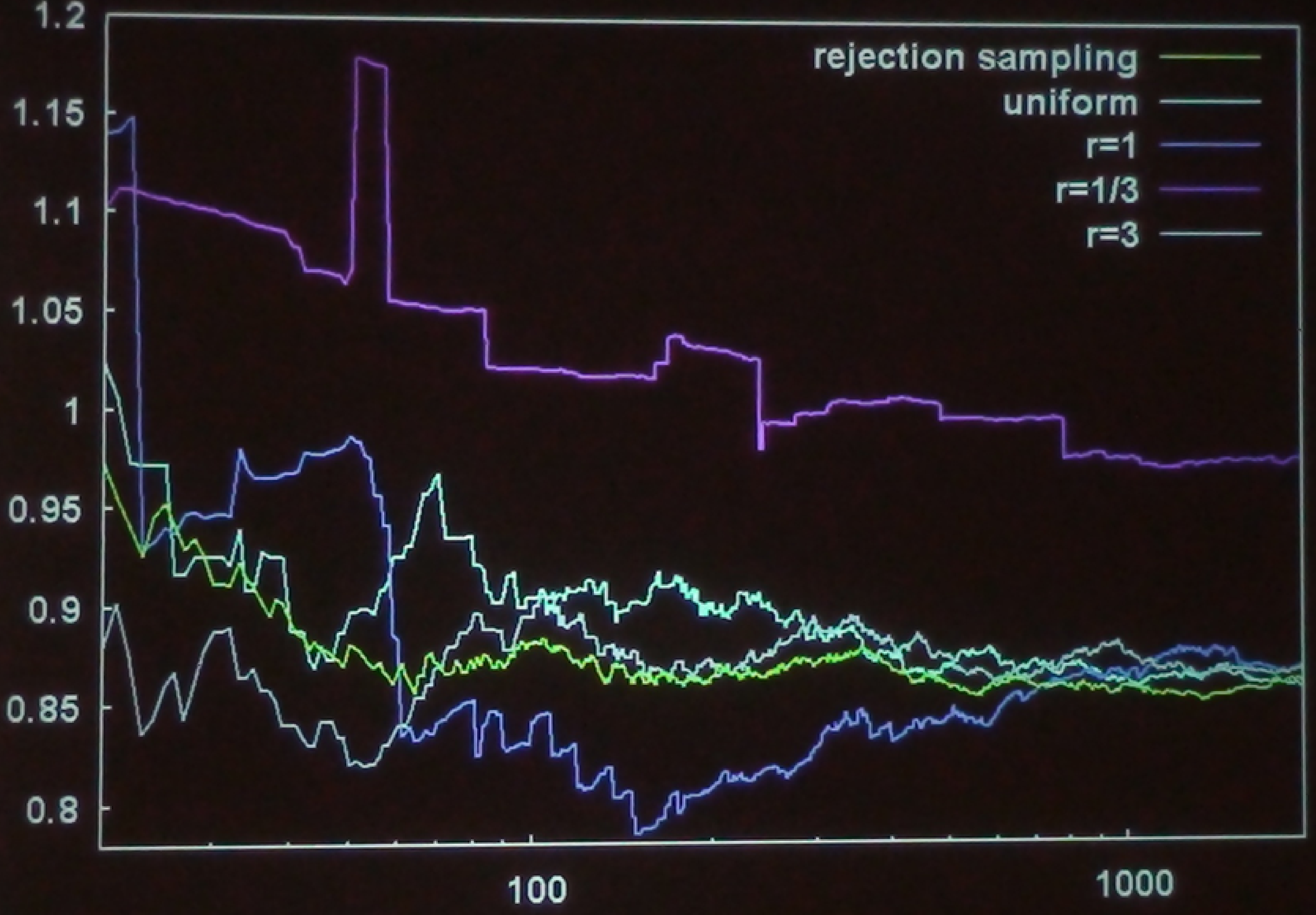

3
2.5
2
1.5
1
0.5
0
-4
-2
0
2
4
P*(x)
Q*(x)
phi(x)

50 samples, 19 accepted

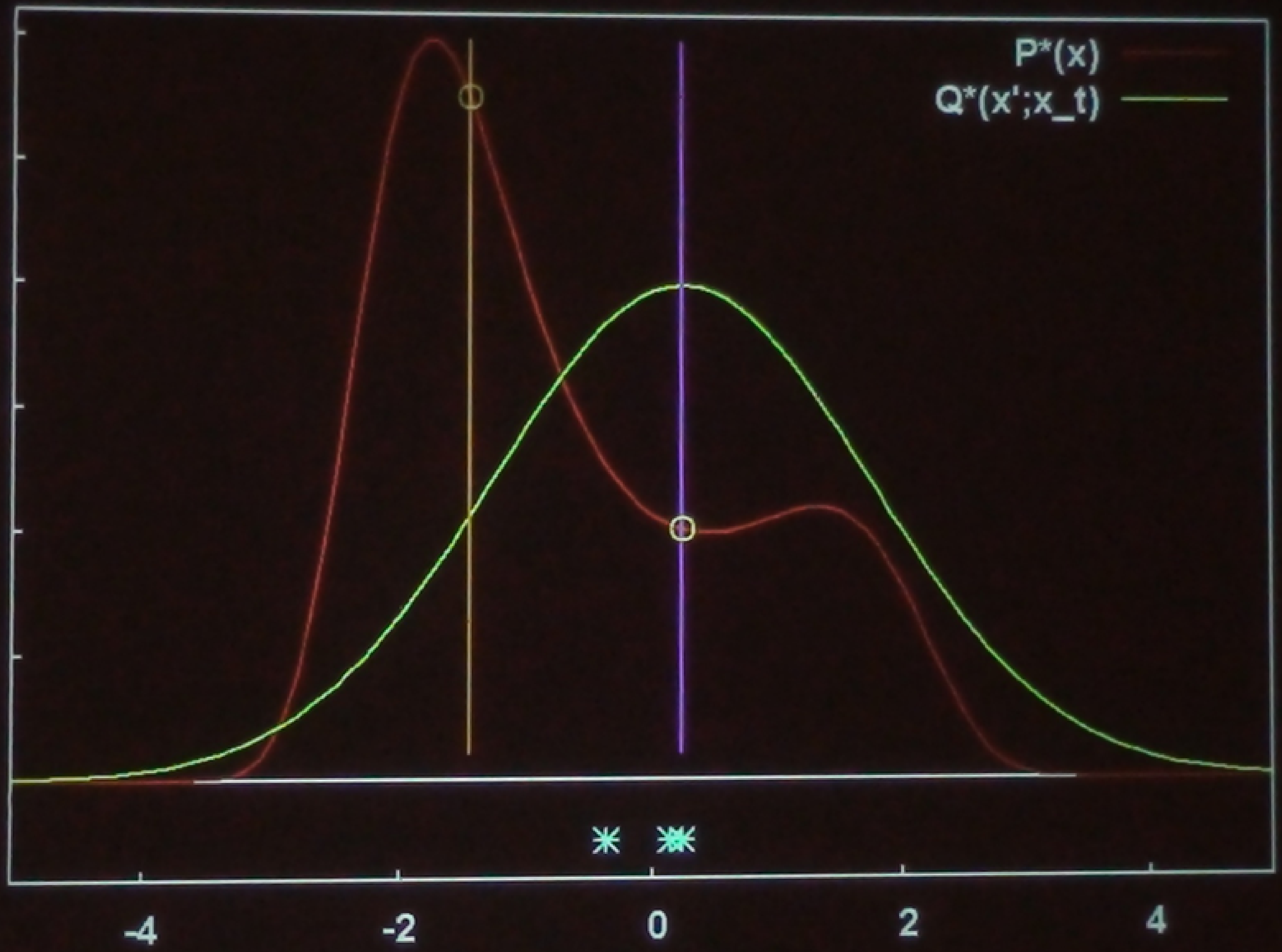


50 samples, 27 accepted




1.2
1.15
1.1
1.05
1
0.95
0.9
0.85
0.8
100
1000
rejection sampling
uniform
r=1
r=1/3
r=3

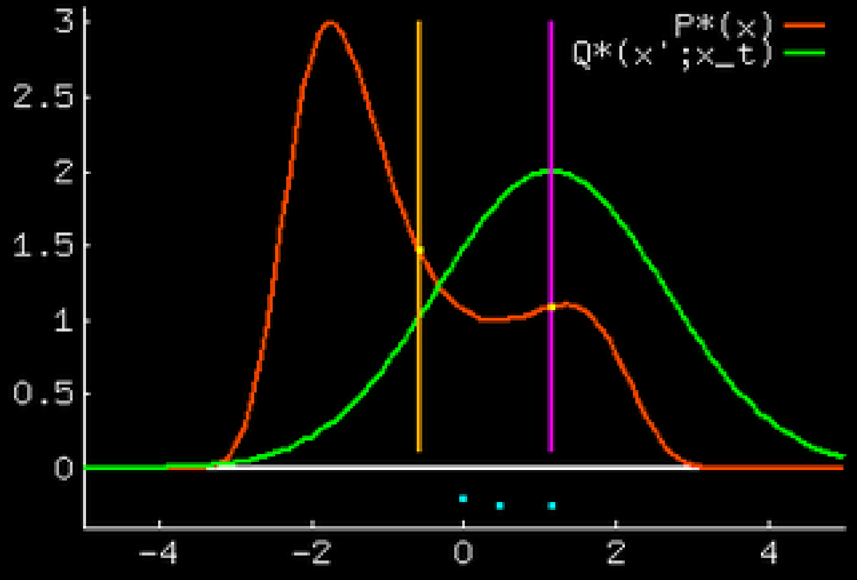
3 samples

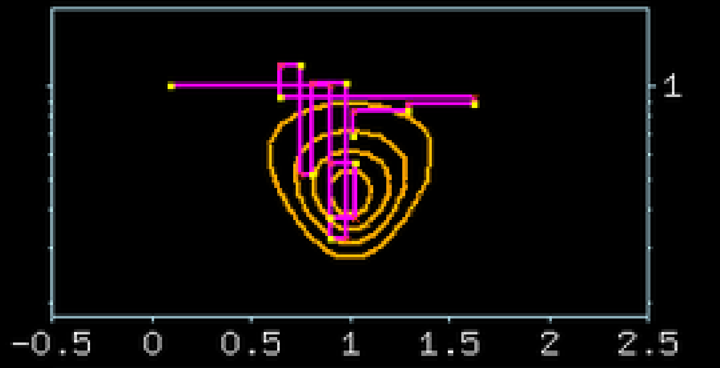


Monte Carlo methods

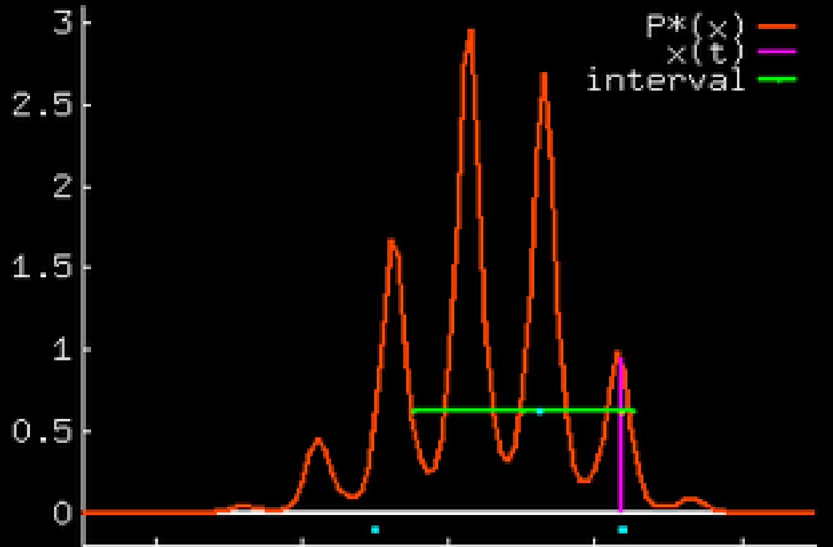
- Simple Monte Carlo methods
  - Importance sampling
  - Rejection sampling
- Markov-chain Monte Carlo methods
  - Metropolis method
  - Gibbs sampling
  - Slice sampling
- Reducing random-walk behaviour
  - Hamiltonian Monte Carlo
  - Overrelaxation
- Exact sampling

itp/exact/rc RUNME





Current hypothesis
Data



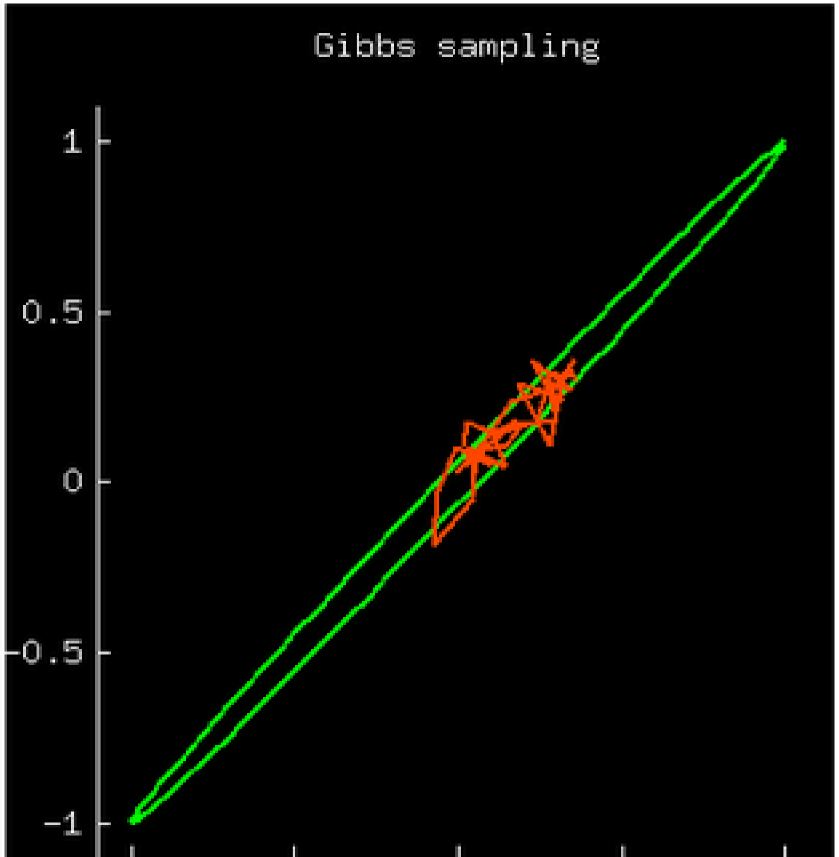
Problems with standard Monte Carlo methods

- Random walk behaviour
- Sensitivity to step size
- When to stop

Efficient methods

Slice sampling

Exact sampling



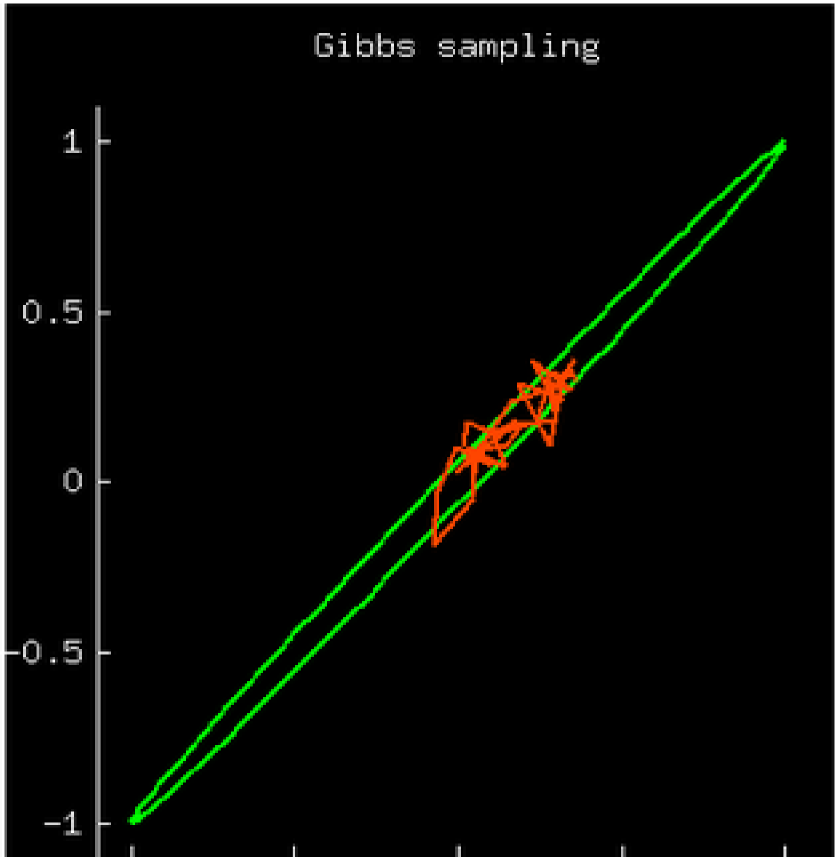
Problems with standard Monte Carlo methods

- Random walk behaviour
- Sensitivity to step size
- When to stop

Efficient methods

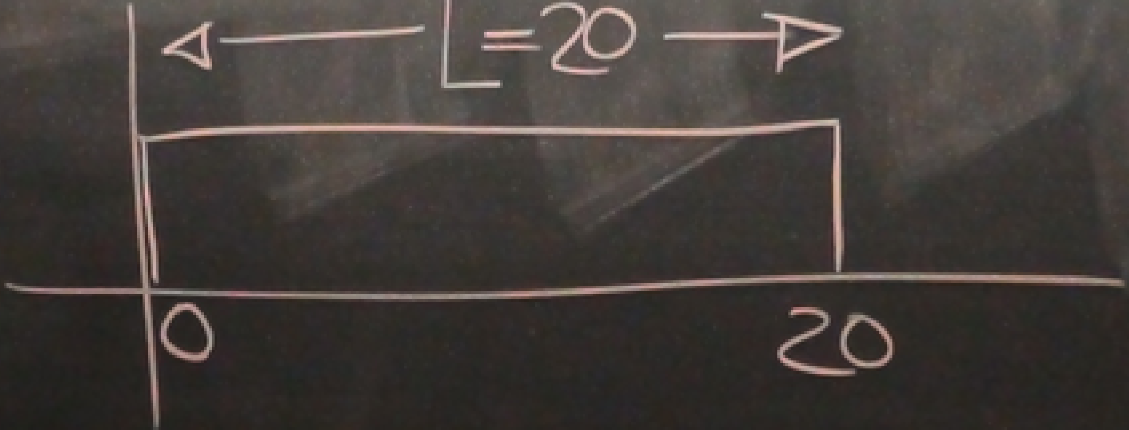
Slice sampling

Exact sampling



$$X \in \{\text{integers}\}$$

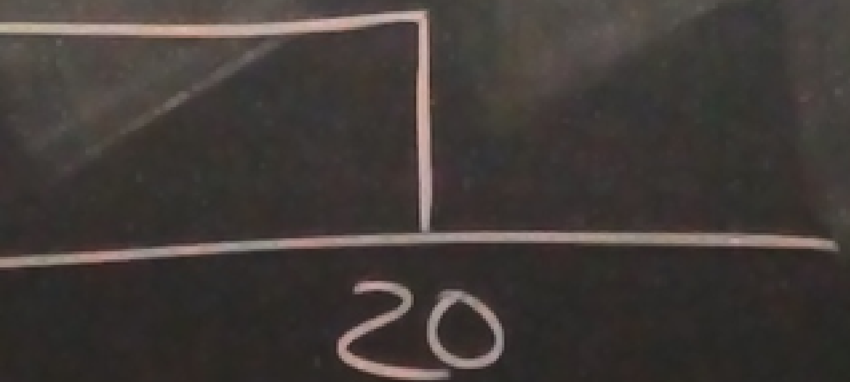
$$P^*(x) = \begin{cases} 1 & x \in \{0, 1, 2, 3 \ldots\ldots 20\} \\ 0 & \text{otherwise} \end{cases}$$



$$Q(\overset{\curvearrowleft}{x';x}) = \begin{cases} 1/2 & x' = x-1 \\ 1/2 & x' = x+1 \end{cases}$$

herwise

20 →

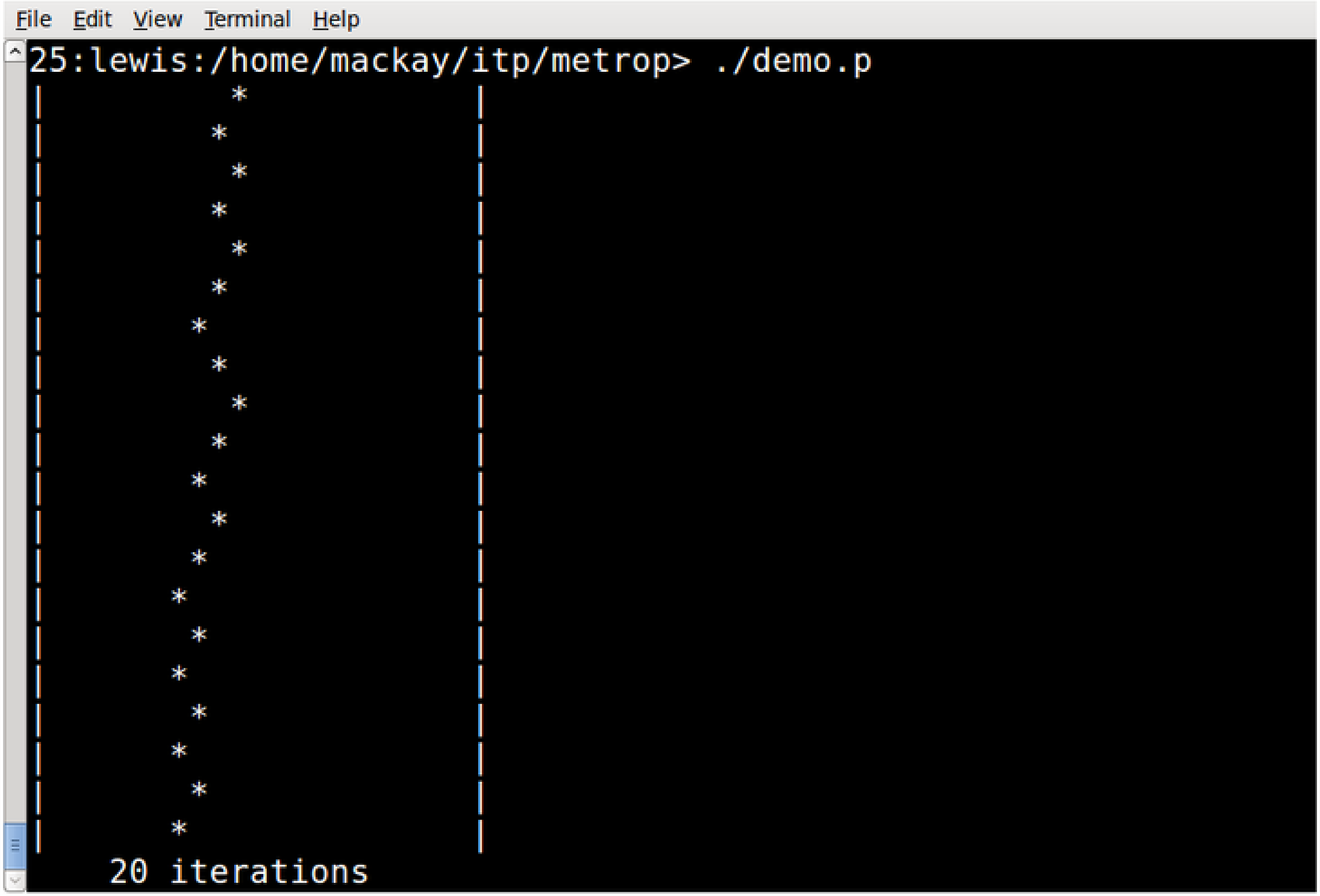


step size

↓

$\varepsilon = 1$

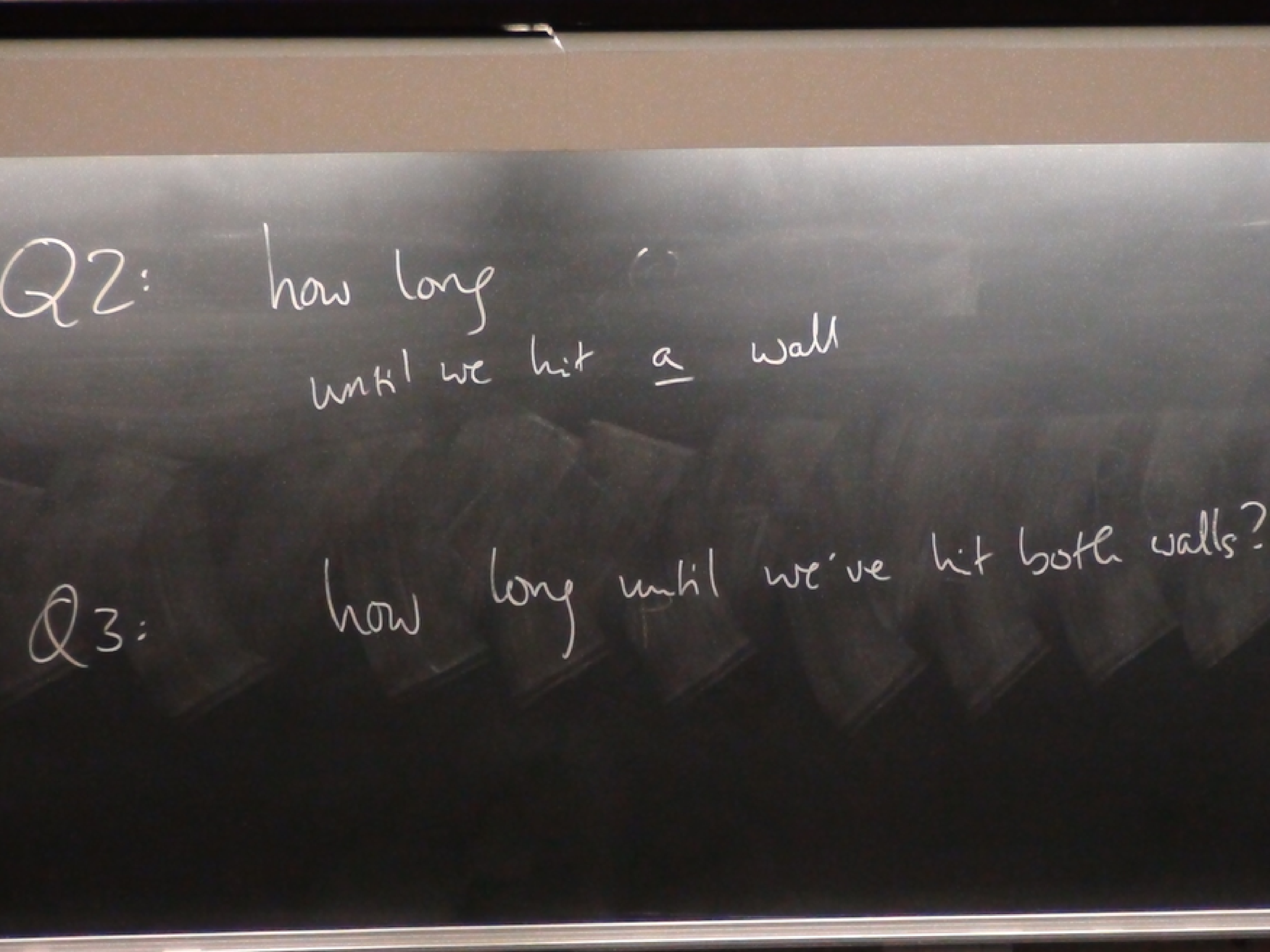
$$= \begin{cases} \frac{1}{2} & x' = x \boxed{-1} \\ \frac{1}{2} & x' = x \boxed{+1} \end{cases}$$



```
25:lewis:/home/mackay/itp/metrop> ./demo.p
|         *           |
|        *            |
|         *           |
|        *            |
|         *           |
|        *            |
|       *             |
|        *            |
|         *           |
|        *            |
|       *             |
|        *            |
|       *             |
|      *              |
|       *             |
|      *              |
|       *             |
|      *              |
|       *             |
|      *              |
    20 iterations
```

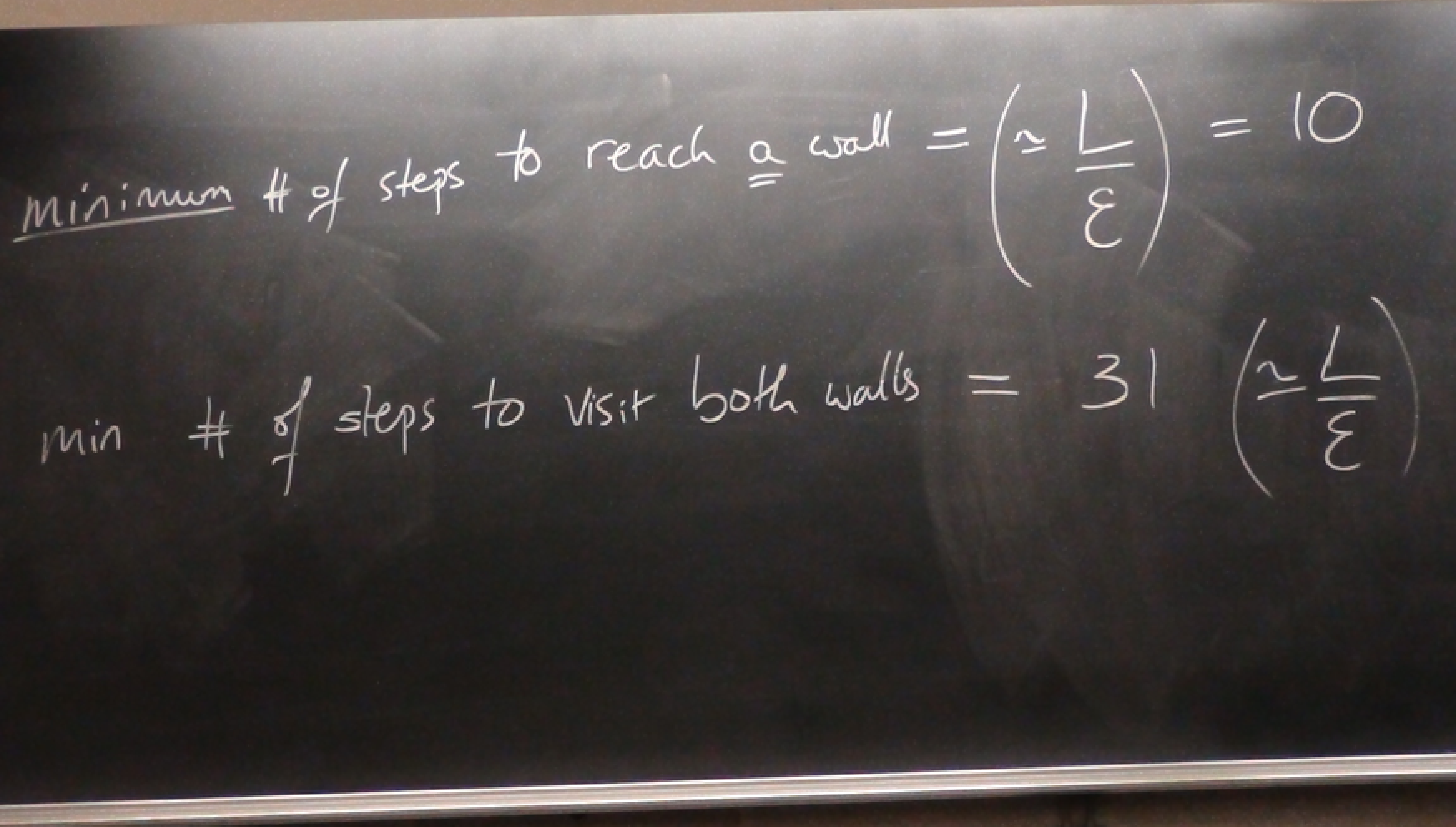
Q1: roughly how long until the chain generates a "good" sample from

P ?



Q2: how long
until we hit a wall

Q3: how long until we've hit both walls?



minimum # of steps to reach a wall $= \left(\simeq \frac{L}{\varepsilon}\right) = 10$

min # of steps to visit both walls $= 31 \quad \left(\simeq \frac{L}{\varepsilon}\right)$

distance after time T

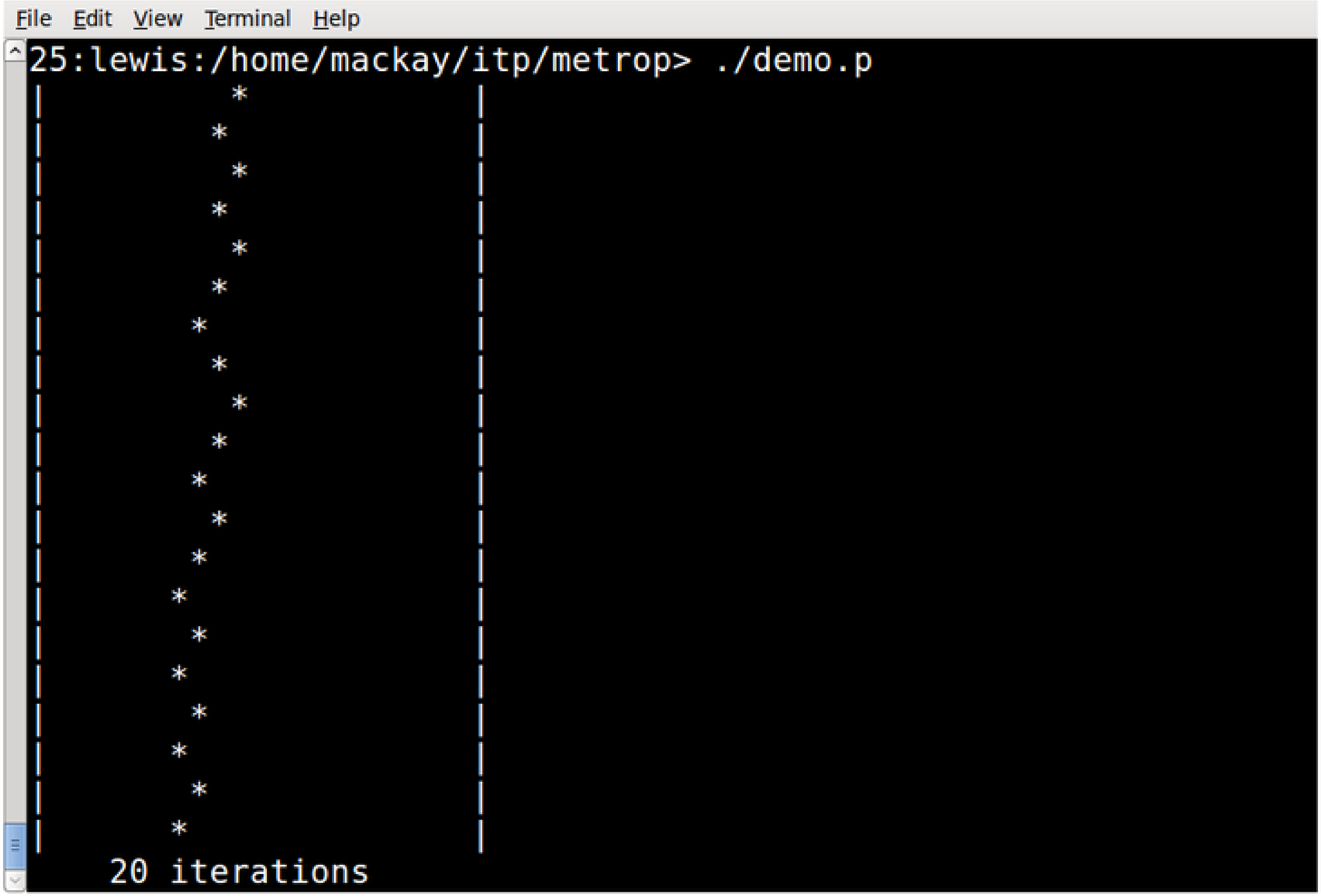
$$\Delta x = \sum_{t=1}^{T} s_t \qquad s_t = \pm 1$$

neglect walls

Var(Δx)

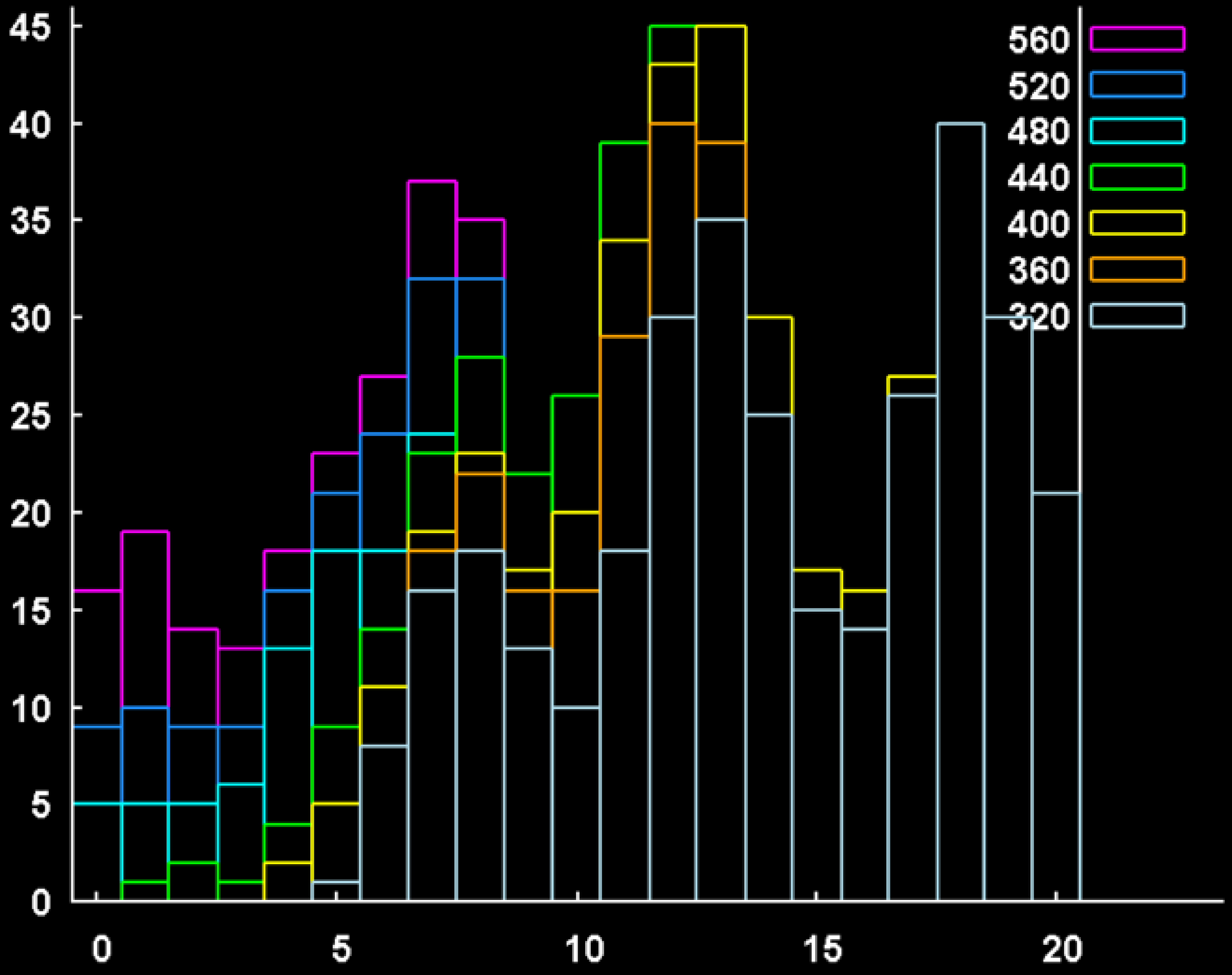
$$\langle \Delta x^2 \rangle = \sum_{1}^{T} \langle s_t^2 \rangle = T$$

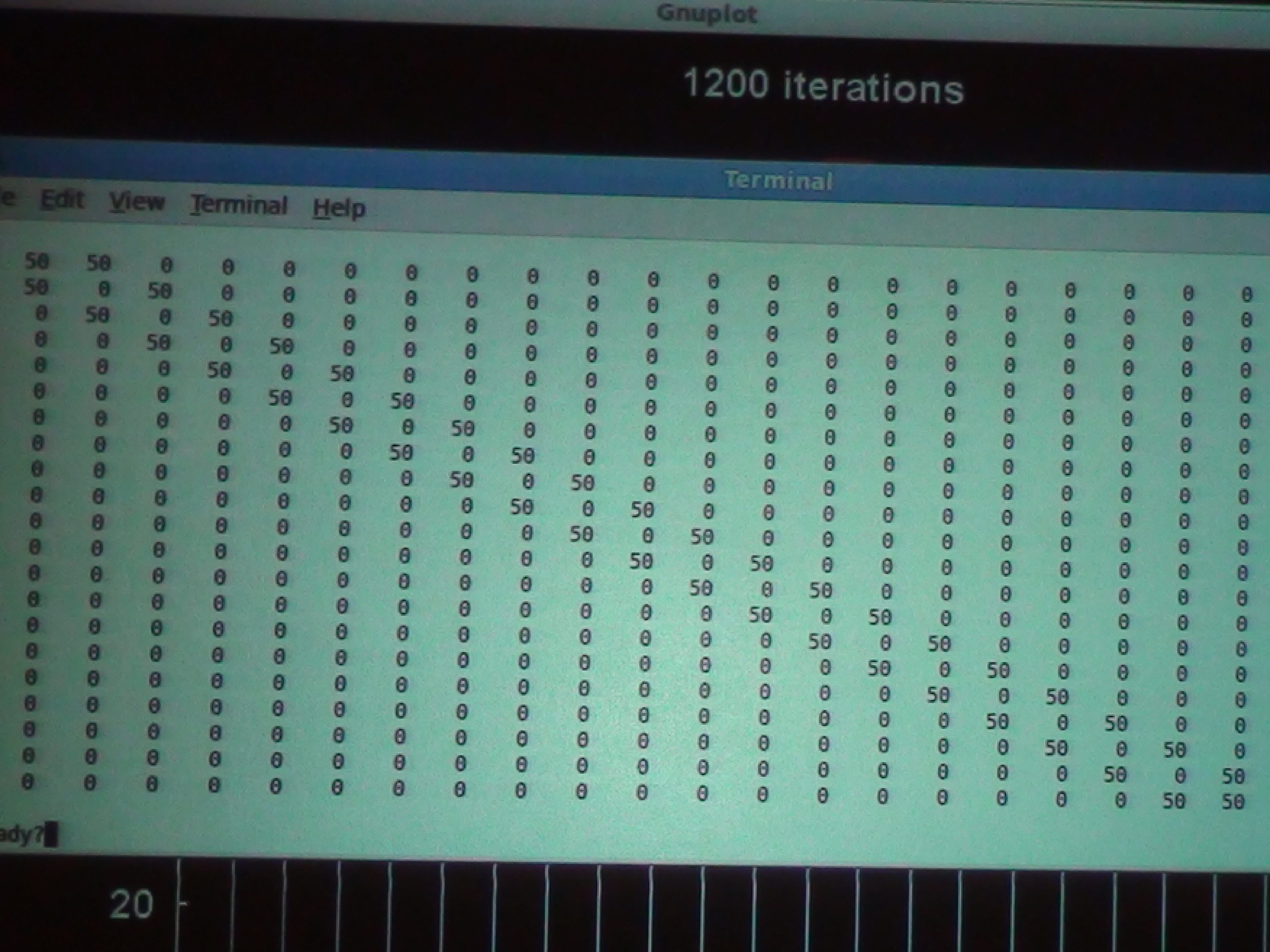
So $\langle \Delta x^2 \rangle = L^2$ need $T = \left(L/\varepsilon \right)^2$



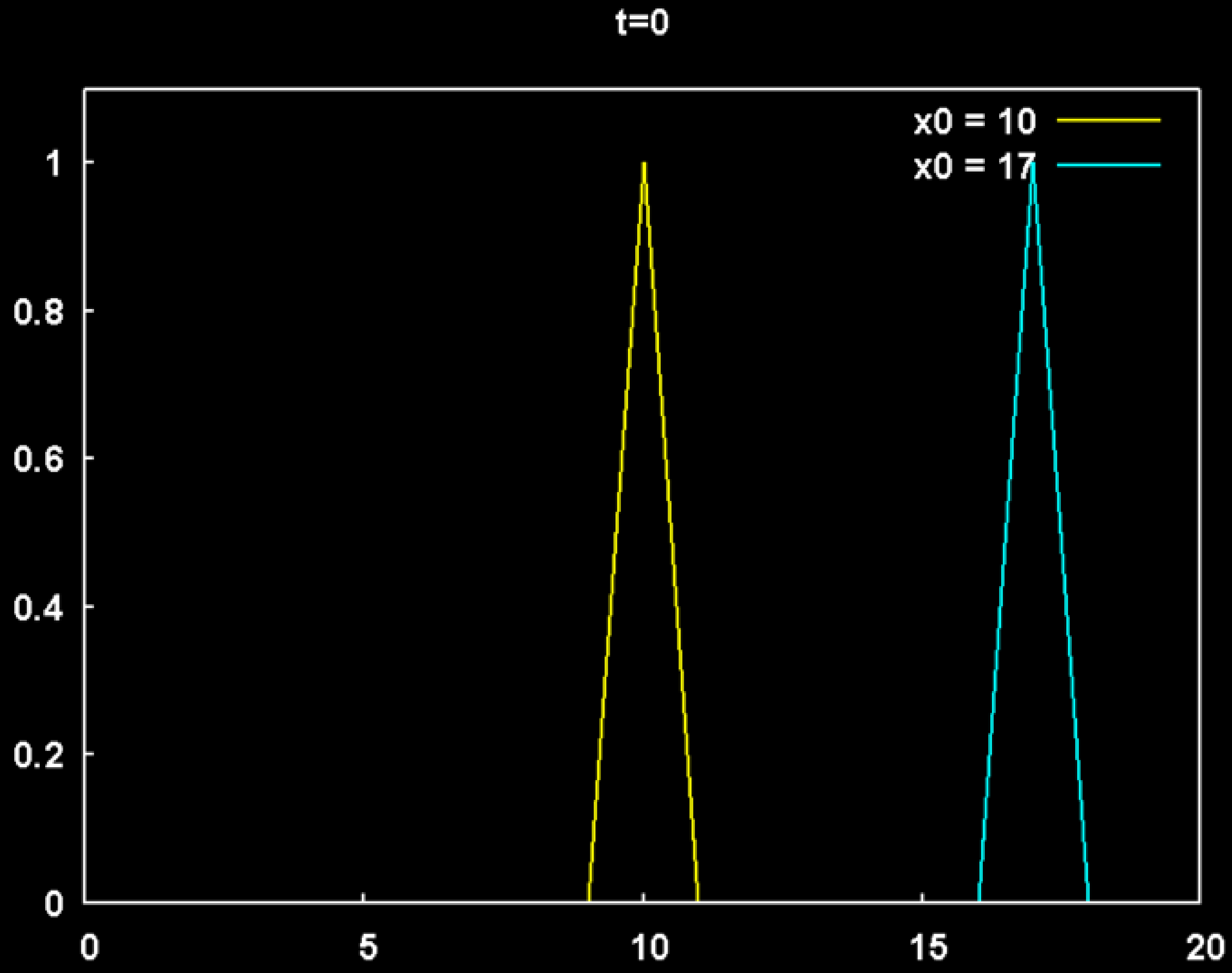
```
25:lewis:/home/mackay/itp/metrop> ./demo.p
|         *           |
|        *            |
|         *           |
|        *            |
|         *           |
|        *            |
|       *             |
|        *            |
|         *           |
|        *            |
|       *             |
|        *            |
|       *             |
|      *              |
|       *             |
|      *              |
|       *             |
|      *              |
|       *             |
|      *              |
    20 iterations
```

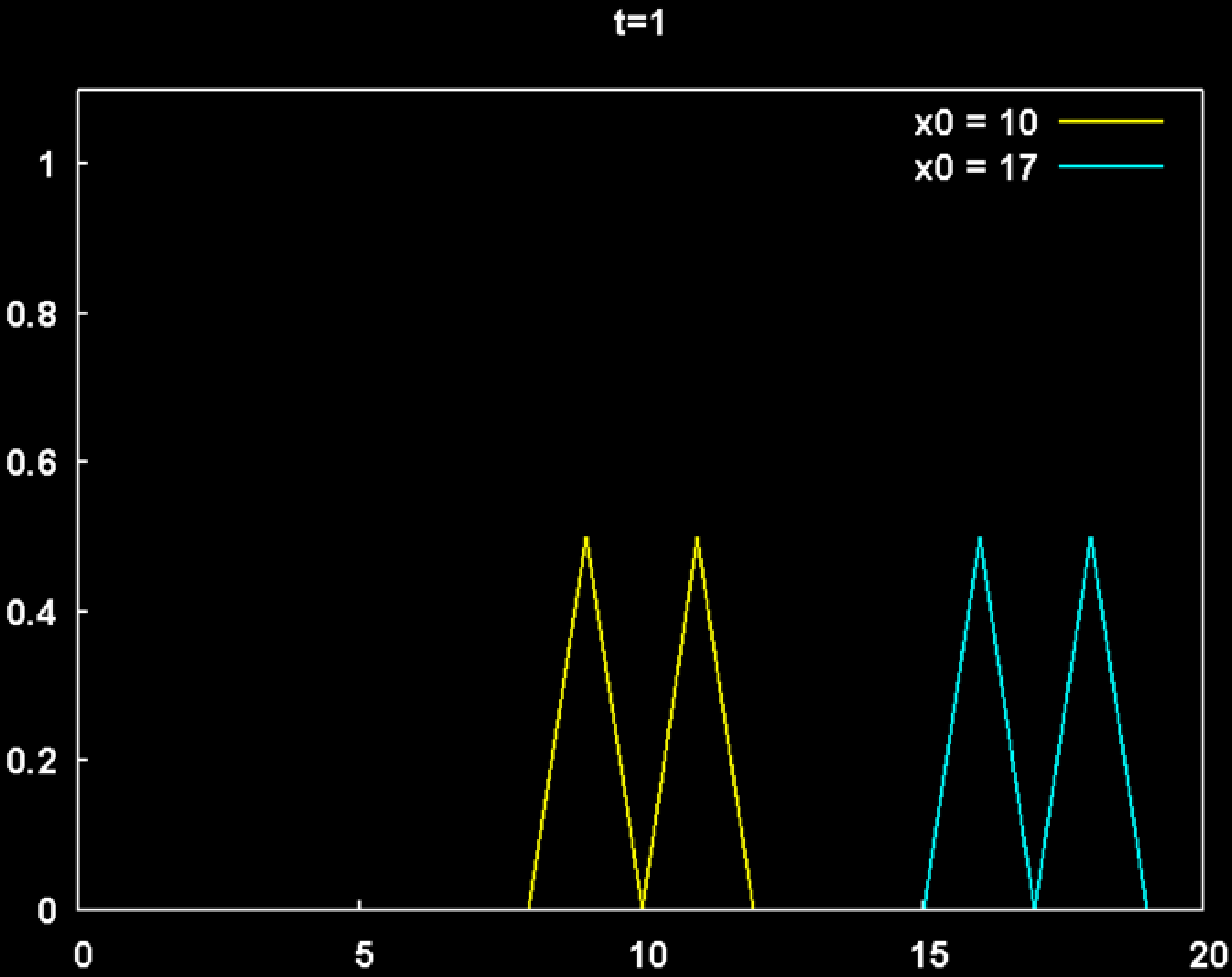
560 iterations

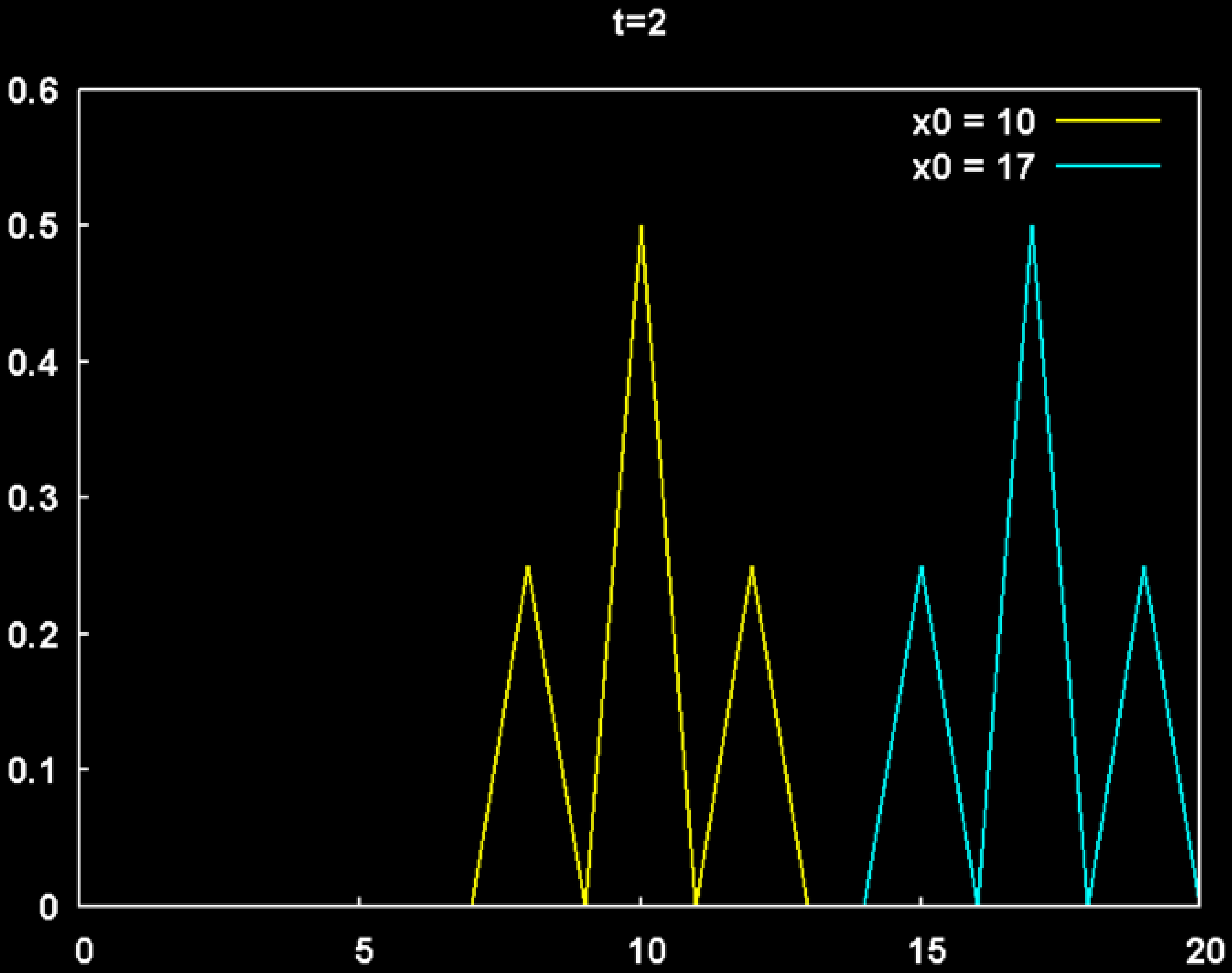


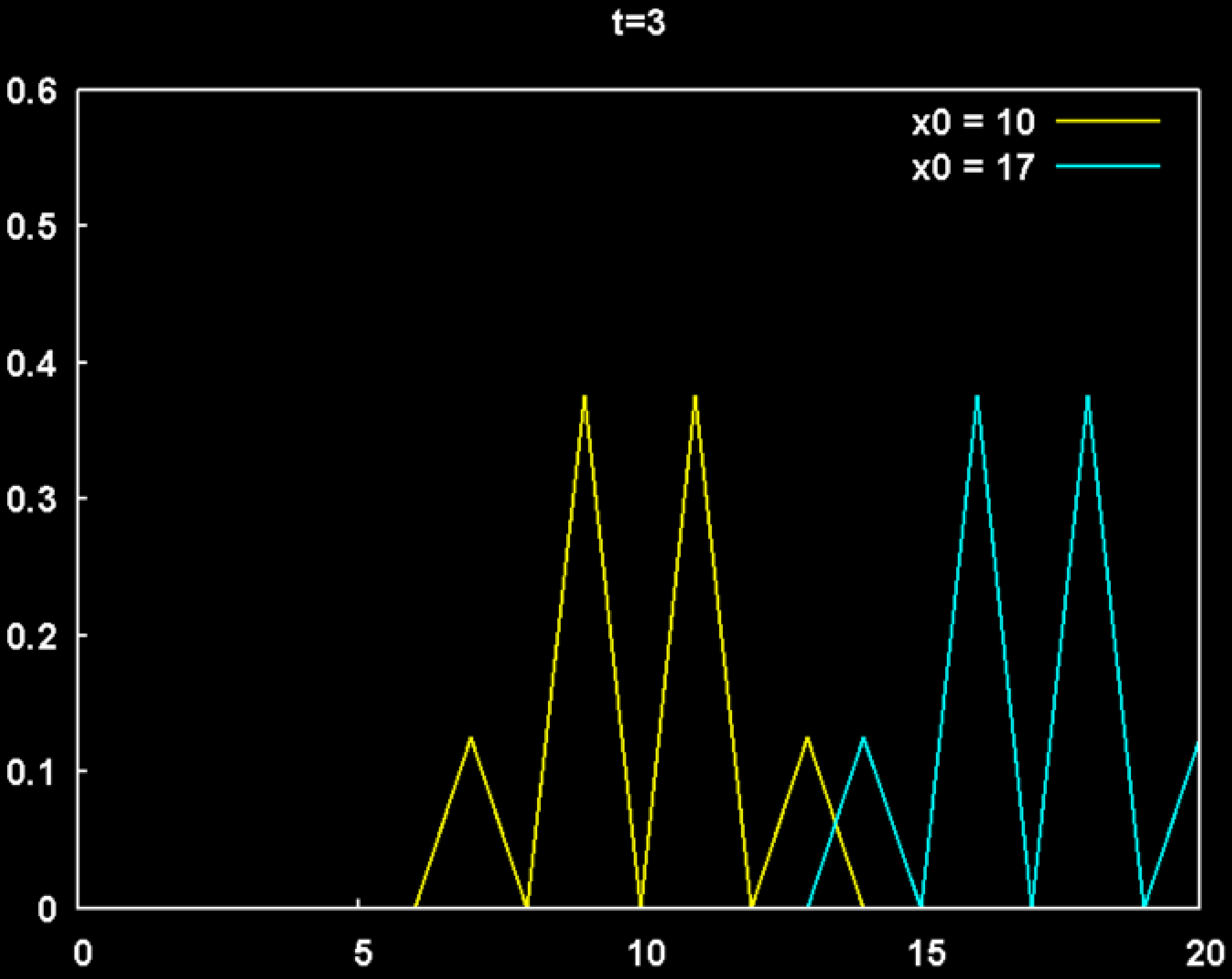


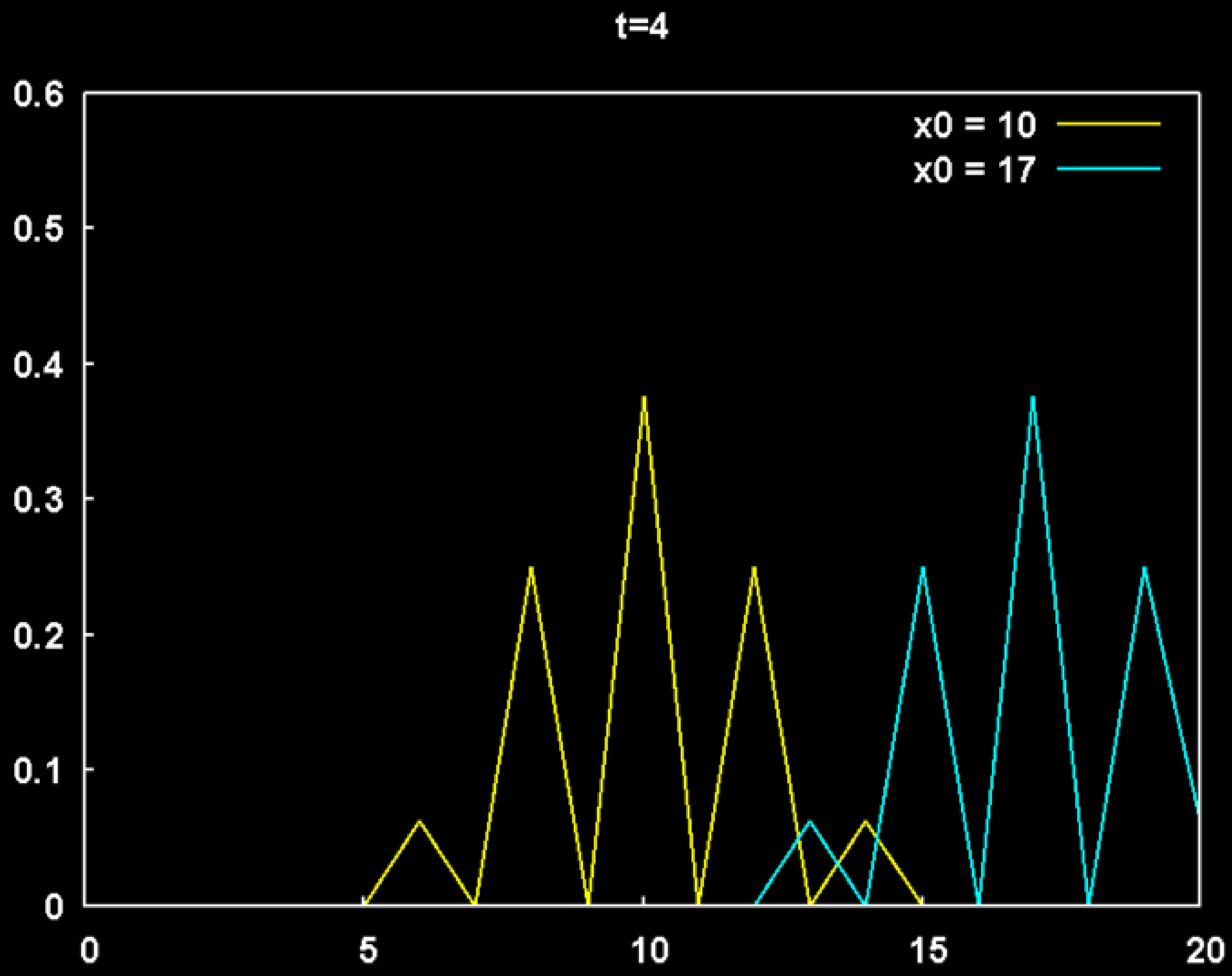
```
   50   50    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
   50    0   50    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0   50    0   50    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0   50    0   50    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0   50    0   50    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0   50    0   50    0    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0   50    0   50    0    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0   50    0   50    0    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0   50    0   50    0    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0   50    0   50    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0   50    0   50    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0   50    0   50    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0   50    0   50    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0   50    0   50    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0   50    0   50    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0   50    0   50    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   50    0   50    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   50    0   50    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   50    0   50    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   50    0   50
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0   50   50
ady?
```

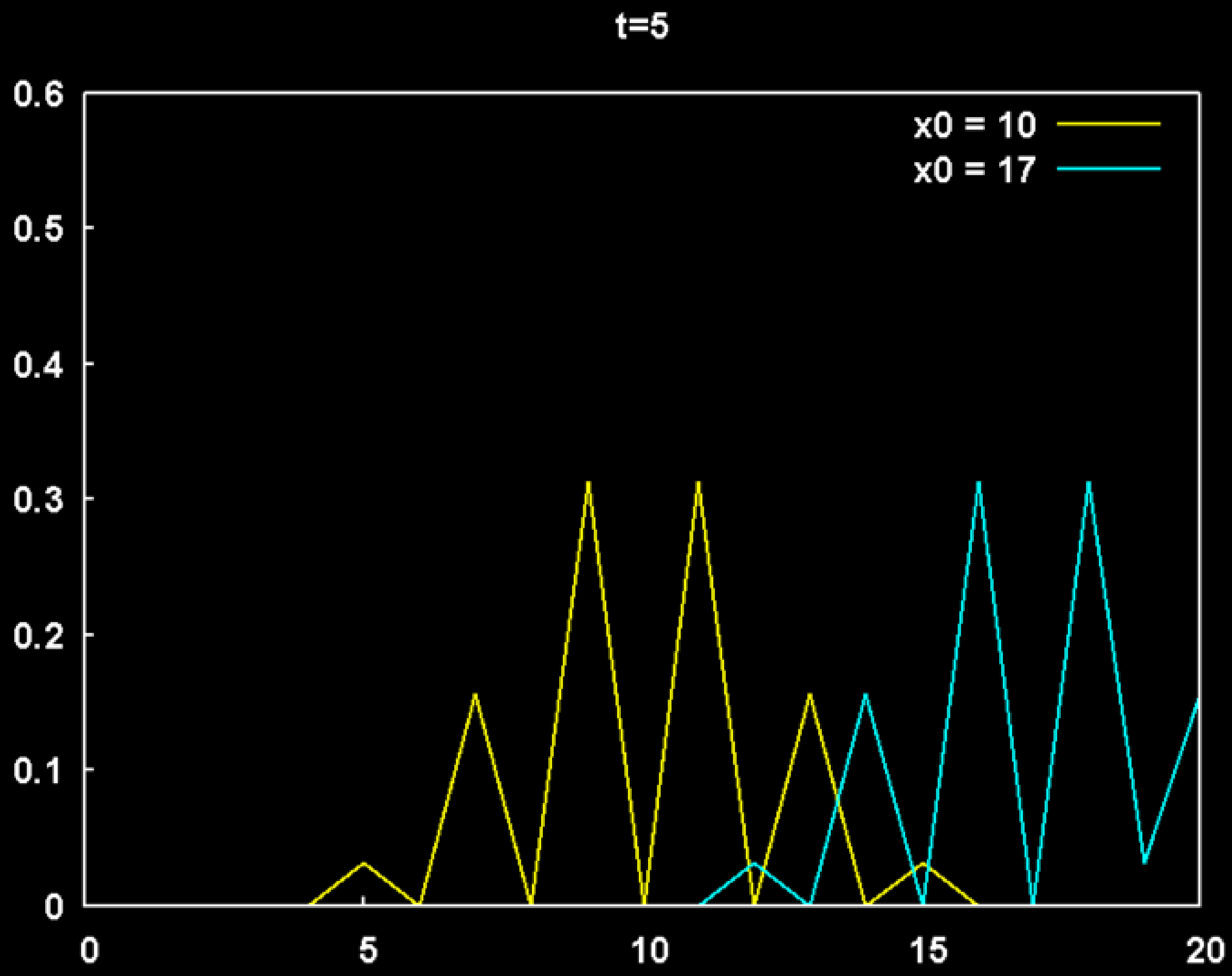

t=0
x0 = 10
x0 = 17
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20

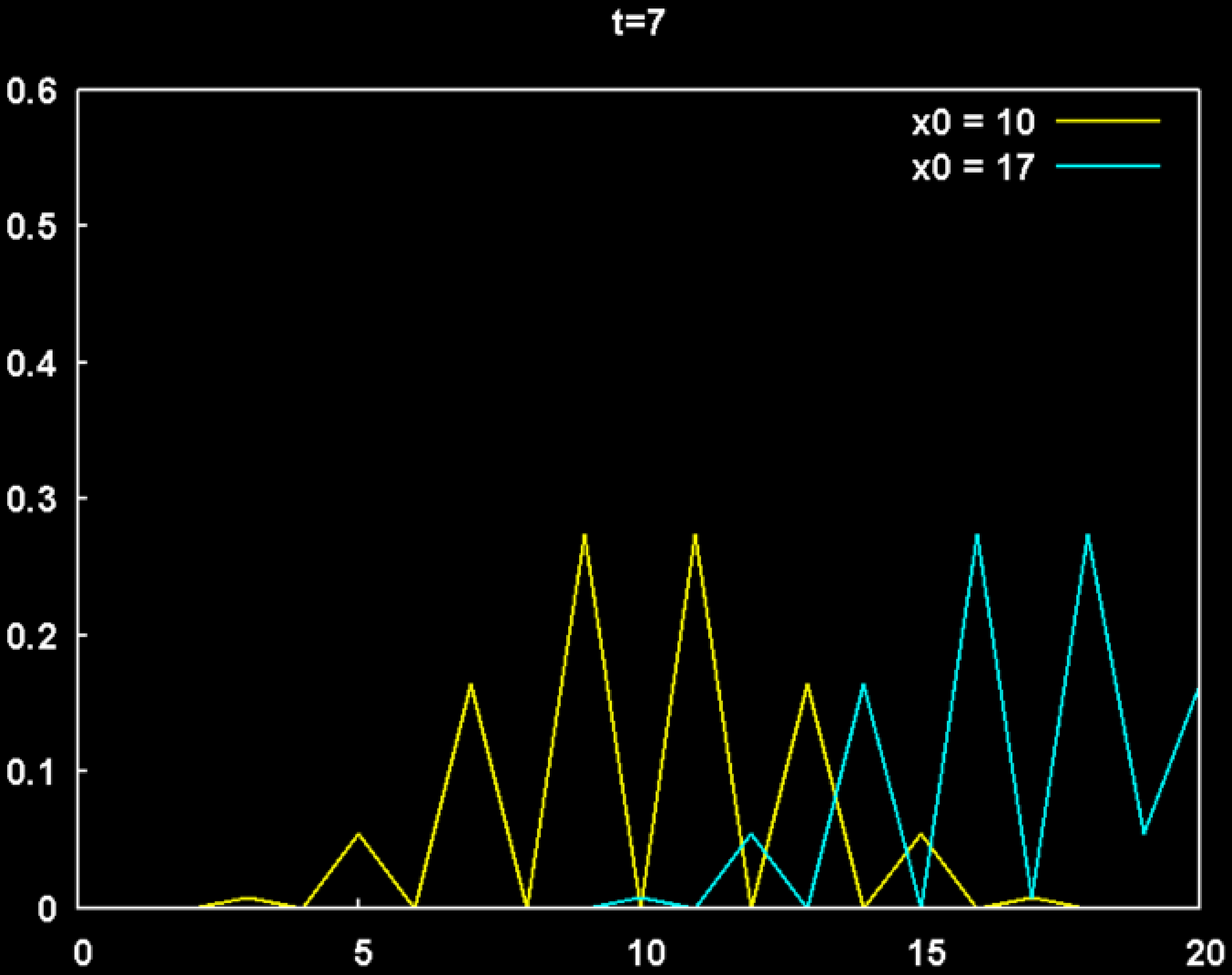

t=1
x0 = 10
x0 = 17
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20

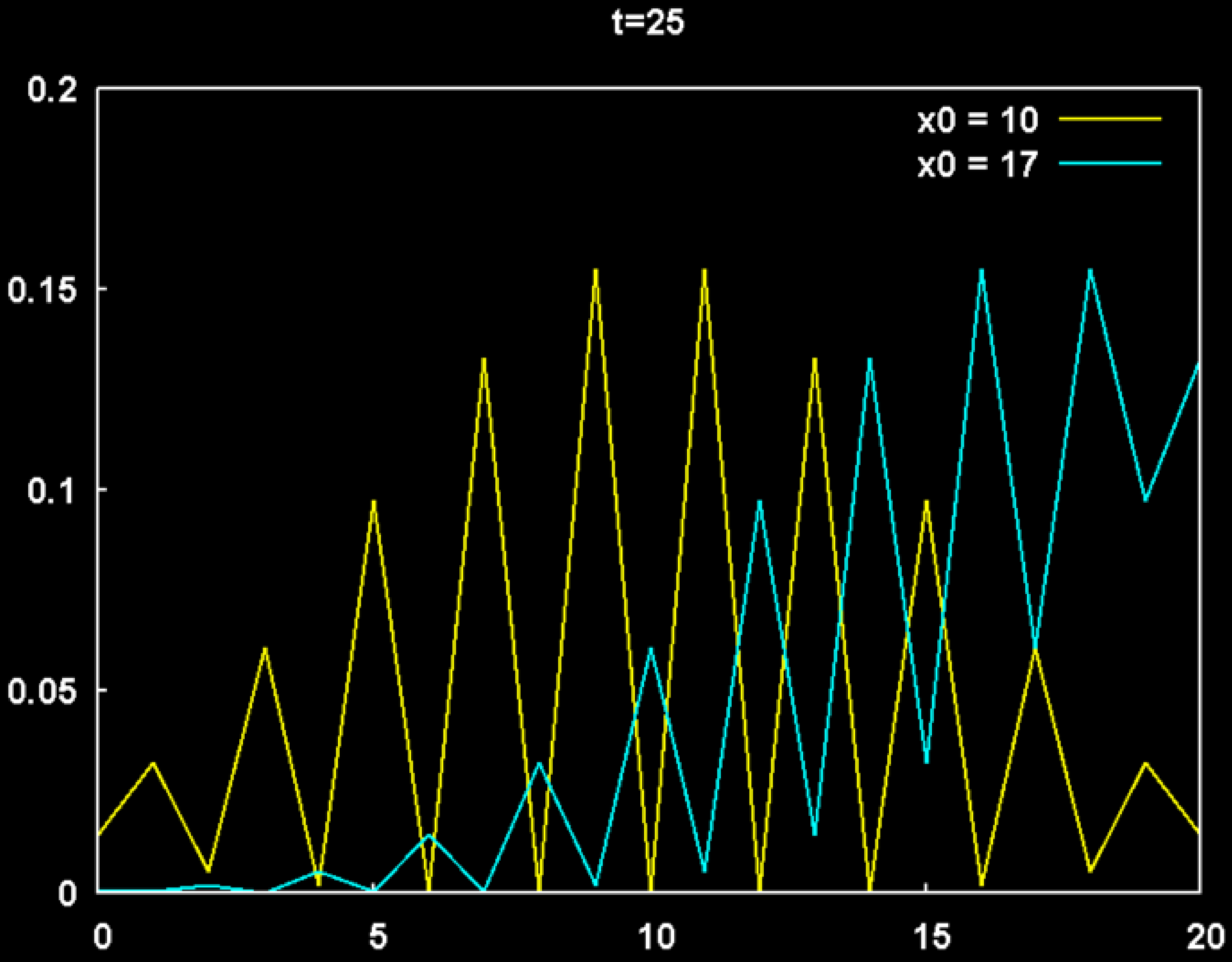

t=2
x0 = 10
x0 = 17
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20

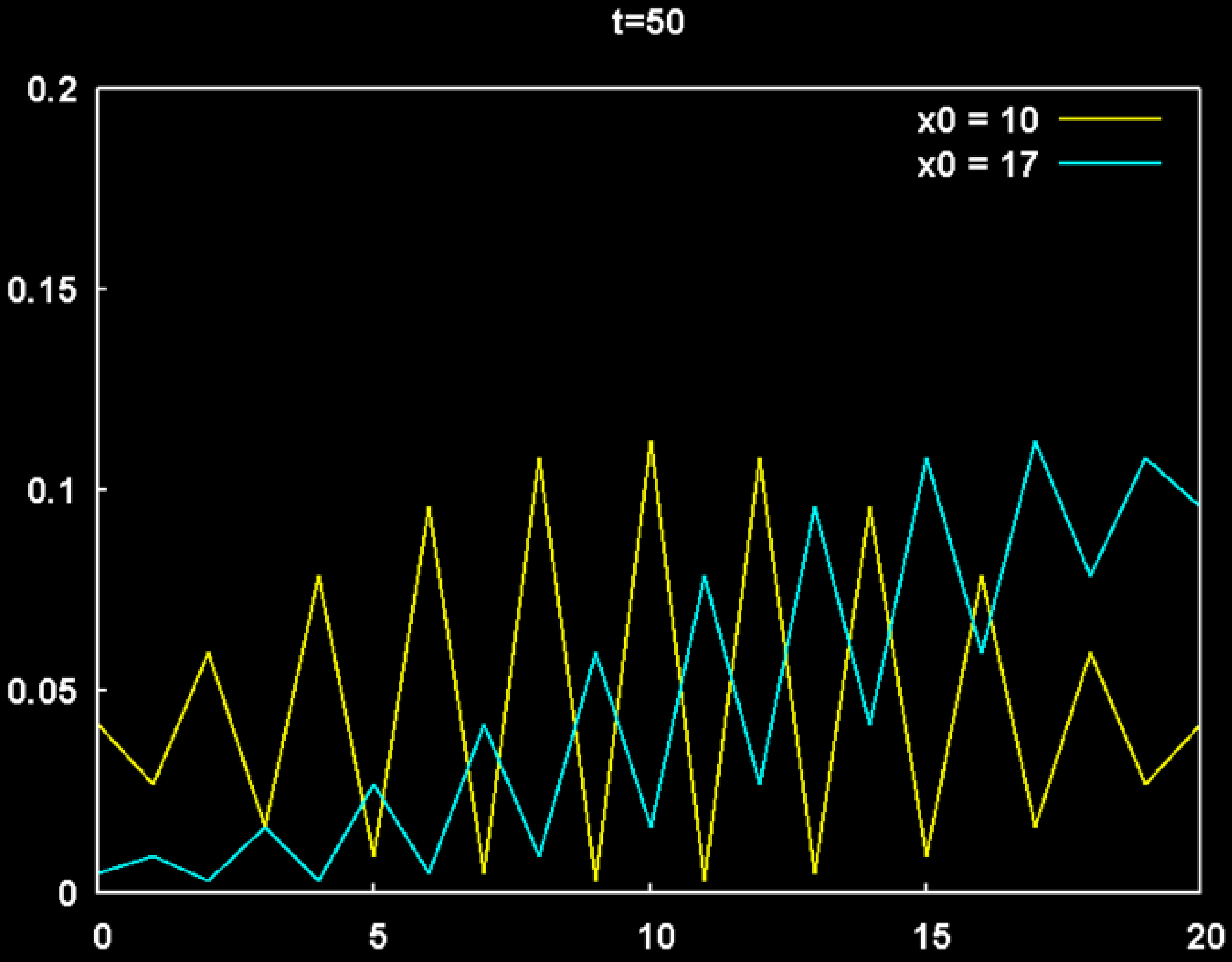

t=3
x0 = 10
x0 = 17
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20

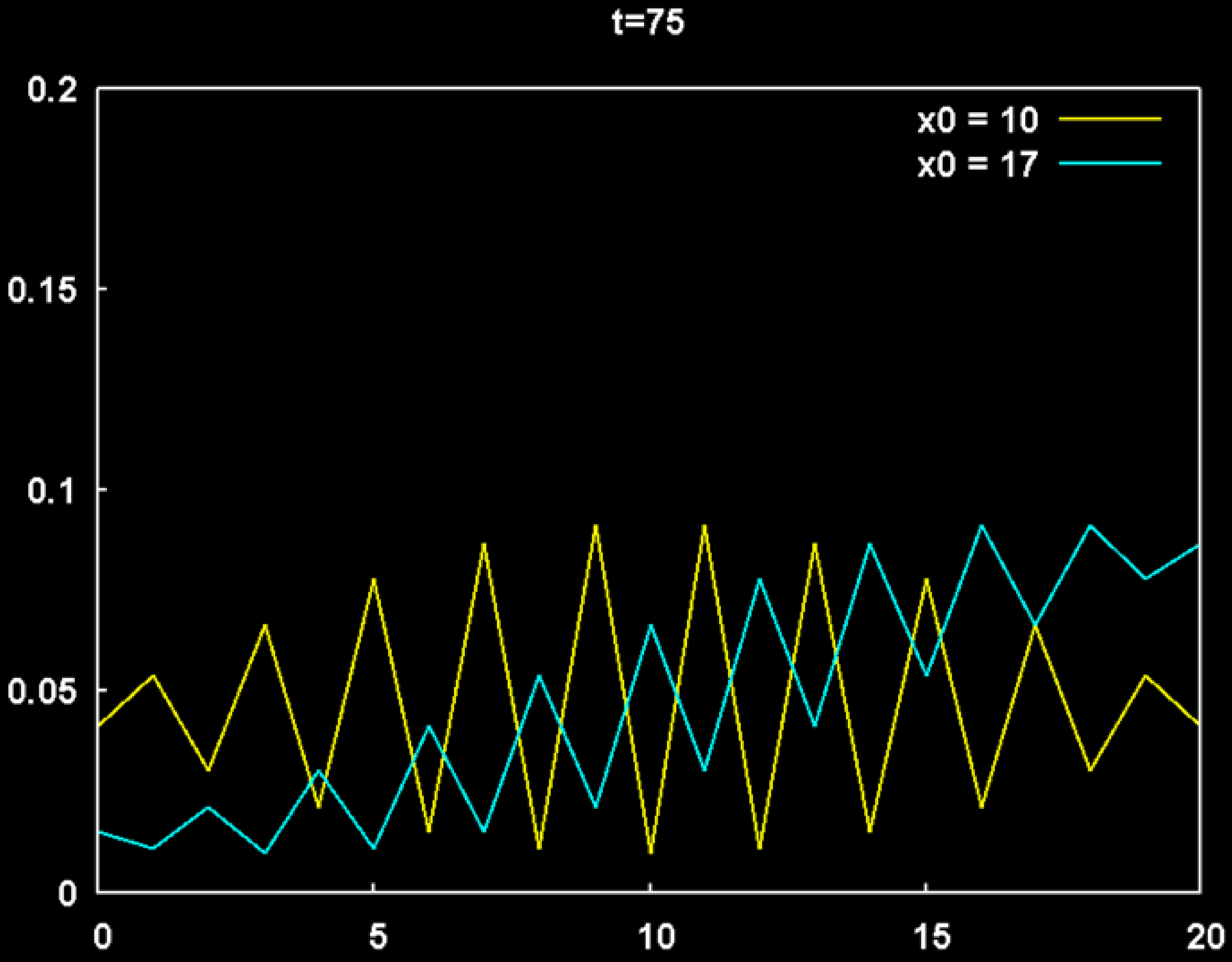

t=4
x0 = 10
x0 = 17
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20

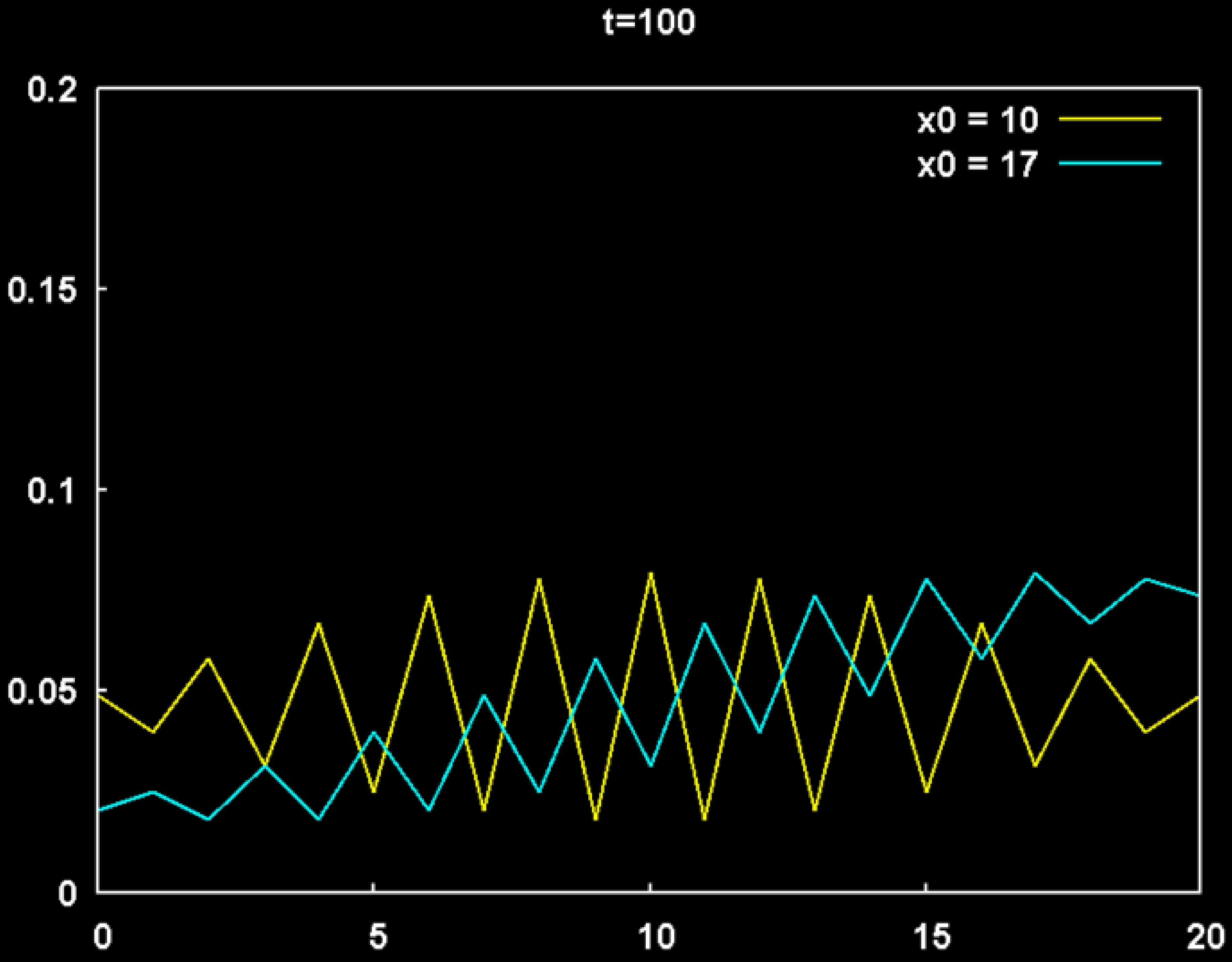

t=5
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
x0 = 10
x0 = 17

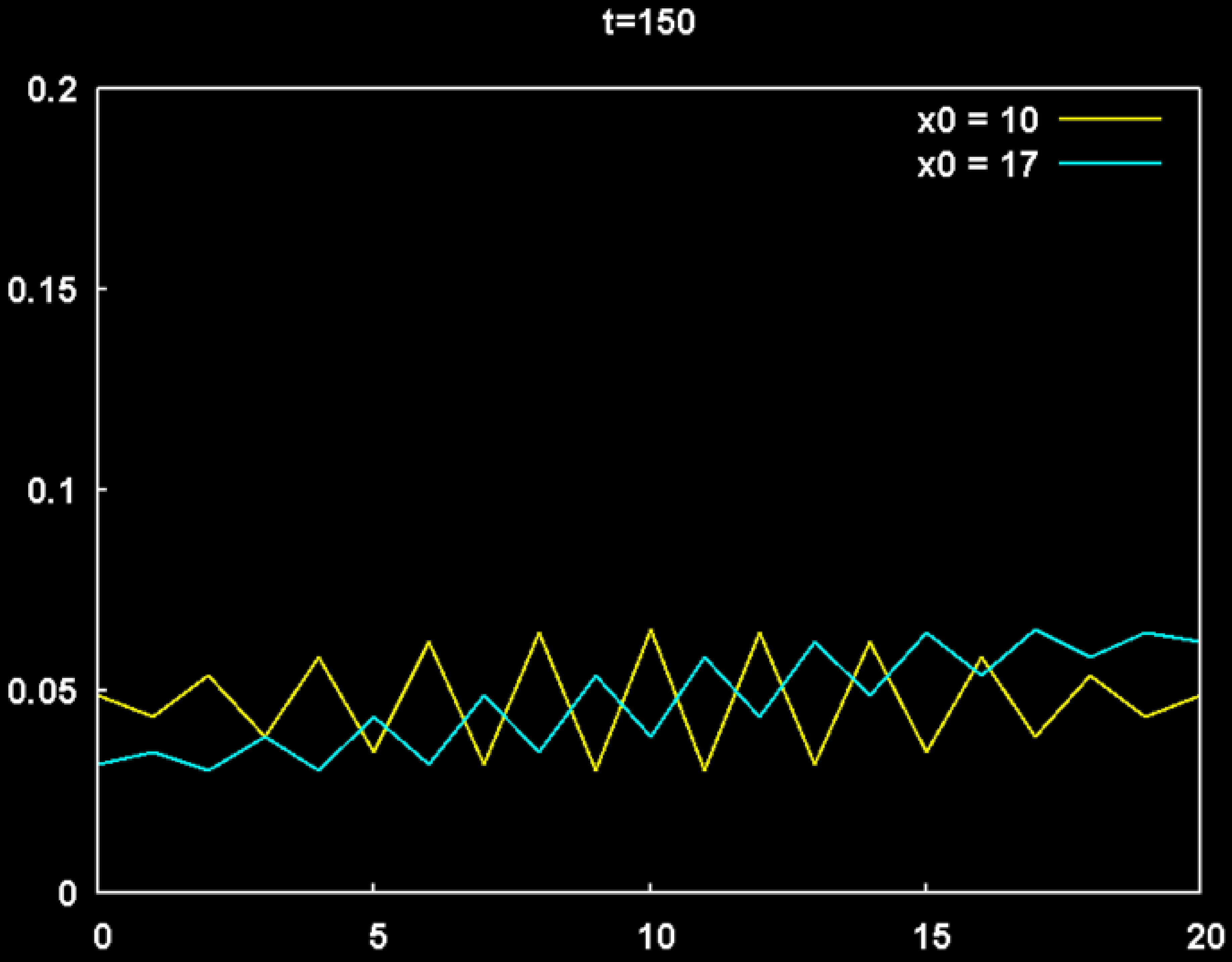

t=7
x0 = 10
x0 = 17
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20

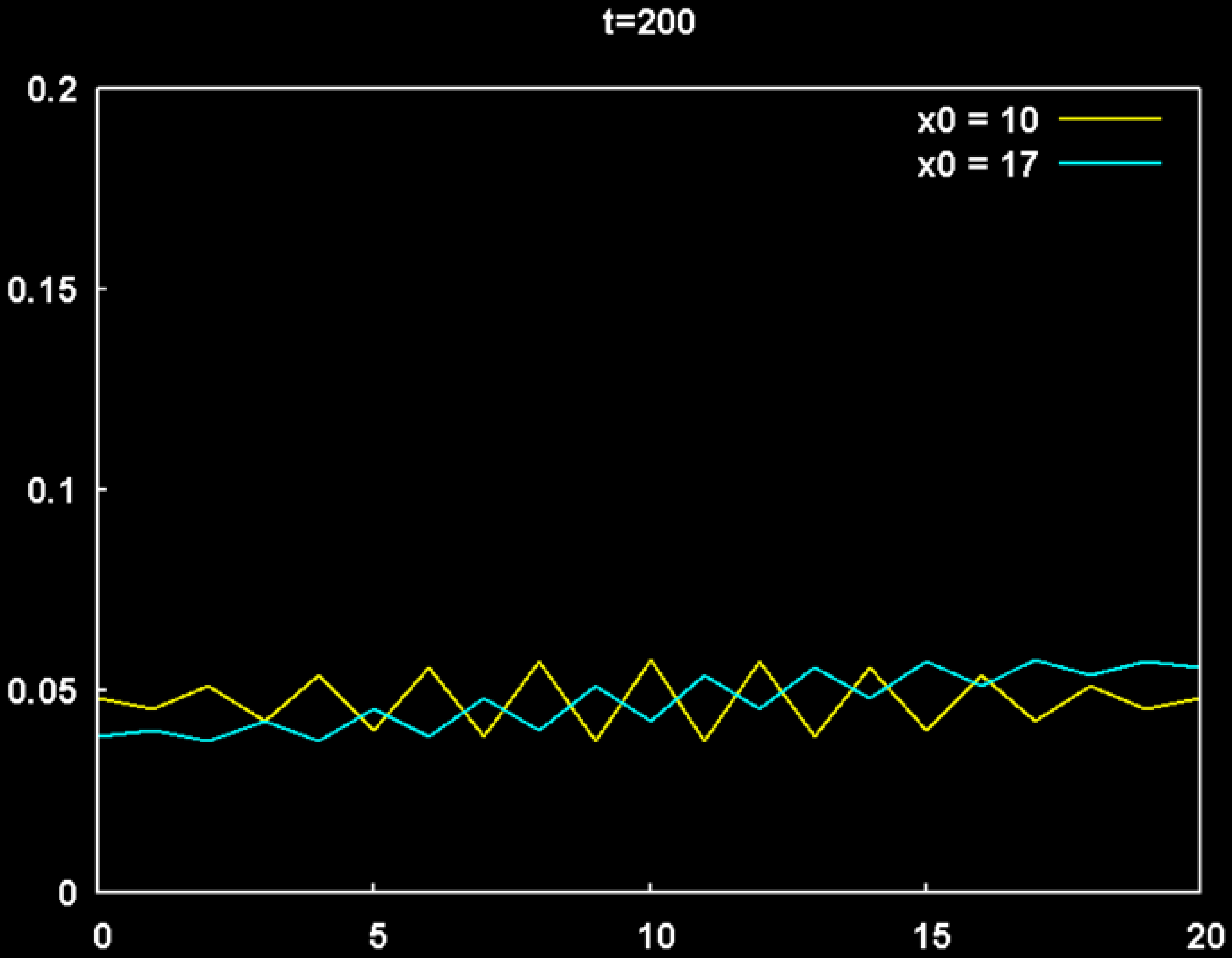

t=25
x0 = 10
x0 = 17
0.2
0.15
0.1
0.05
0
0
5
10
15
20

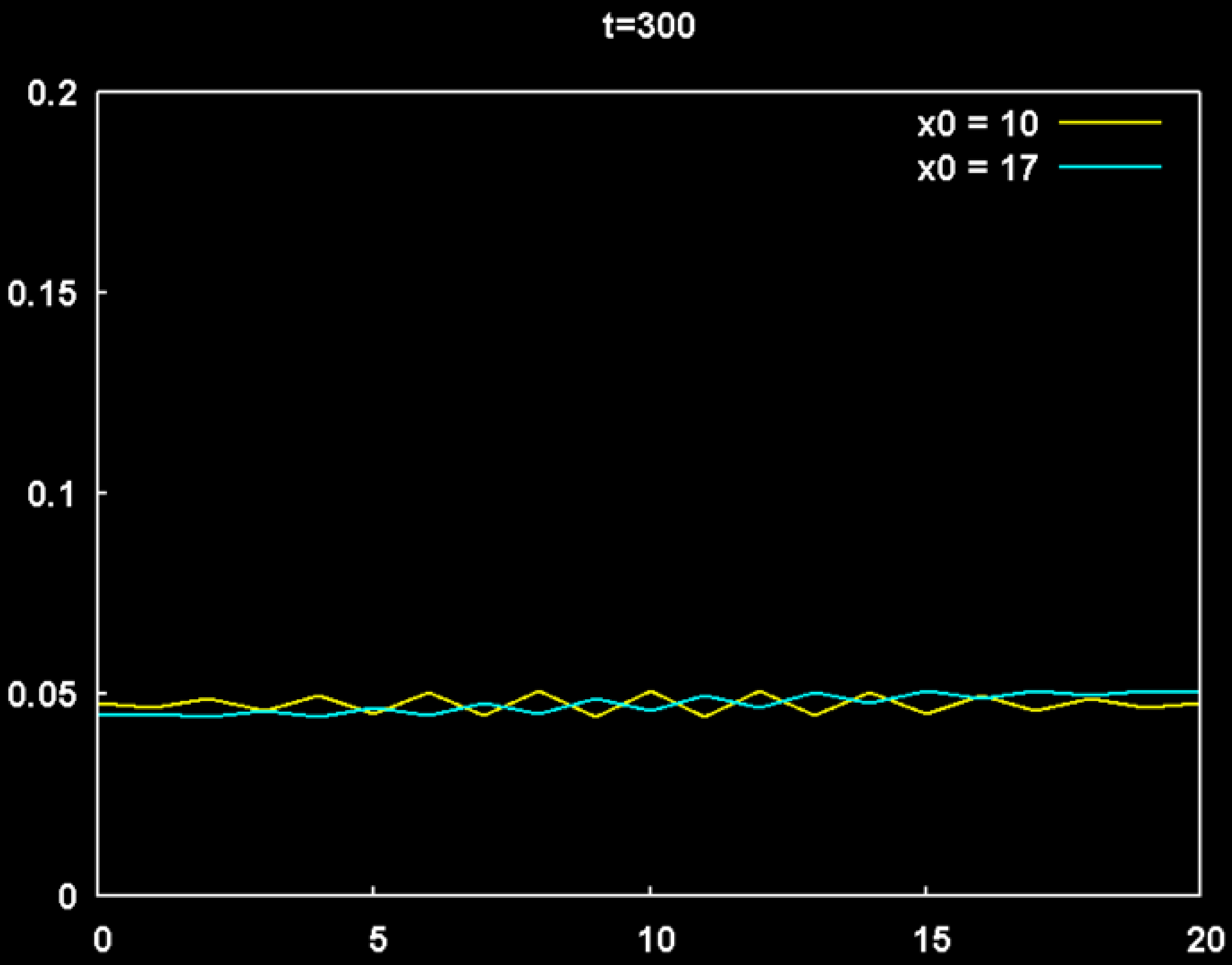

t=50
x0 = 10
x0 = 17
0.2
0.15
0.1
0.05
0
0
5
10
15
20


t=75
x0 = 10
x0 = 17
0.2
0.15
0.1
0.05
0
0
5
10
15
20

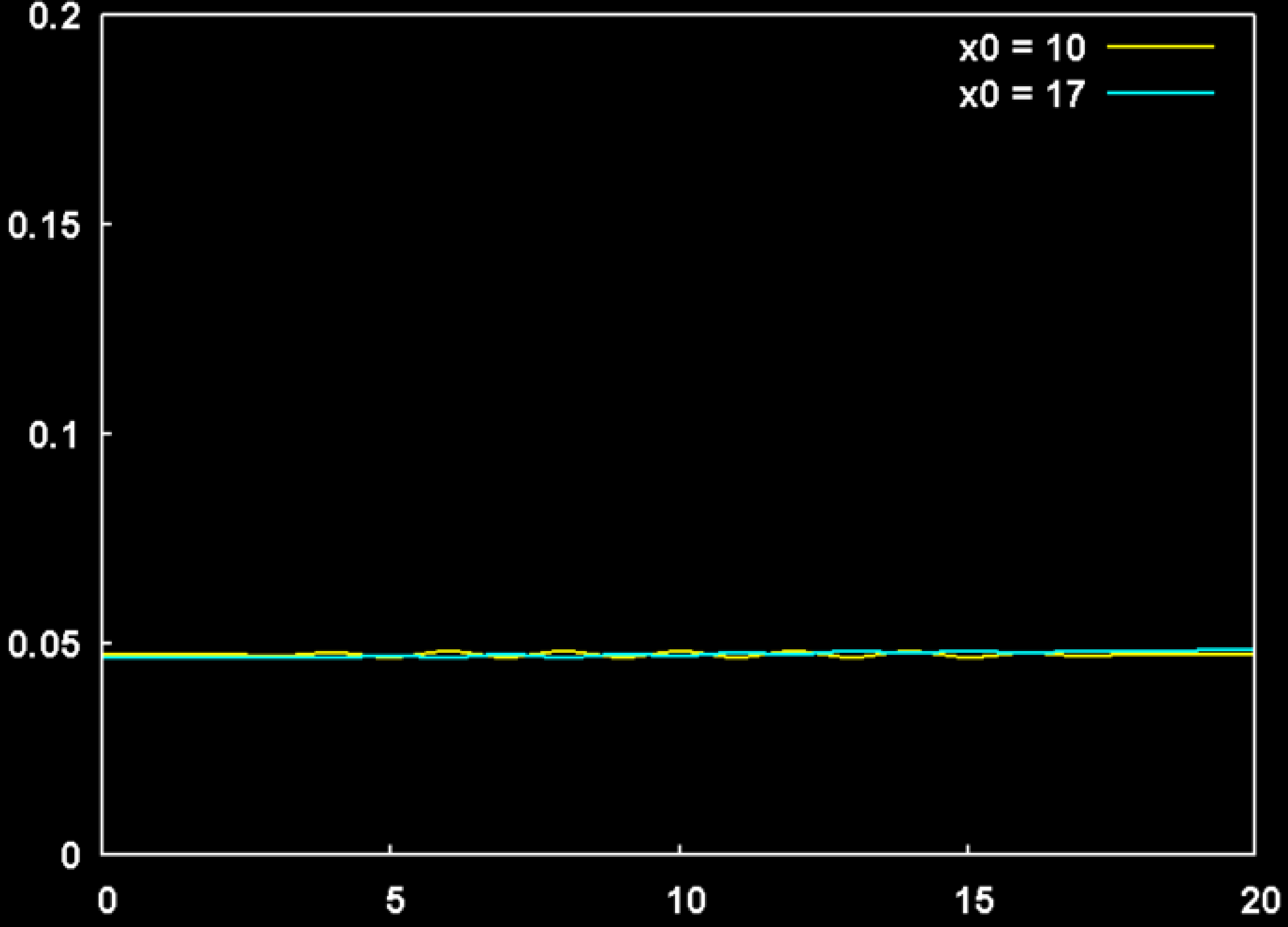

t=100
x0 = 10
x0 = 17
0.2
0.15
0.1
0.05
0
0
5
10
15
20

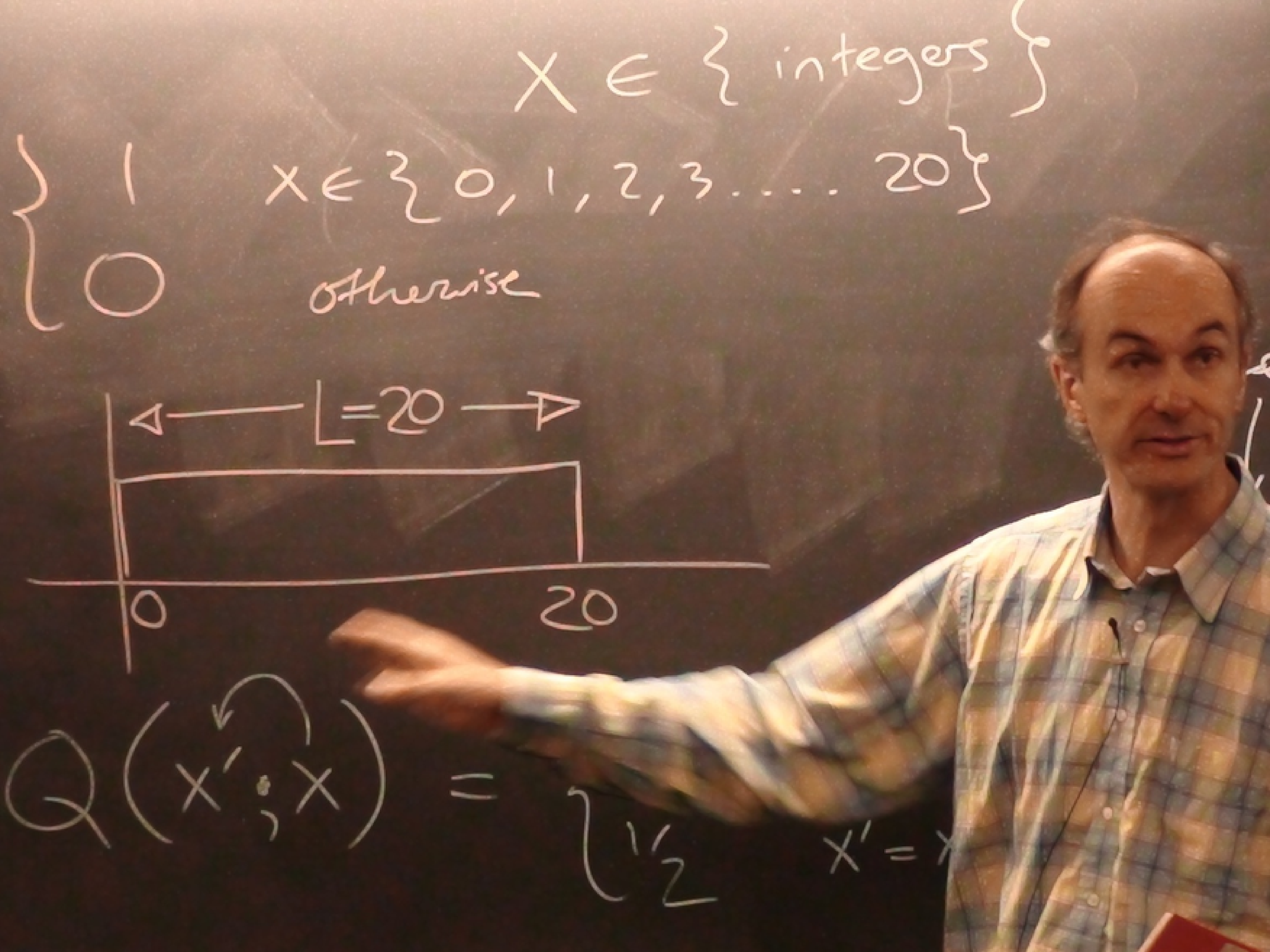

t=150
x0 = 10
x0 = 17
0.2
0.15
0.1
0.05
0
0
5
10
15
20


t=200
x0 = 10
x0 = 17
0.2
0.15
0.1
0.05
0
0
5
10
15
20


t=300
x0 = 10
x0 = 17
0.2
0.15
0.1
0.05
0
0
5
10
15
20

t=400

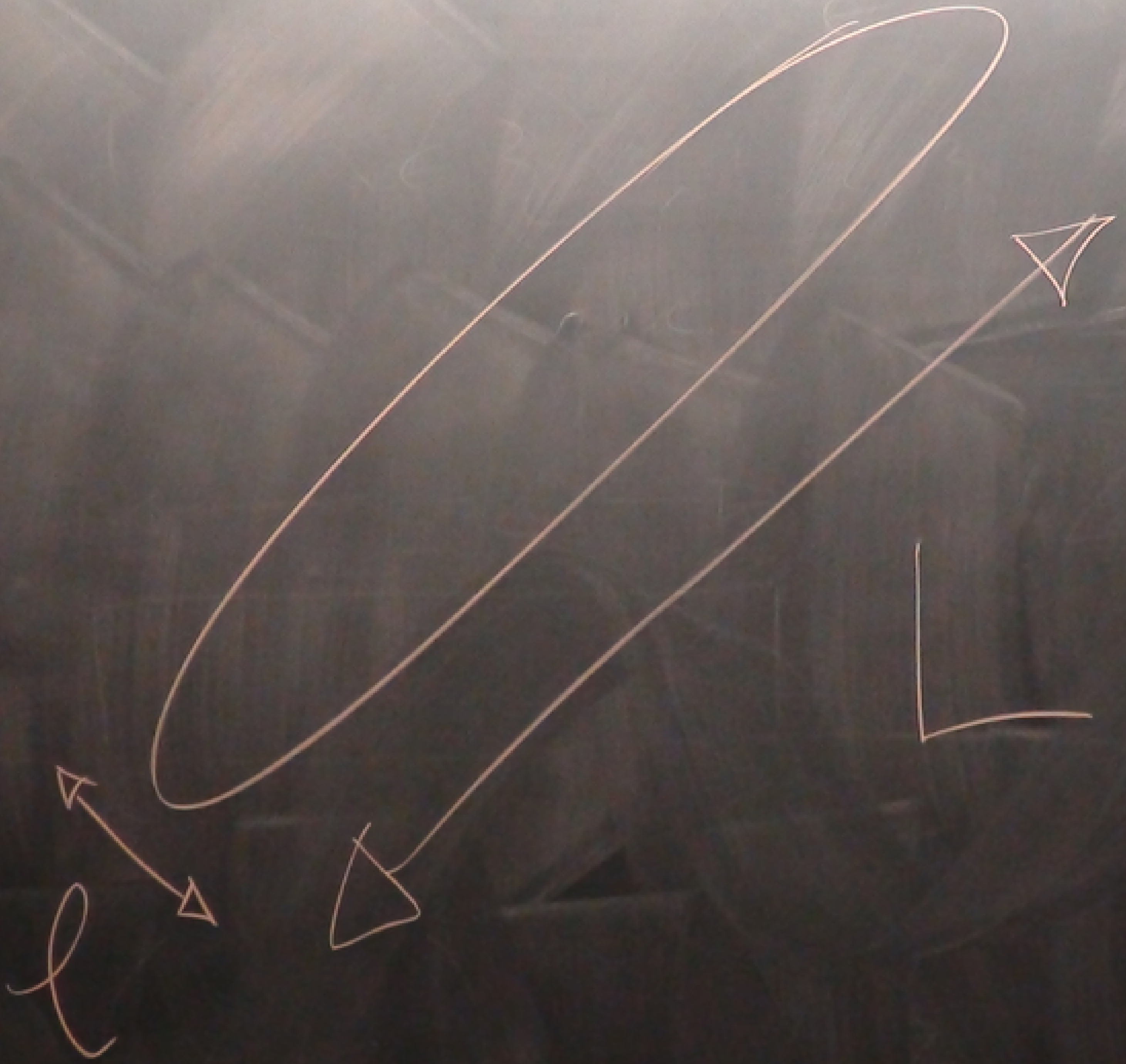


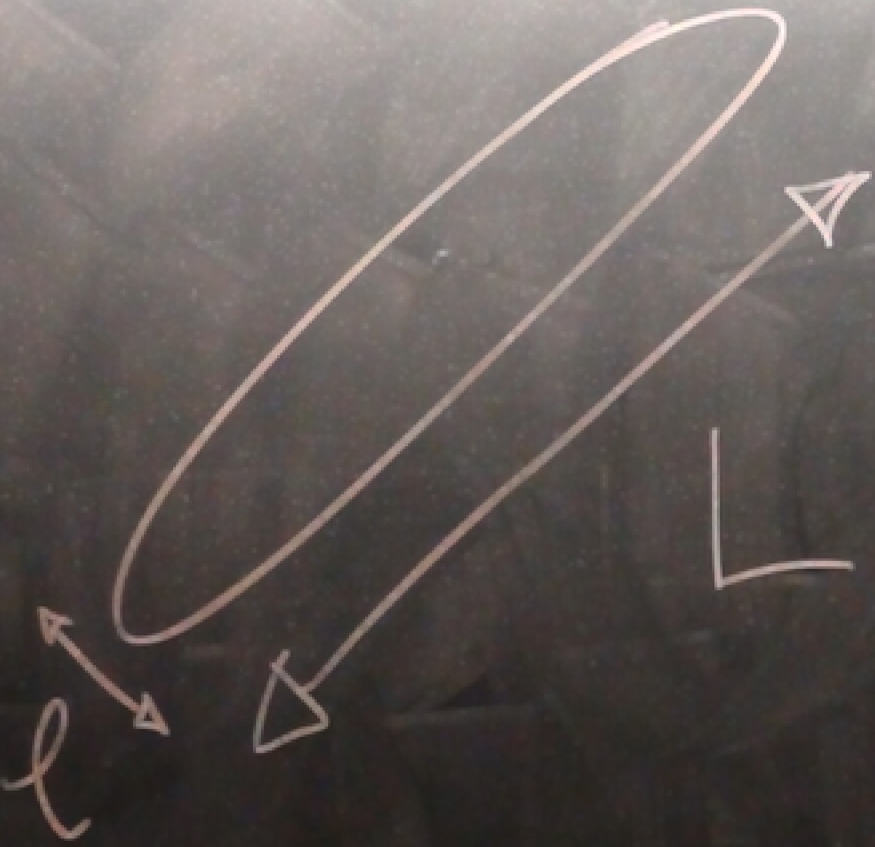


$X \in \{ \text{integers} \}$

$$\begin{cases} 1 & x \in \{0, 1, 2, 3 \ldots\ldots 20\} \\ 0 & \text{otherwise} \end{cases}$$

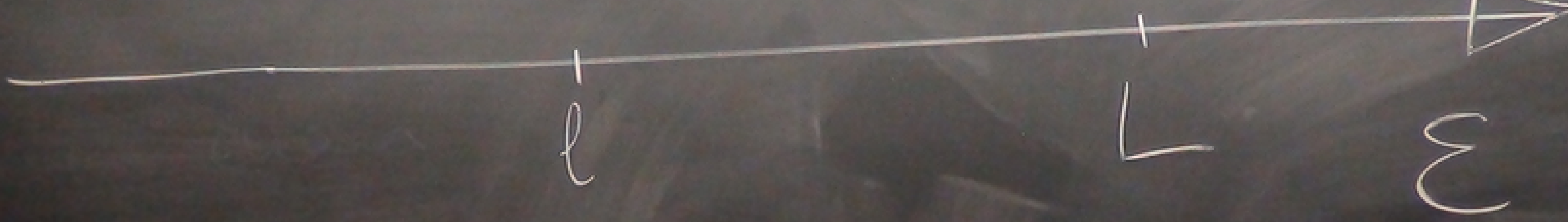
$$Q(x'; x) = \begin{cases} \tfrac{1}{2} & x' = x \end{cases}$$


L
ℓ

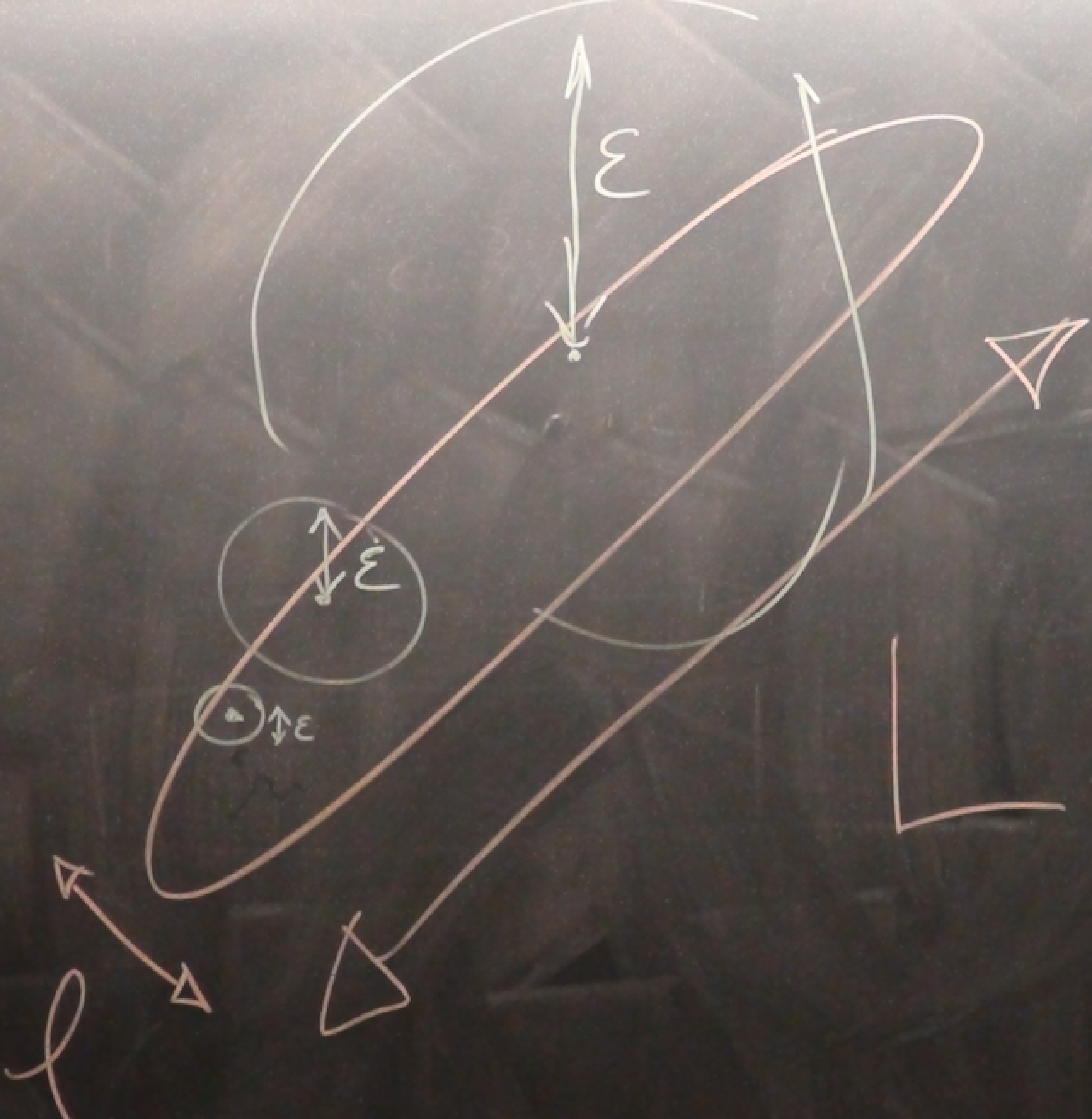


γ dimensions ℓ

$k-\gamma$ dimensions L

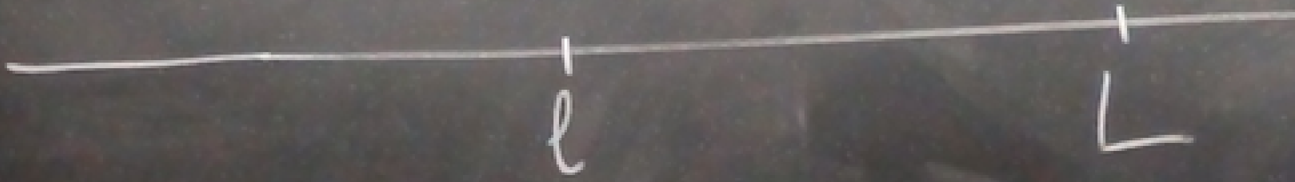


ℓ
L
ε

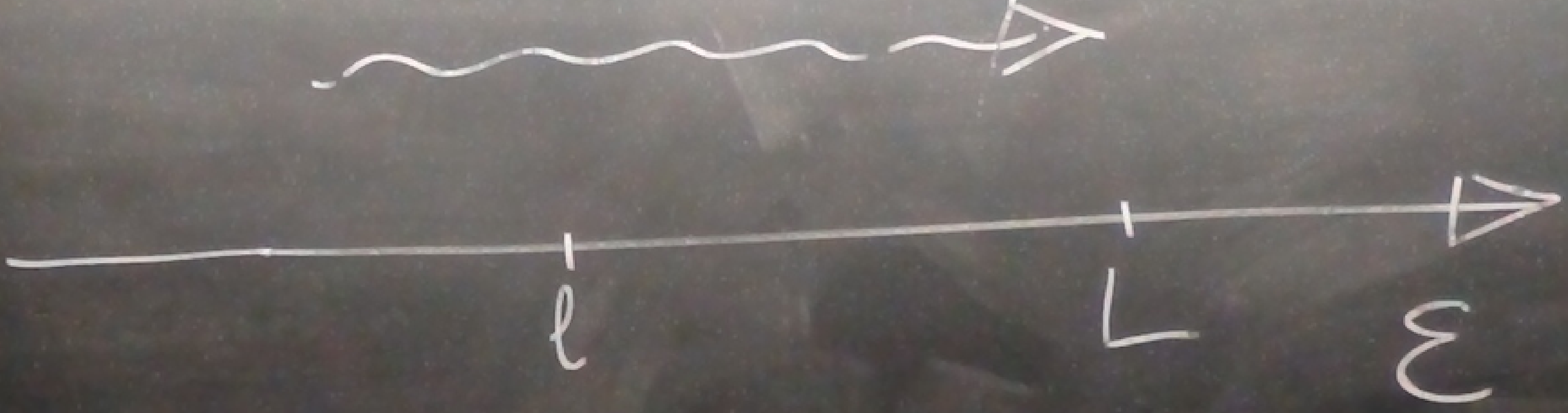

ε
ε̇
ε
L
ℓ

dimensions ℓ

dimensions L

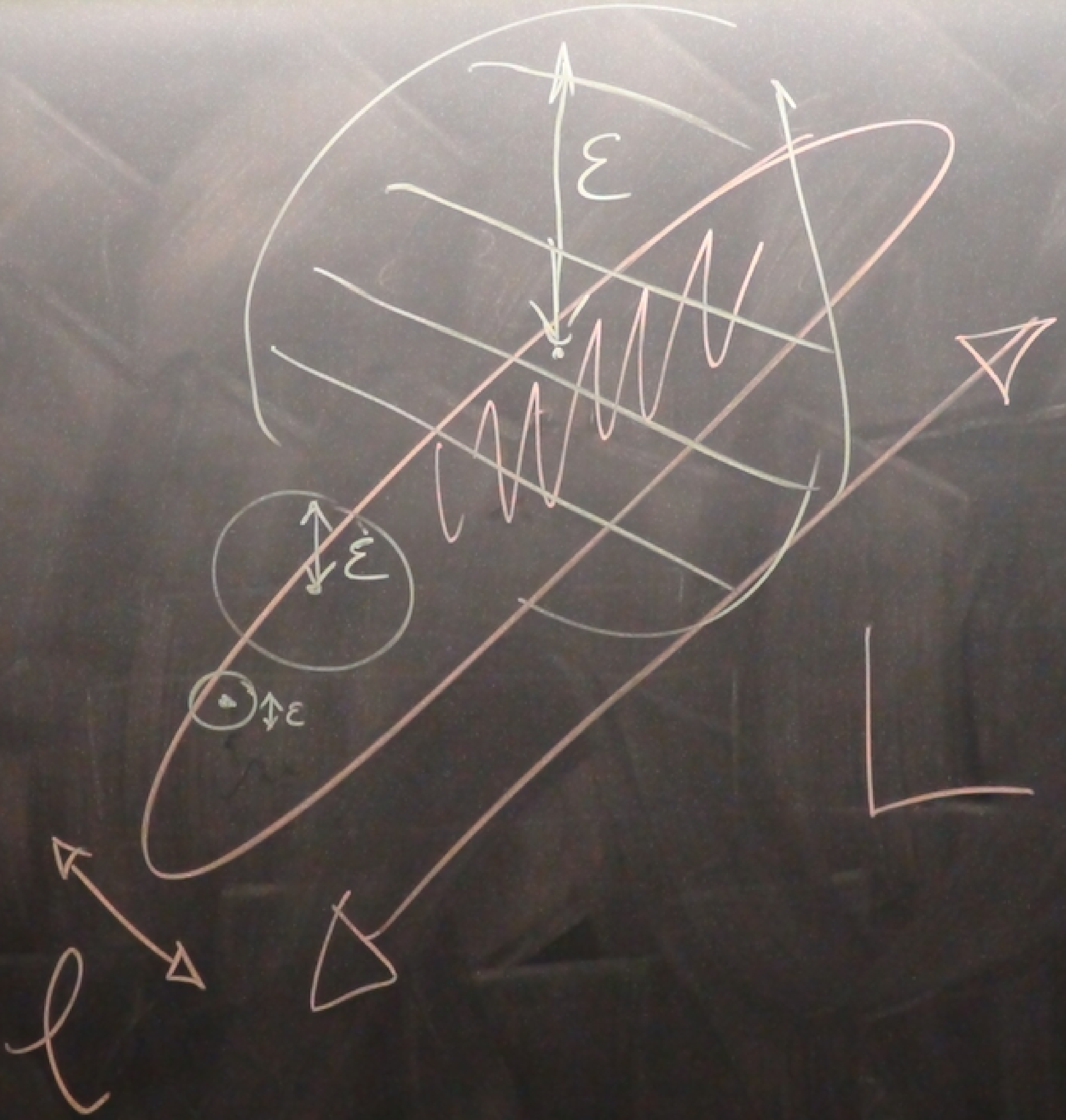


need at least $T \simeq \left(\frac{L}{\varepsilon}\right)^2$ steps to get a fresh independent sample



$\circledast$

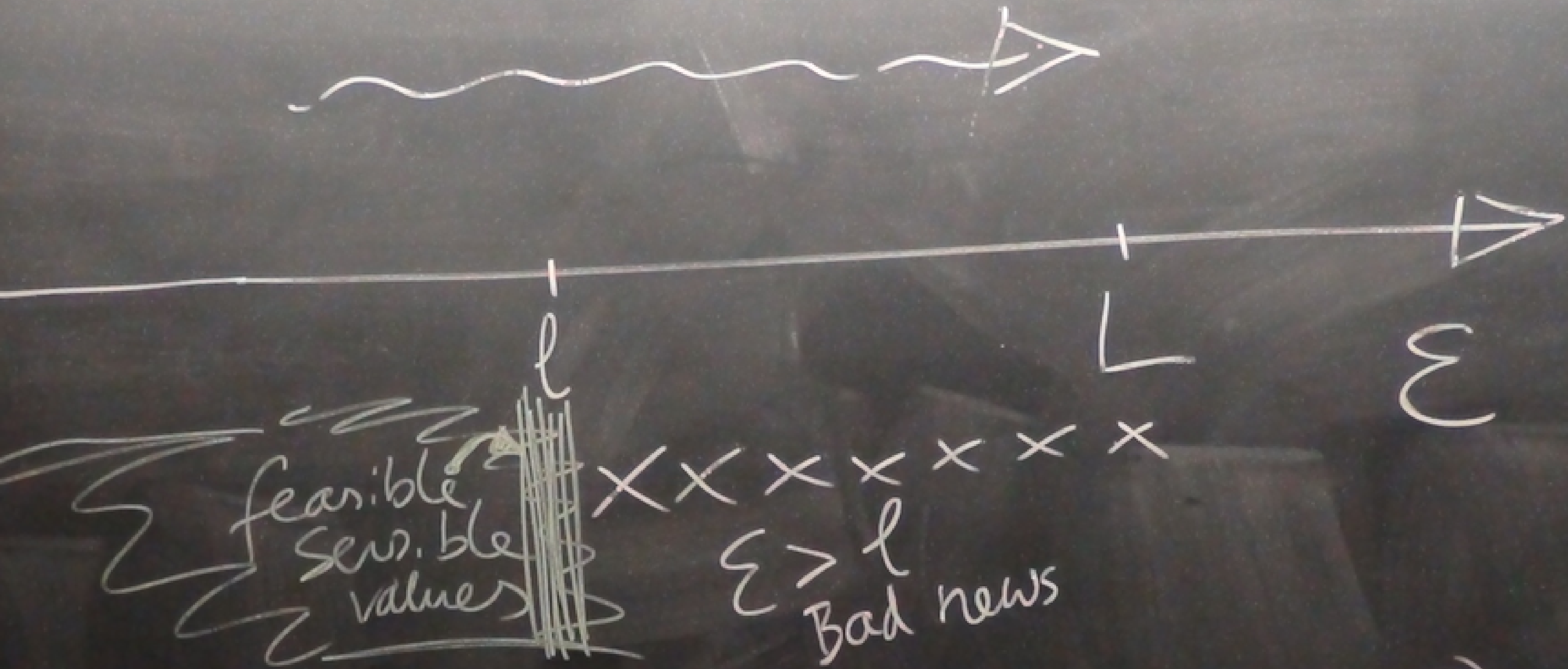
Steps to get a fresh independent sample.


ε
ε
ε
L
ℓ

if $\varepsilon \gg l$

Prob of accepted move $= \dfrac{l^{\gamma}}{\varepsilon^{\gamma}}$

$= \left(l/\varepsilon\right)^{\gamma}$



⊛

Steps to get a fresh independent sample.

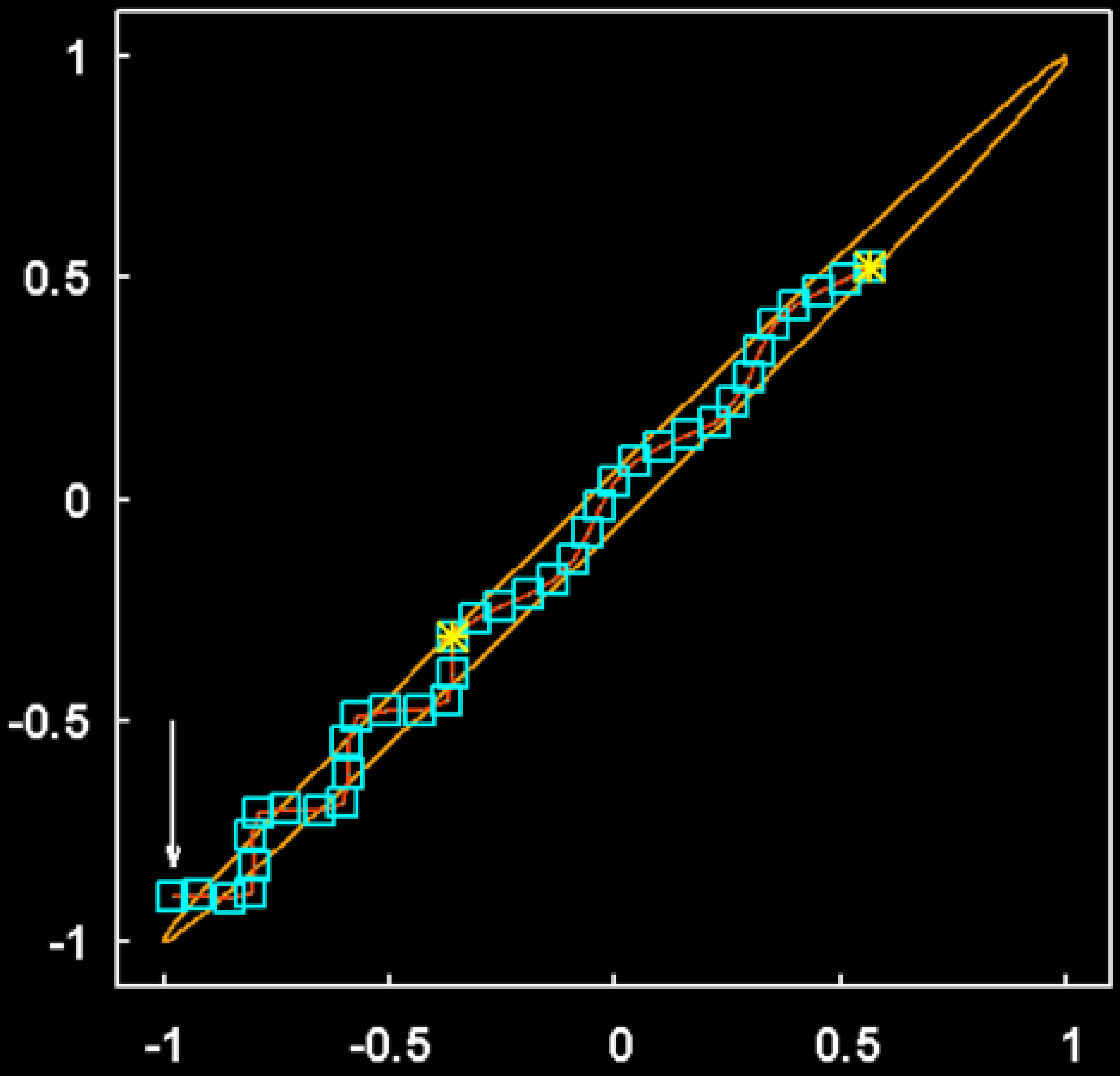
optimal $\varepsilon \simeq \ell$

& resulting optimal acceptance probability $\simeq 1/2$

Efficient Monte Carlo methods

Hamiltonian Monte Carlo

Overrelaxation


1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1

Hamiltonian MC

HMC (or Hybrid)

$$P(x, \overset{\overset{\text{momentum}}{\downarrow}}{p}) = P(x) \times \frac{e^{-\frac{1}{2} p^2/m}}{Z_p}$$

$$= \frac{e^{-\left\{\overset{\downarrow}{E}(x) + \frac{1}{2} p^2/m\right\}}}{Z}$$

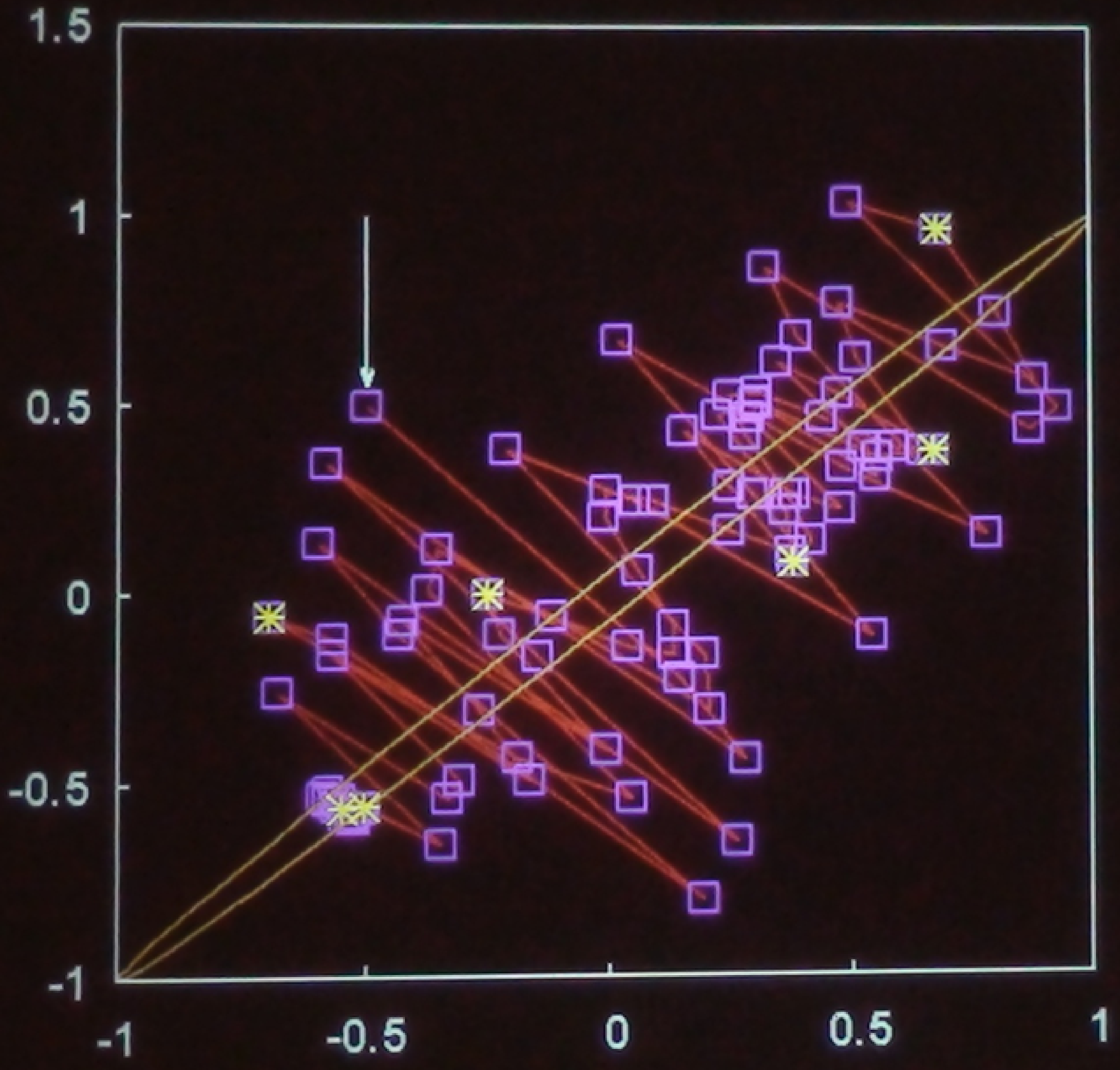
$$\dot{x} \sim p$$

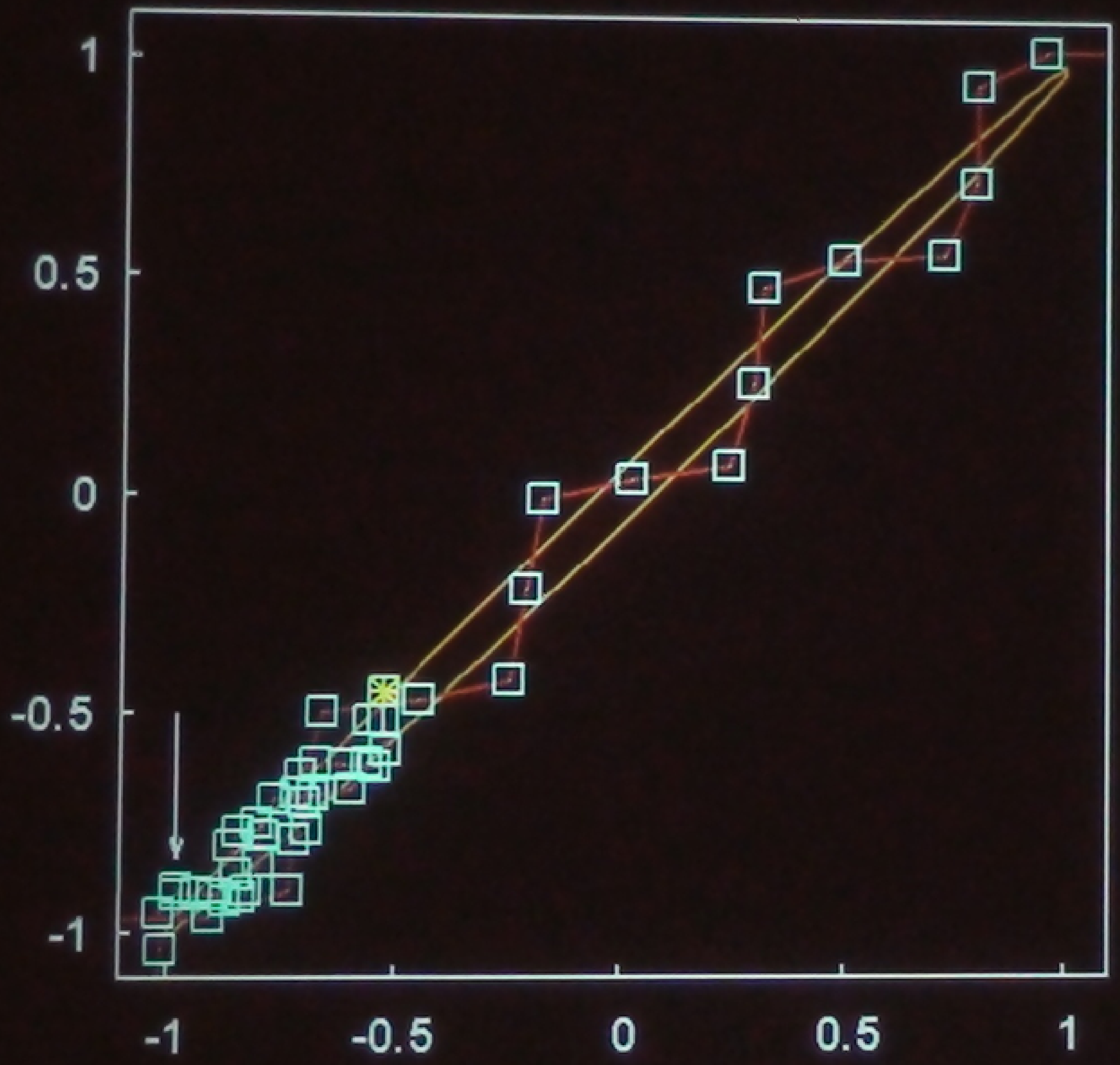
$$\dot{p} \sim \frac{\partial E}{\partial \underline{x}}$$

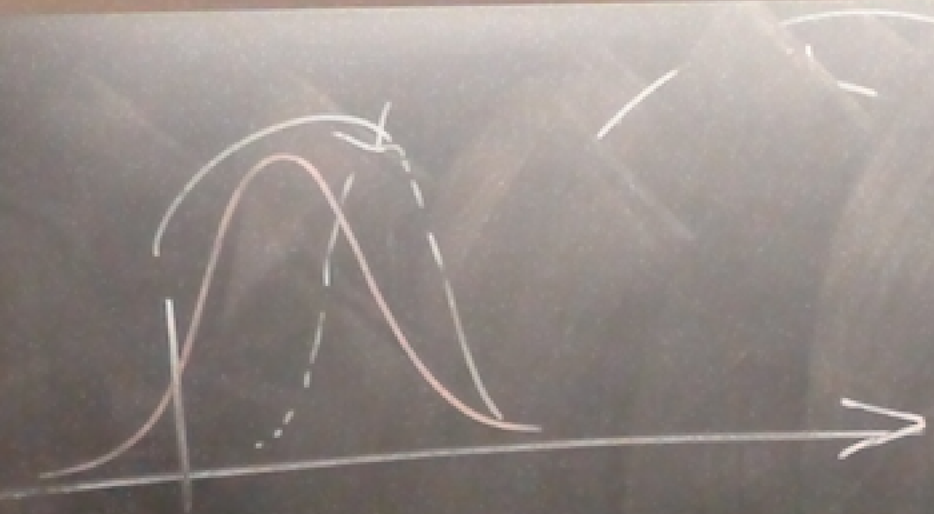
1. Simulate Newton's law (for a time) (approximately)

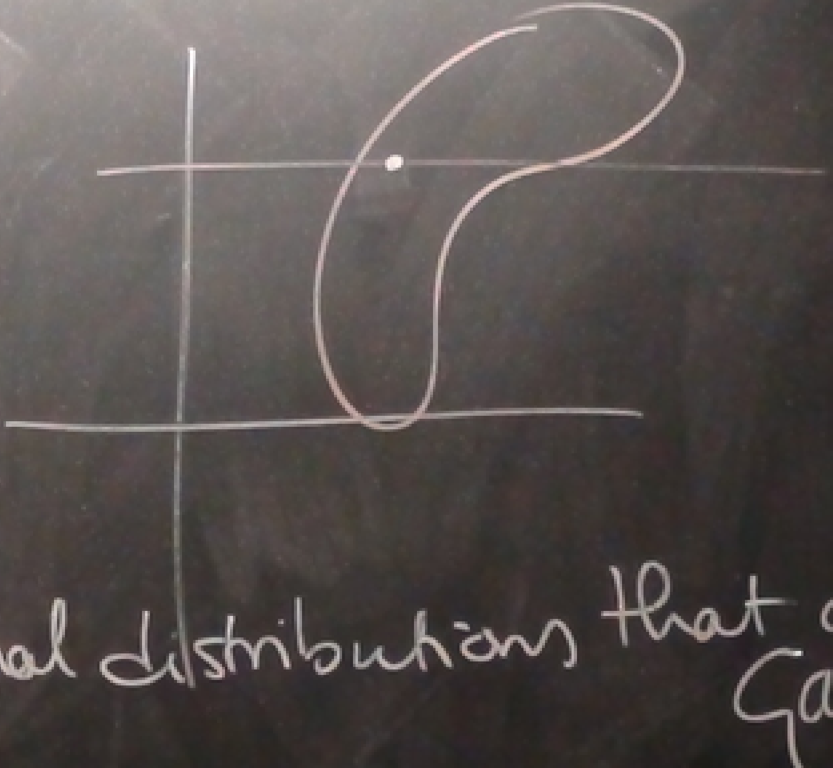
2. accept/reject based on change in total energy

3. randomize the momentum $p \sim e^{-\frac{1}{2}p^2/m}$


1.5
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1


1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1

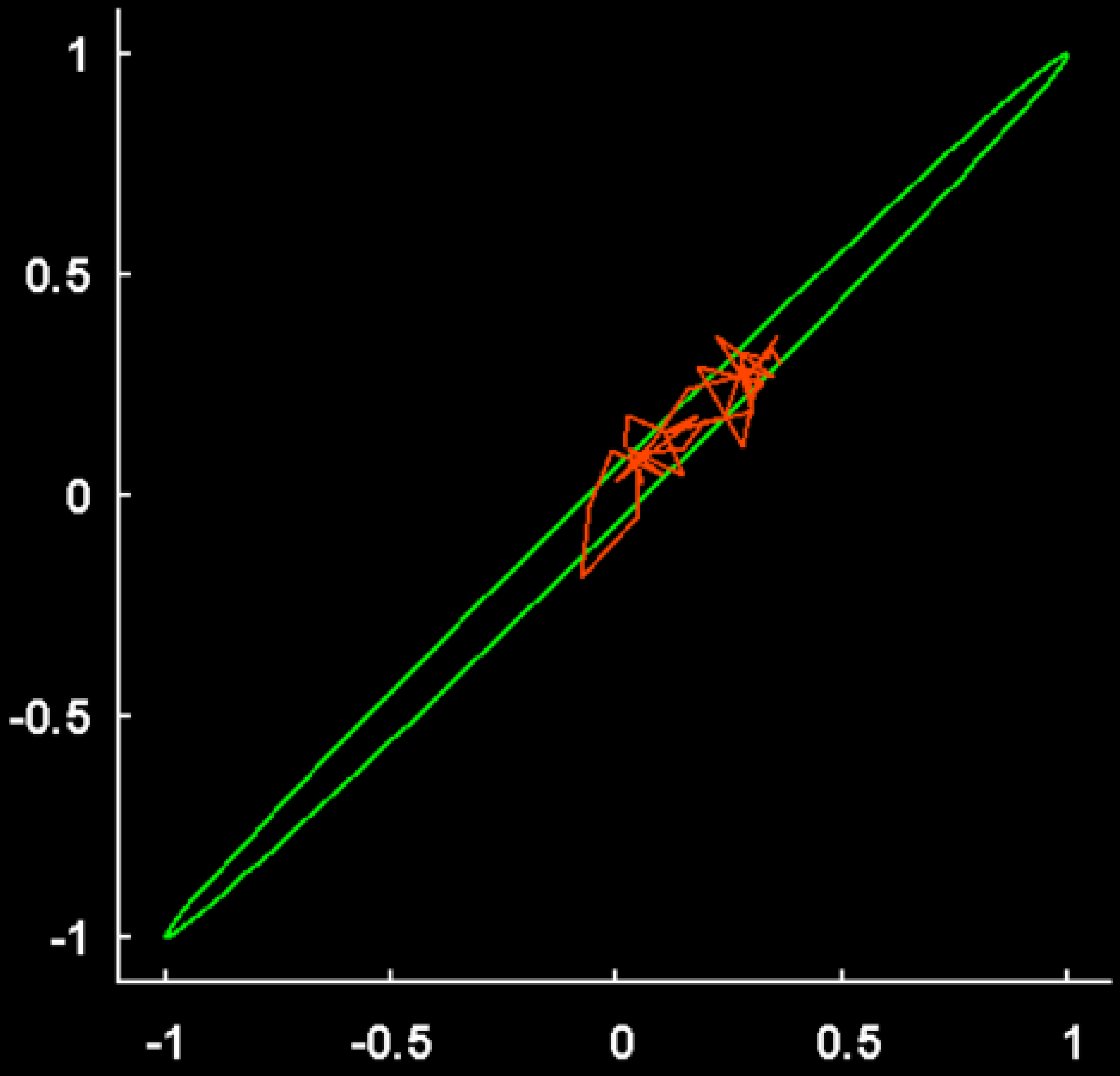




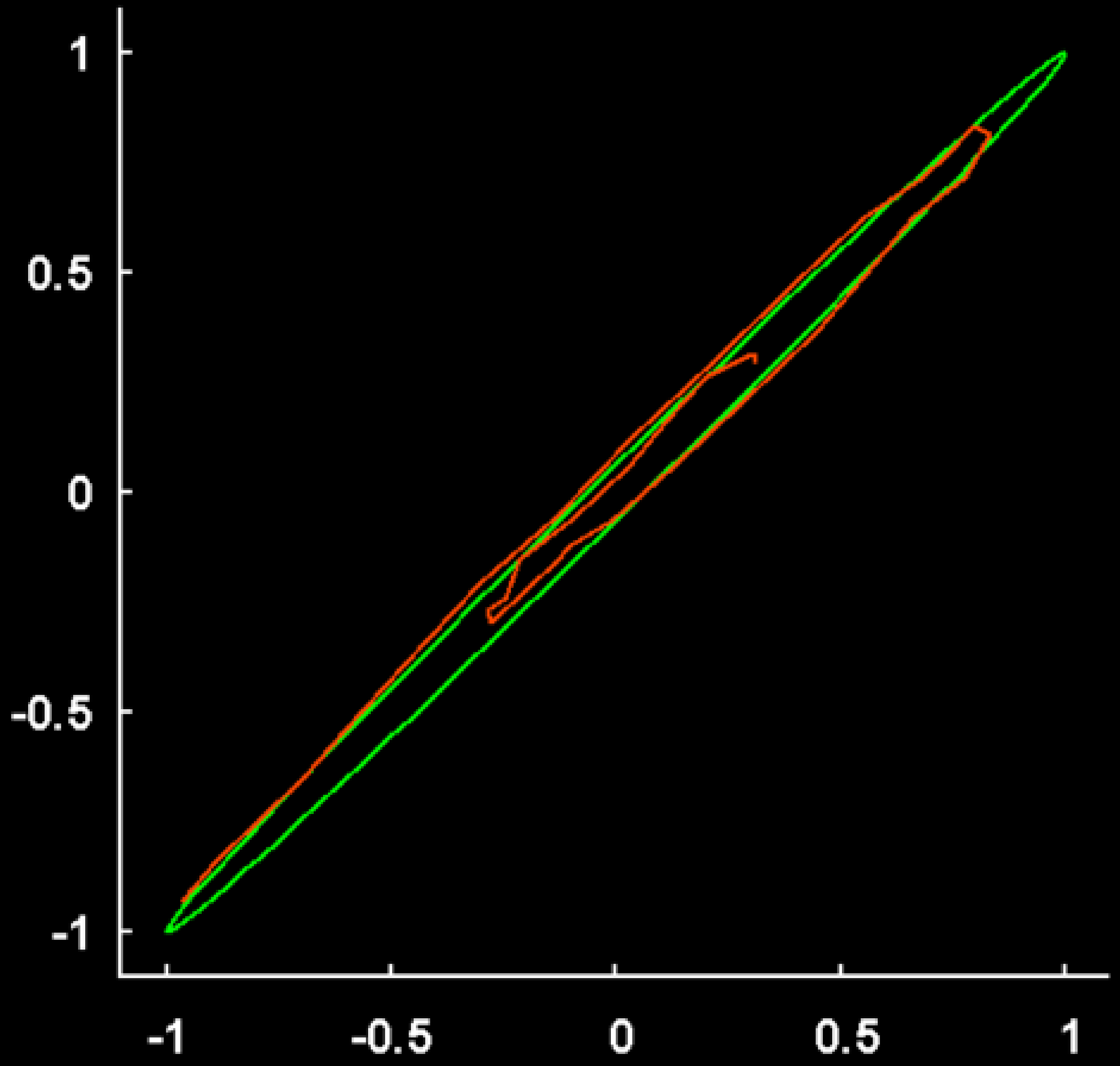
Adler's overrelaxation

for conditional distributions that are Gaussians.

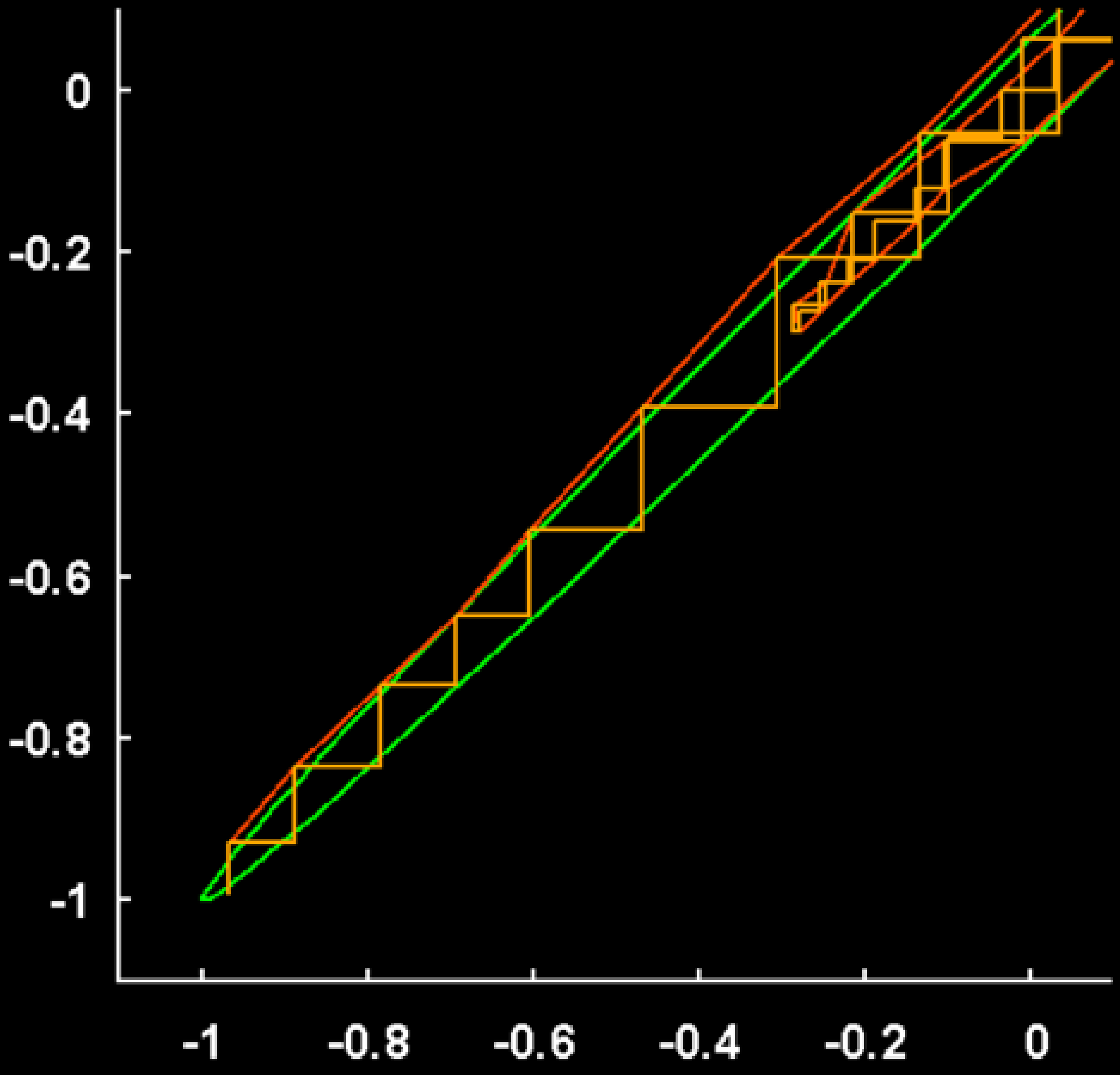
Gibbs sampling

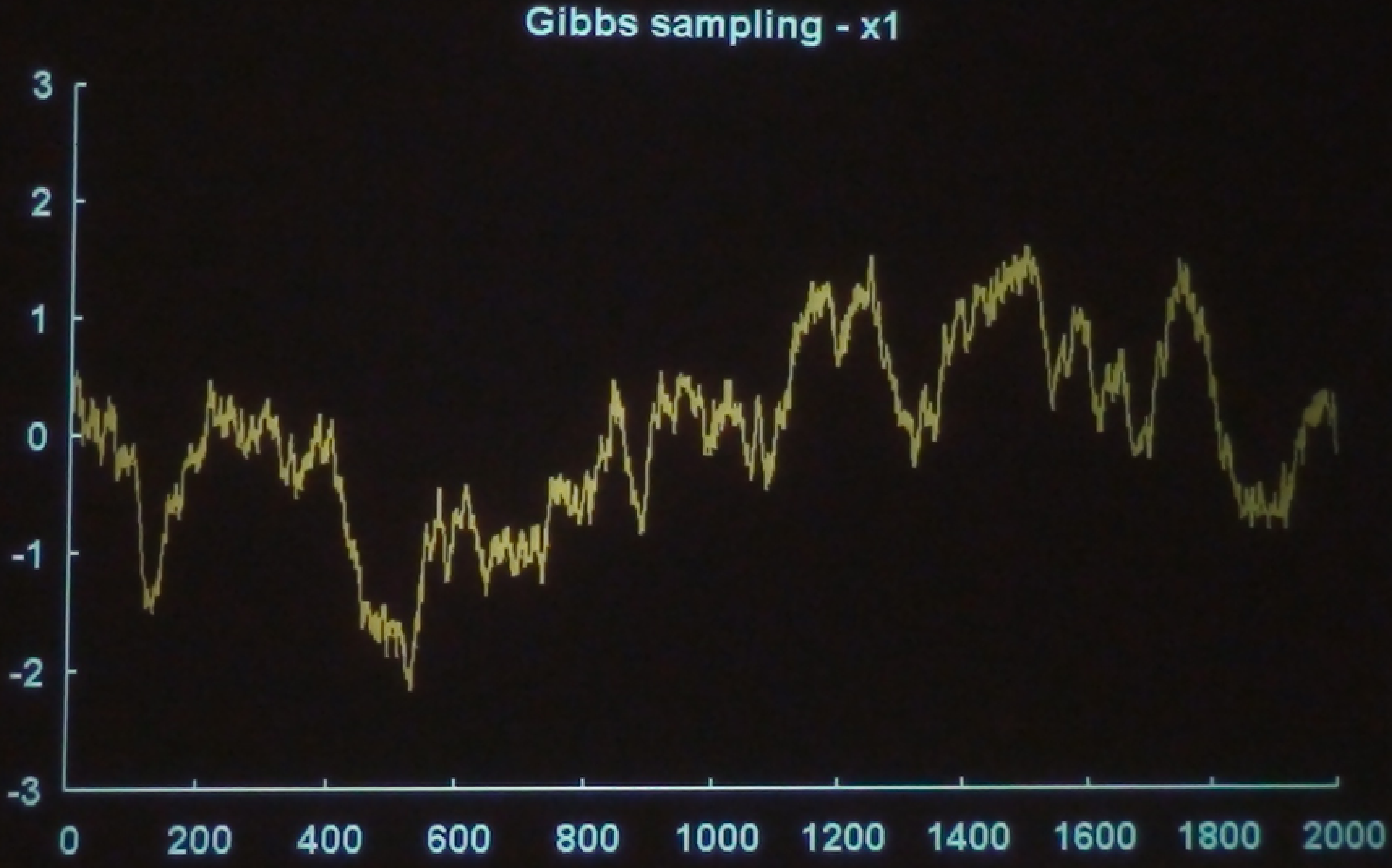


Adler's Overrelaxation



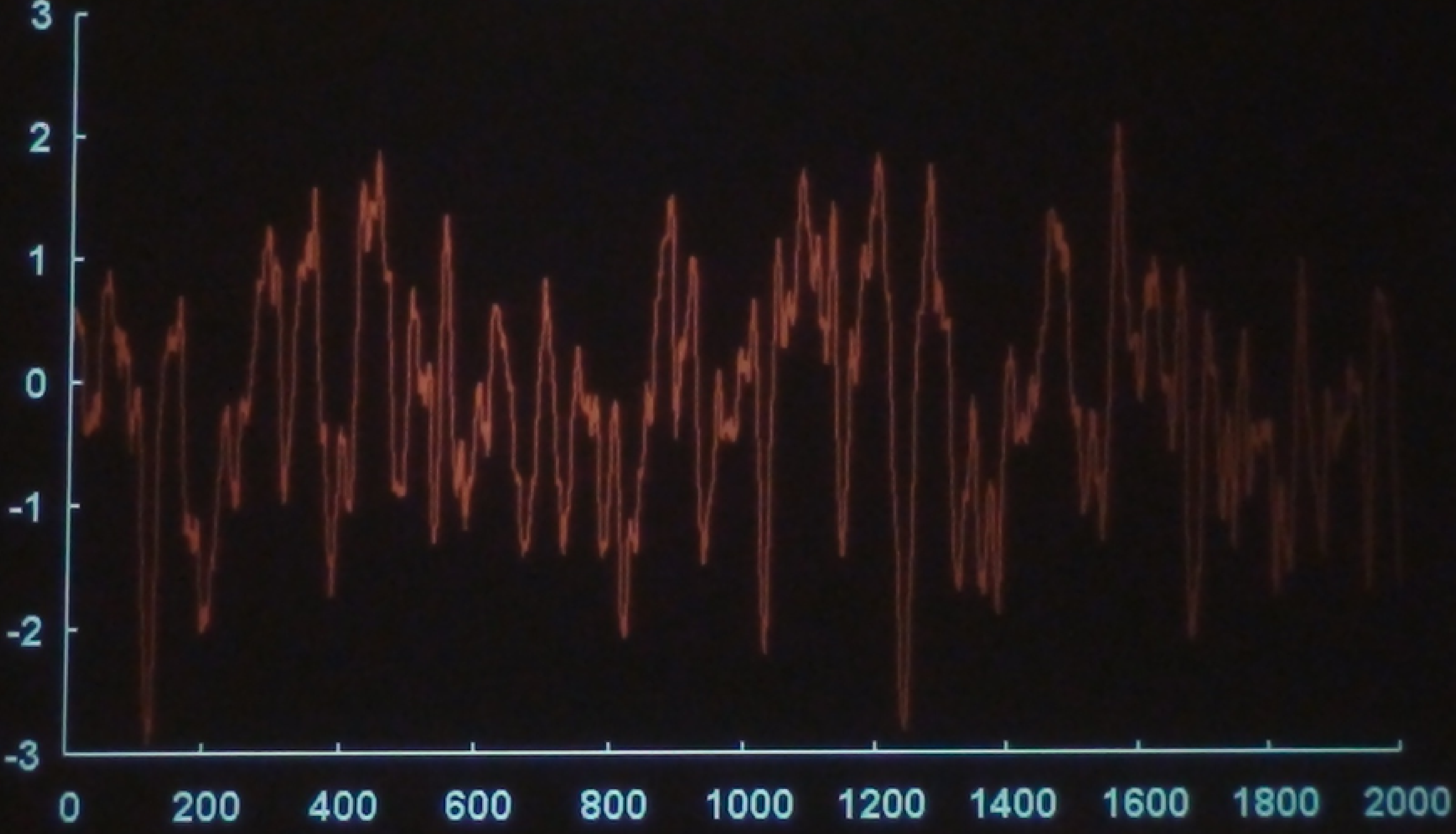
Adler's Overrelaxation



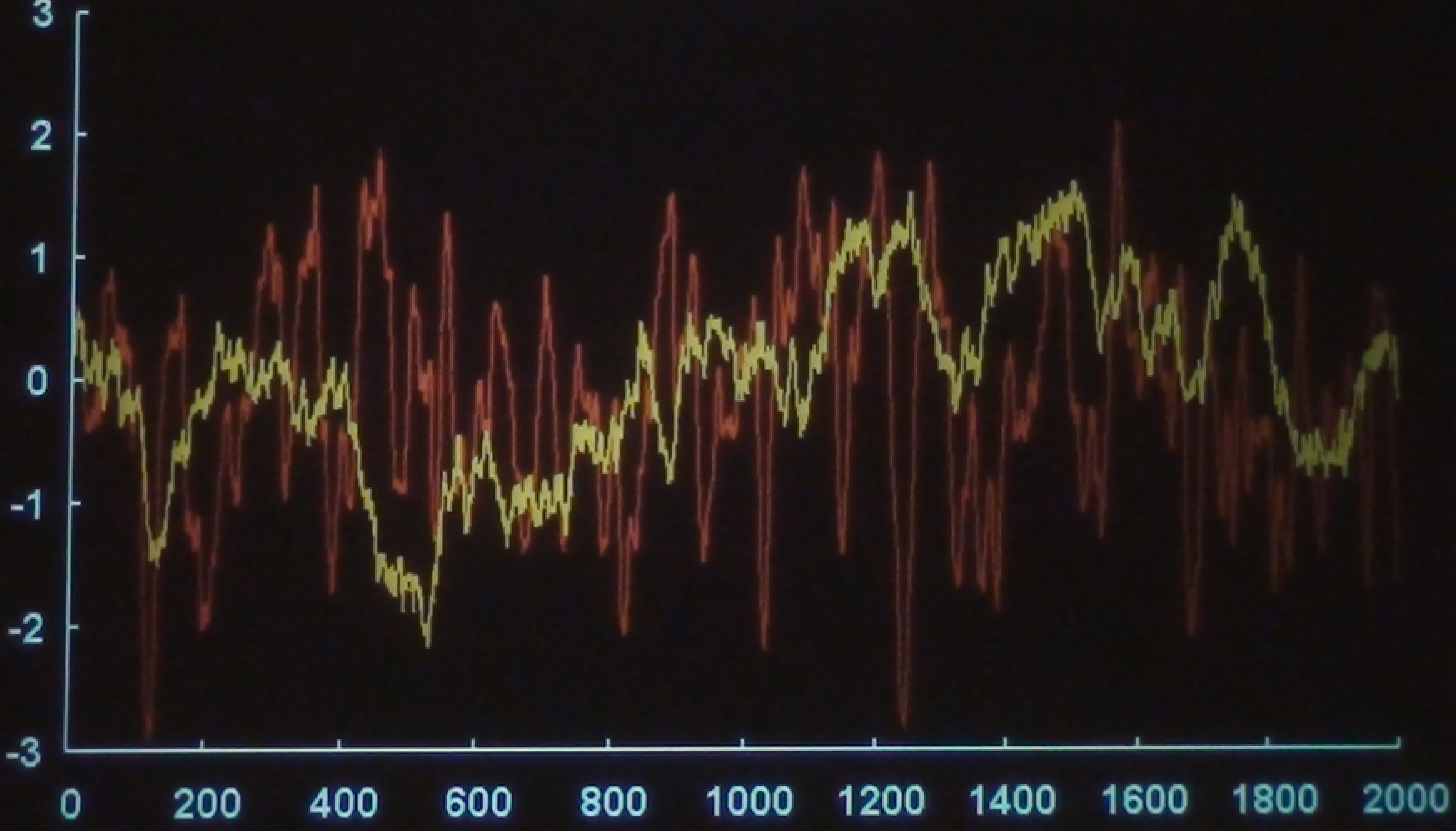


Gibbs sampling - x1

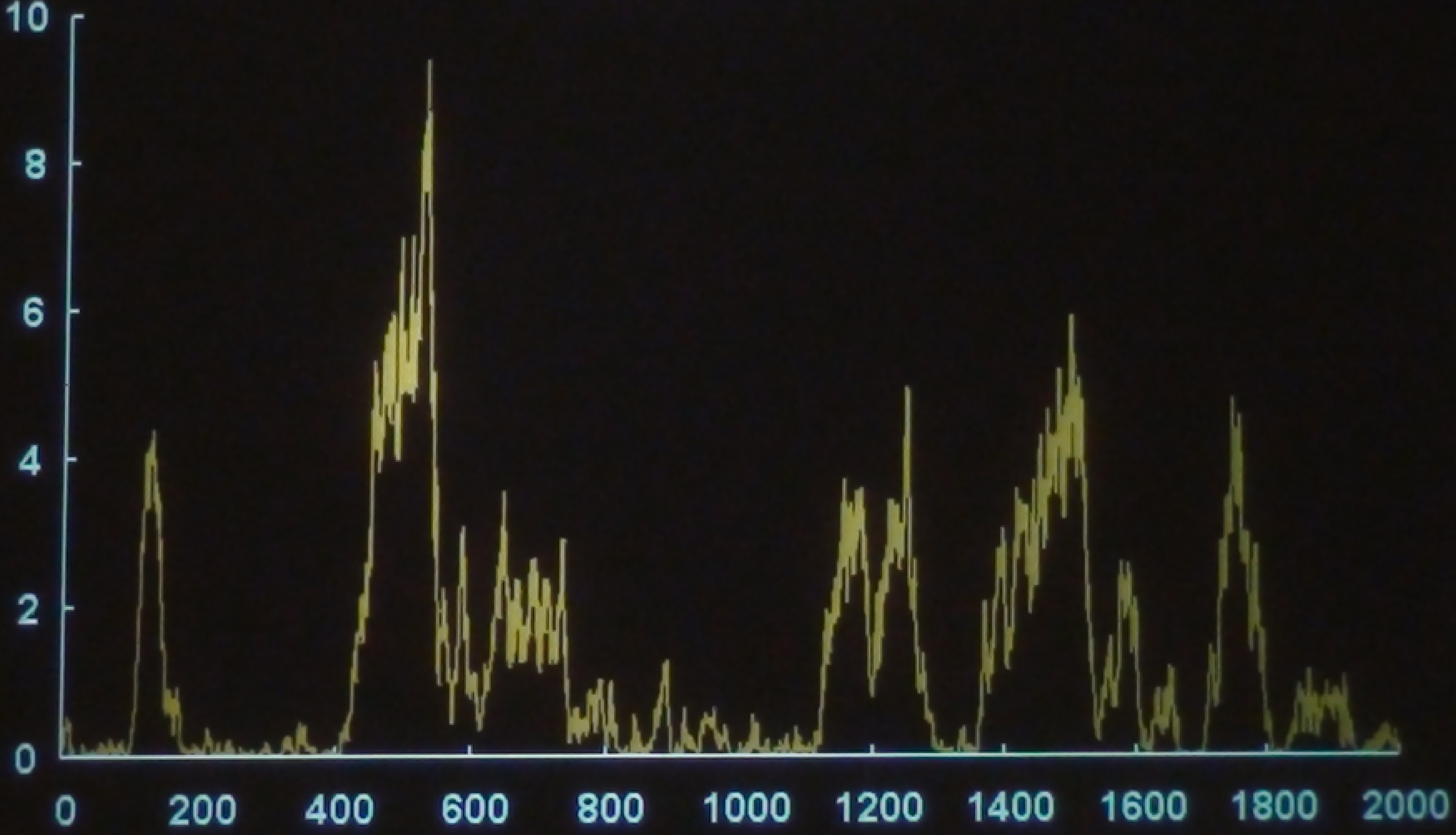
Adler's Overrelaxation - x1



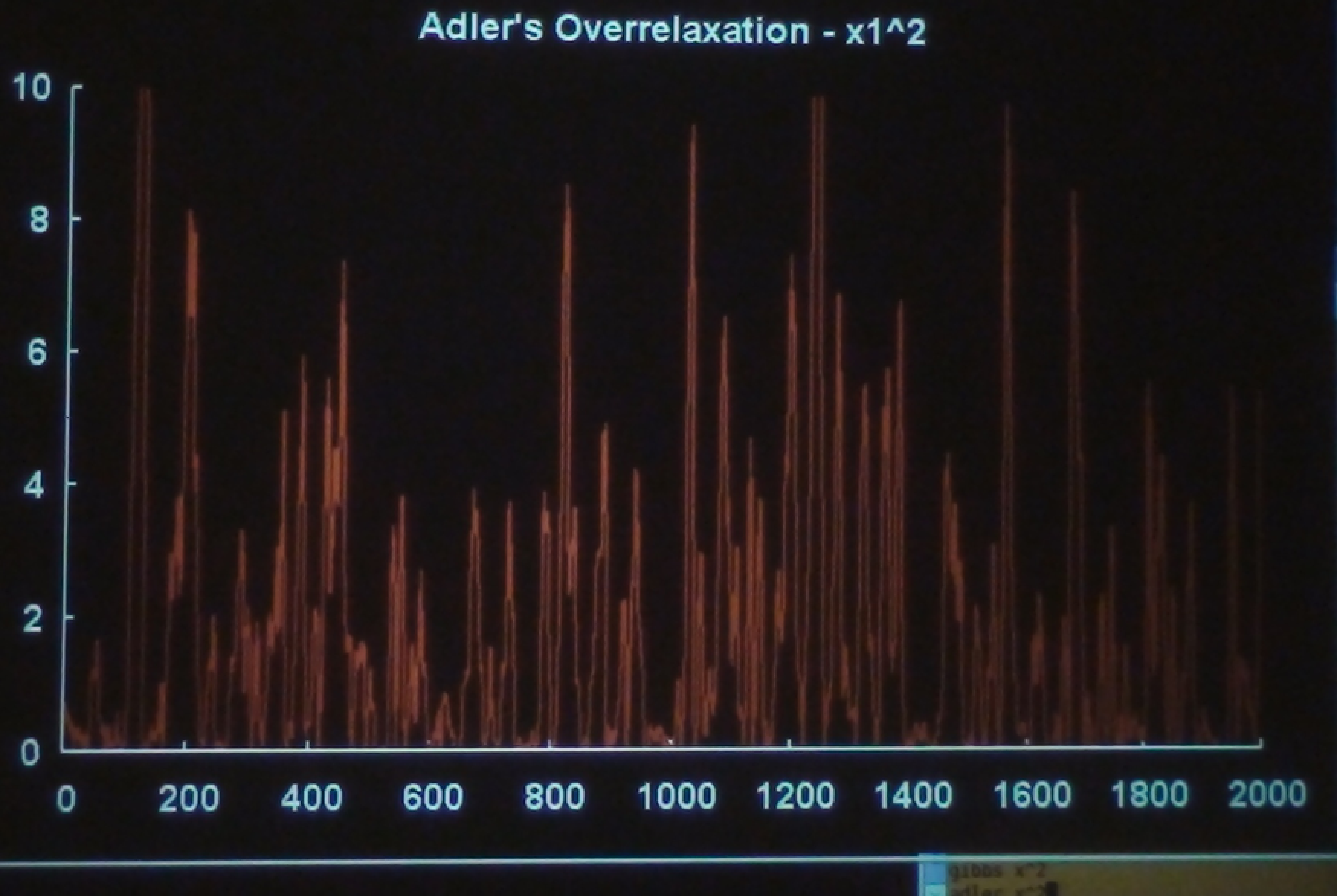
Adler's Overrelaxation - x1

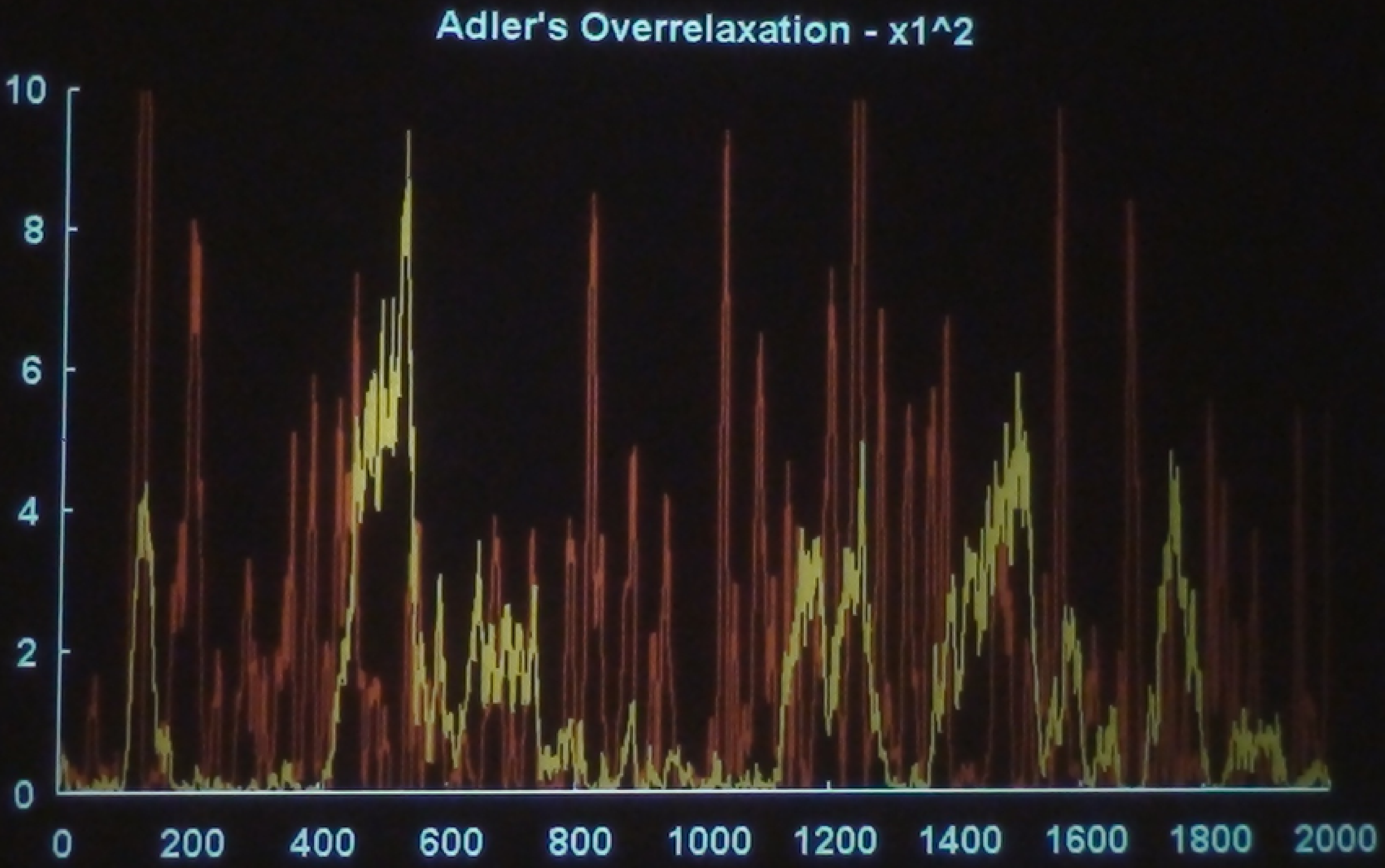


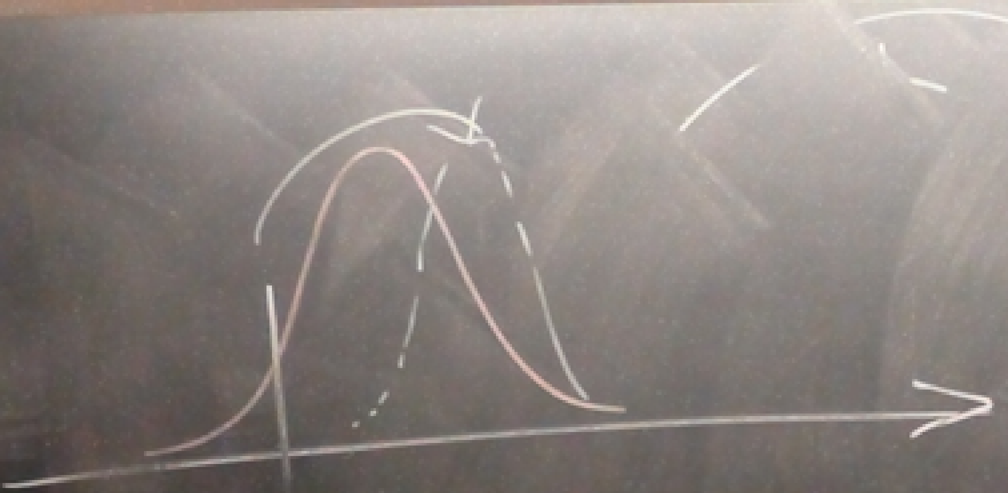
Gibbs sampling - x1^2

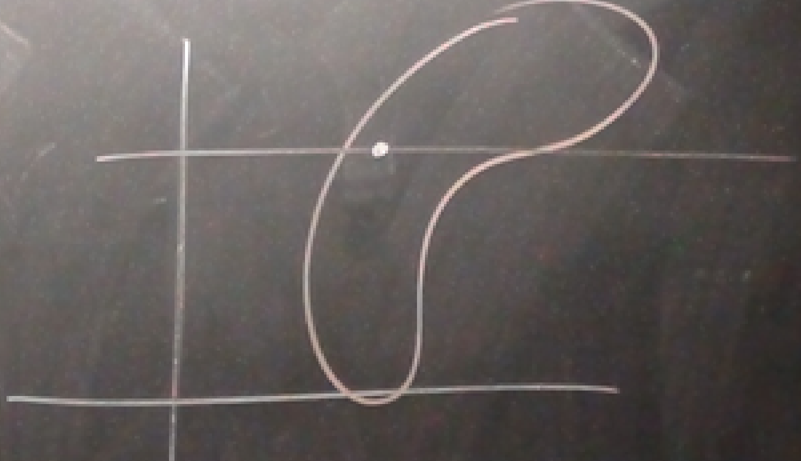


adler and gibbs x1


Adler's Overrelaxation - x1^2
0
2
4
6
8
10
0
200
400
600
800
1000
1200
1400
1600
1800
2000
gibbs x^2
adler x^2


Adler's Overrelaxation - x1^2
10
8
6
4
2
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

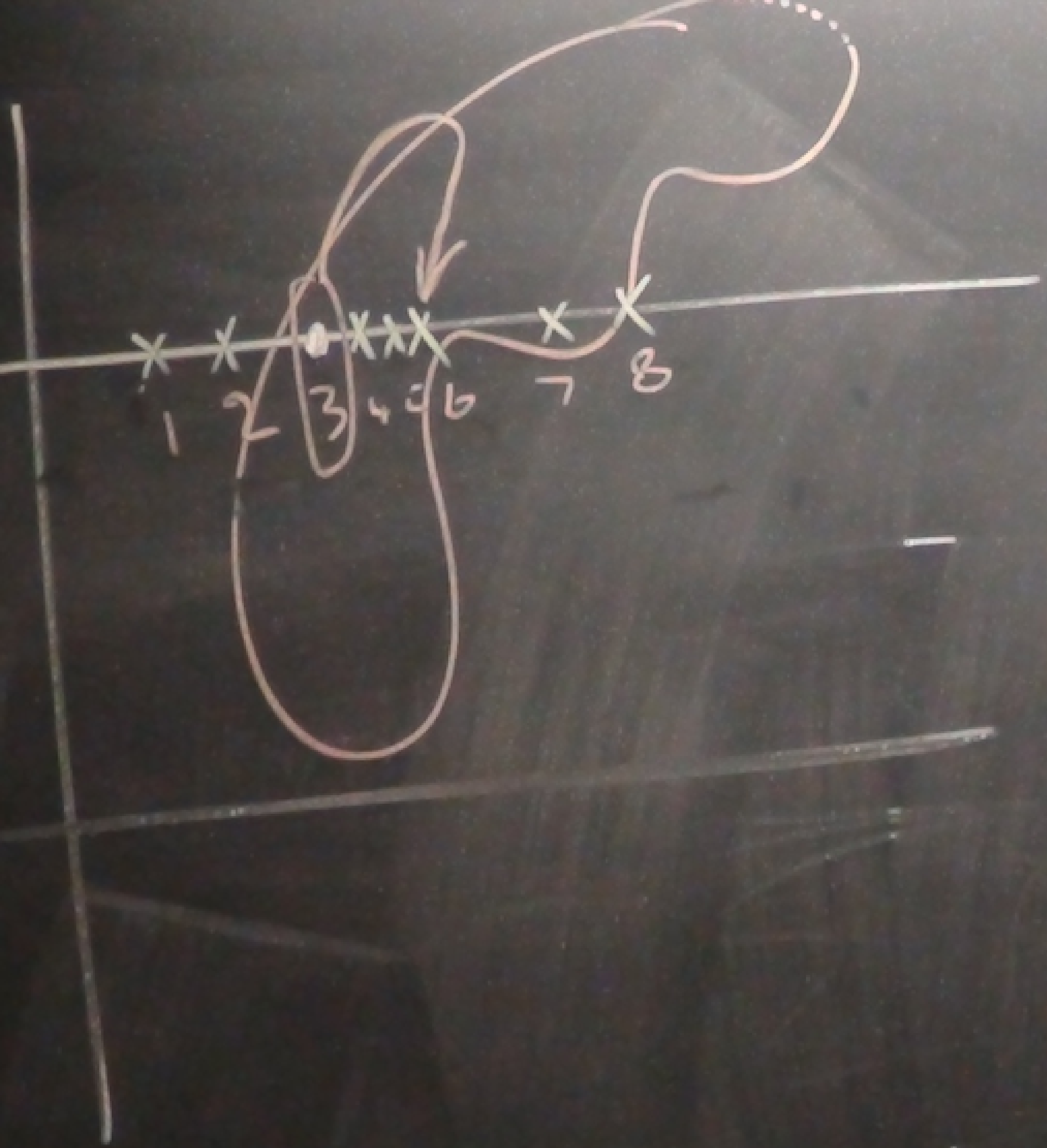




Adler's overrelaxation

for conditional distributions that are Gaussians.

Ordered overrelaxation Radford Neal



draw

K

times

$K \simeq 20$

BUGS

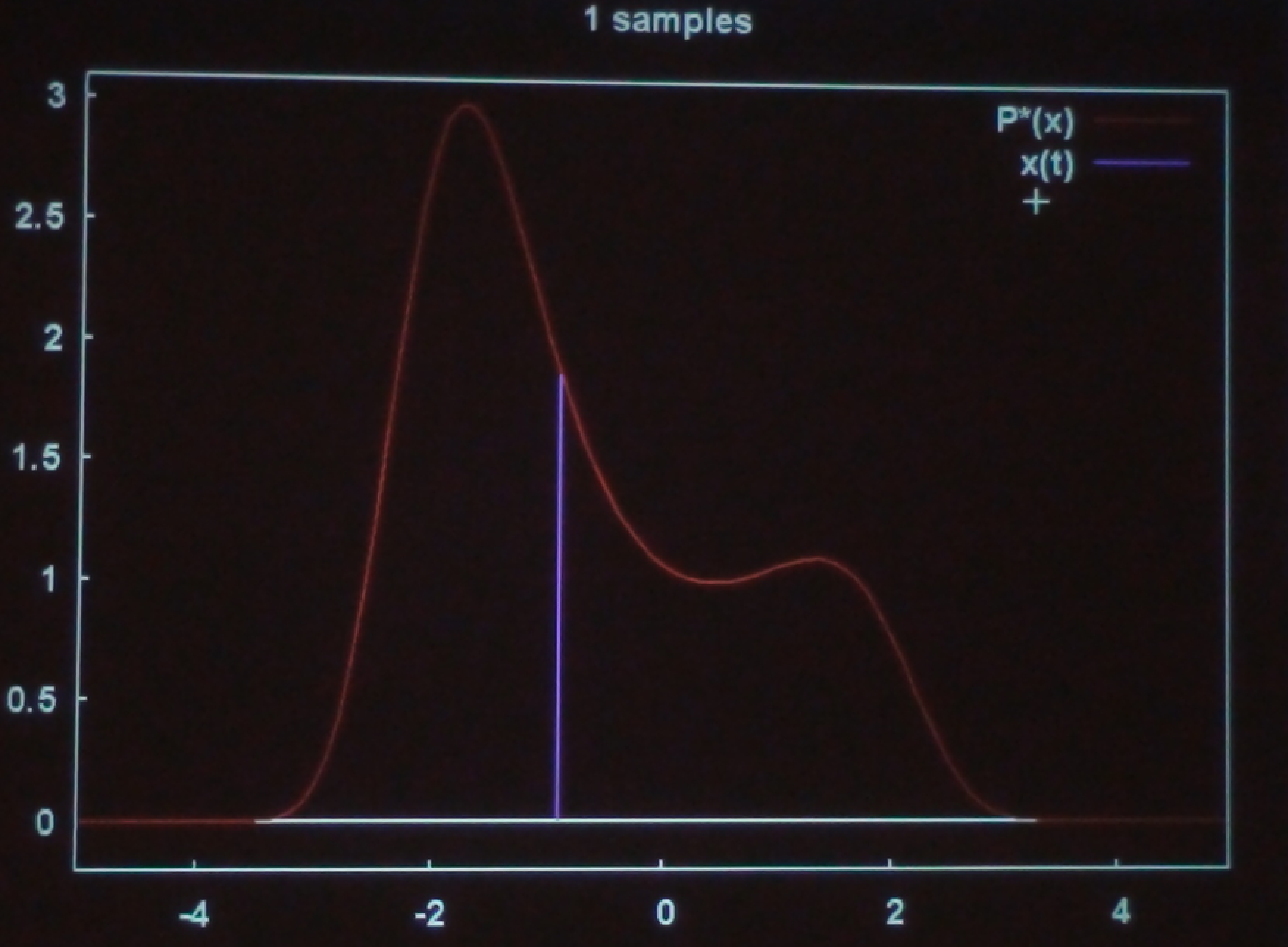
Robust Monte Carlo methods

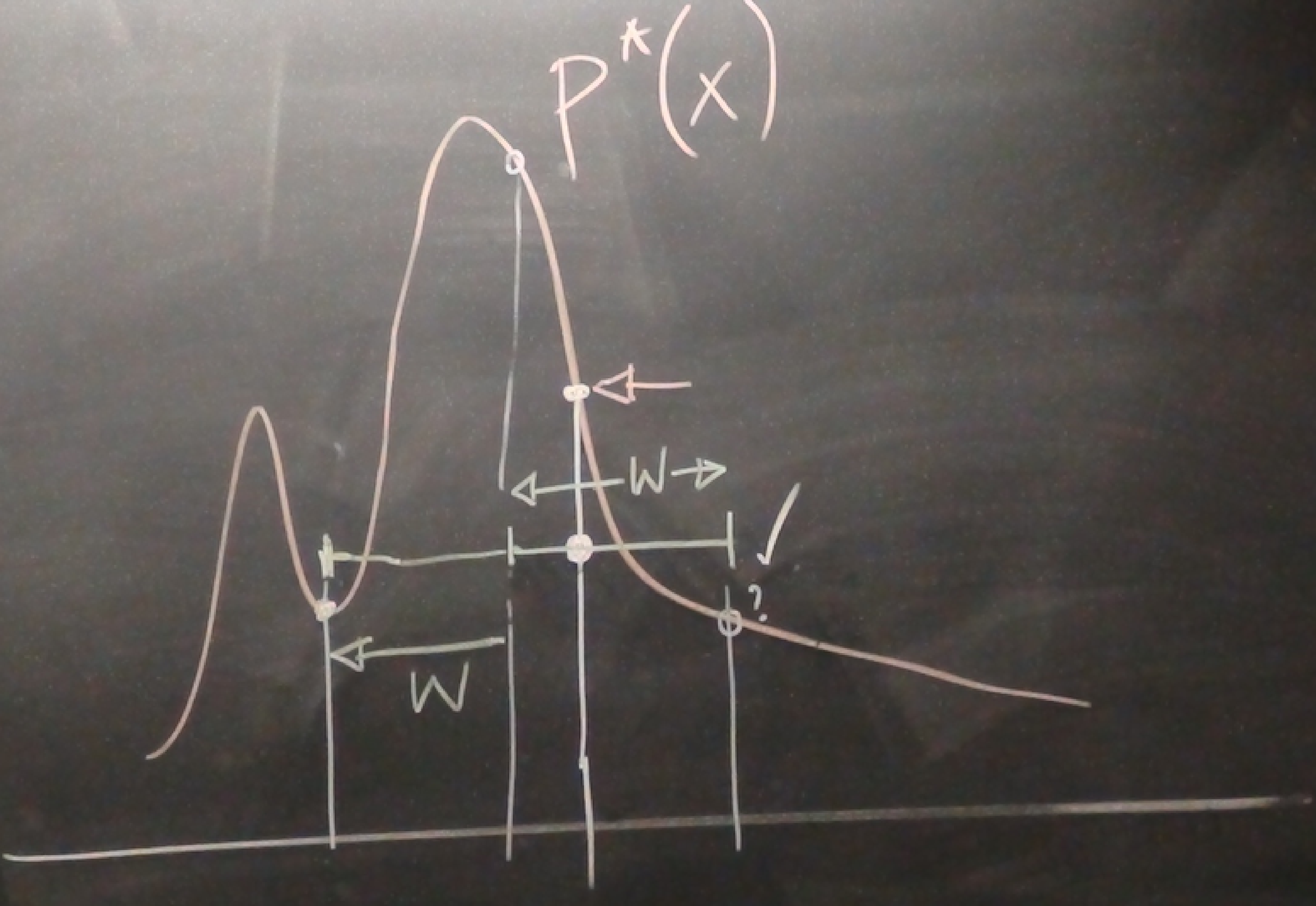
Slice sampling

Slice sampling

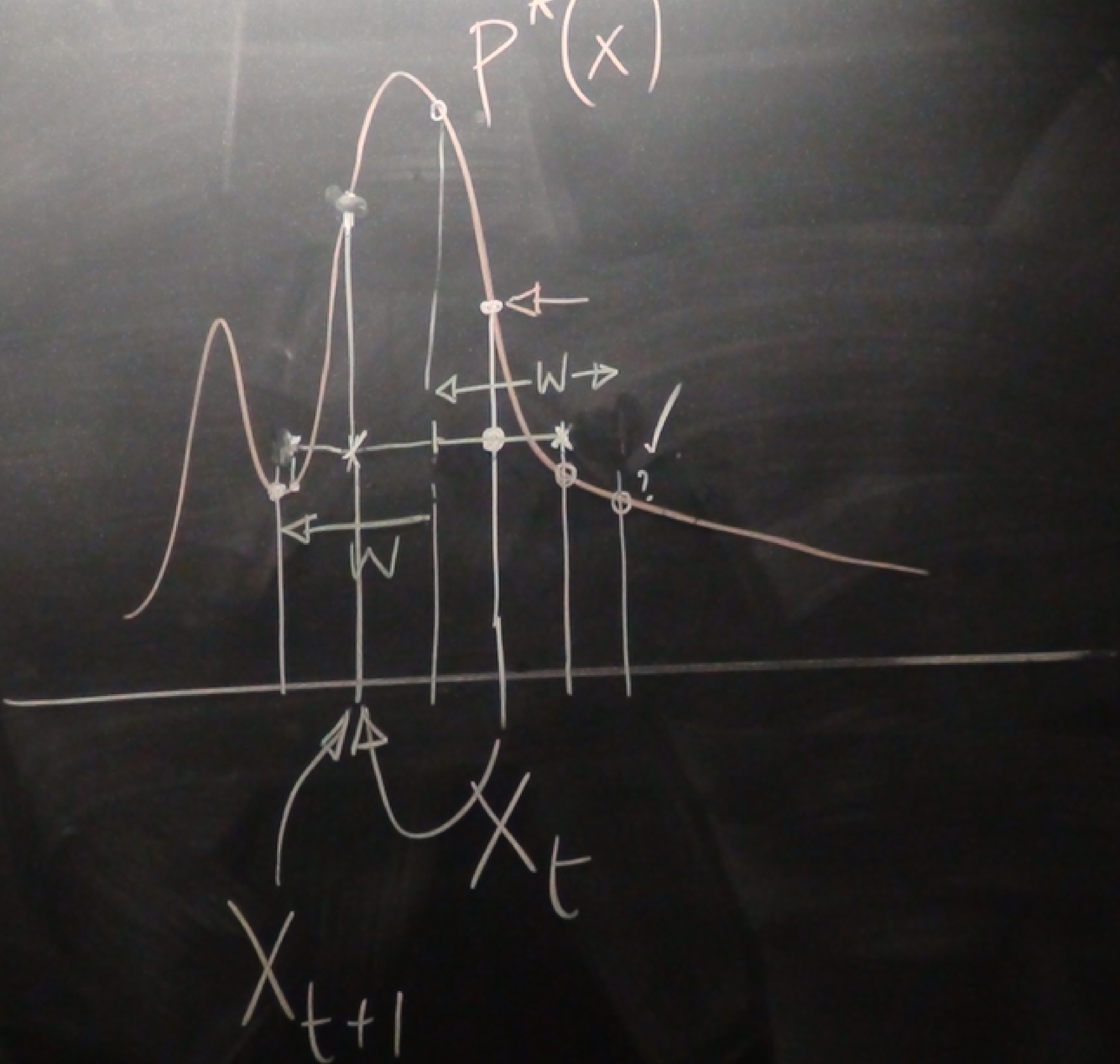
Radford Neal

John Skilling

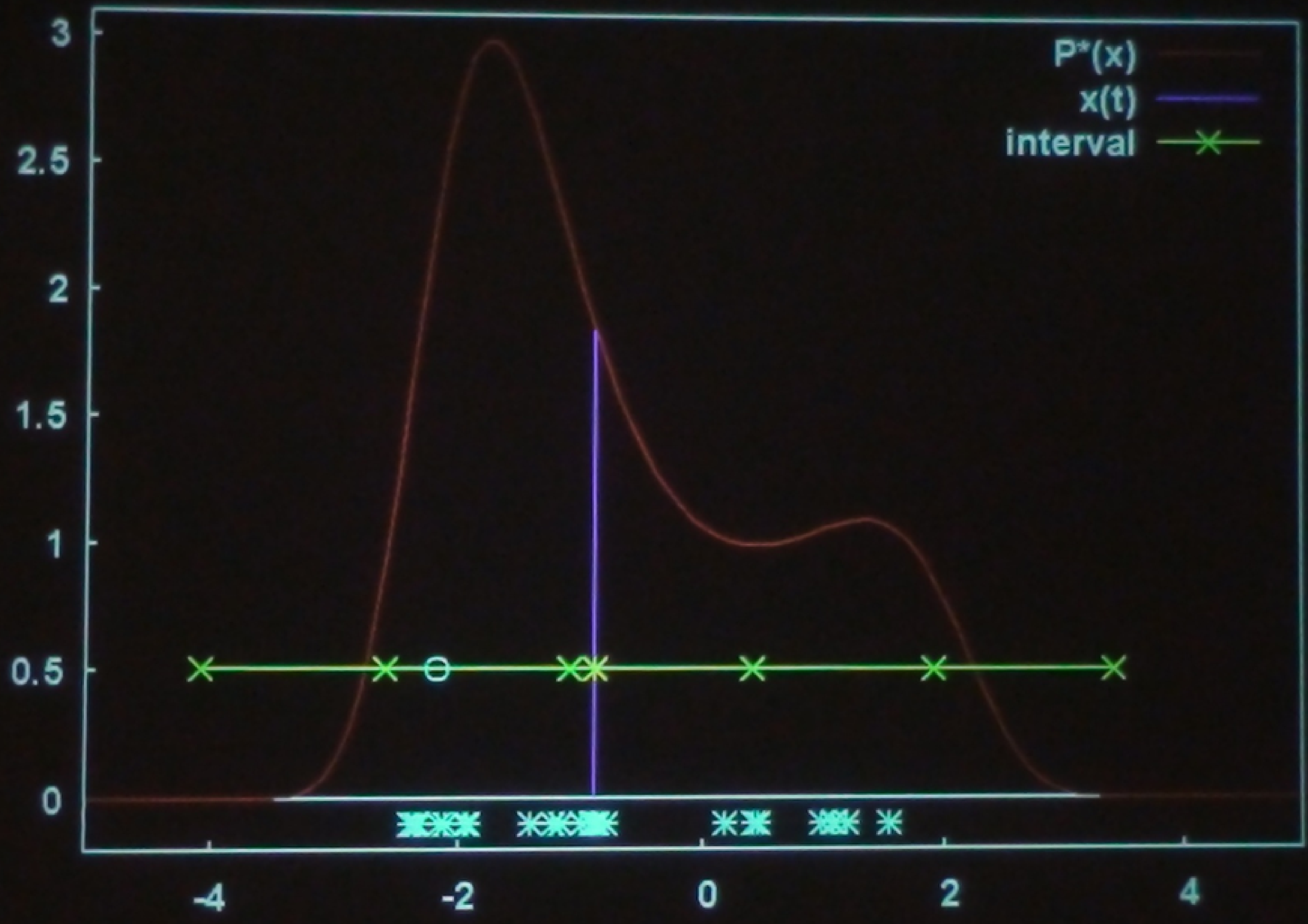

1 samples
P*(x)
x(t)
+
3
2.5
2
1.5
1
0.5
0
-4
-2
0
2
4



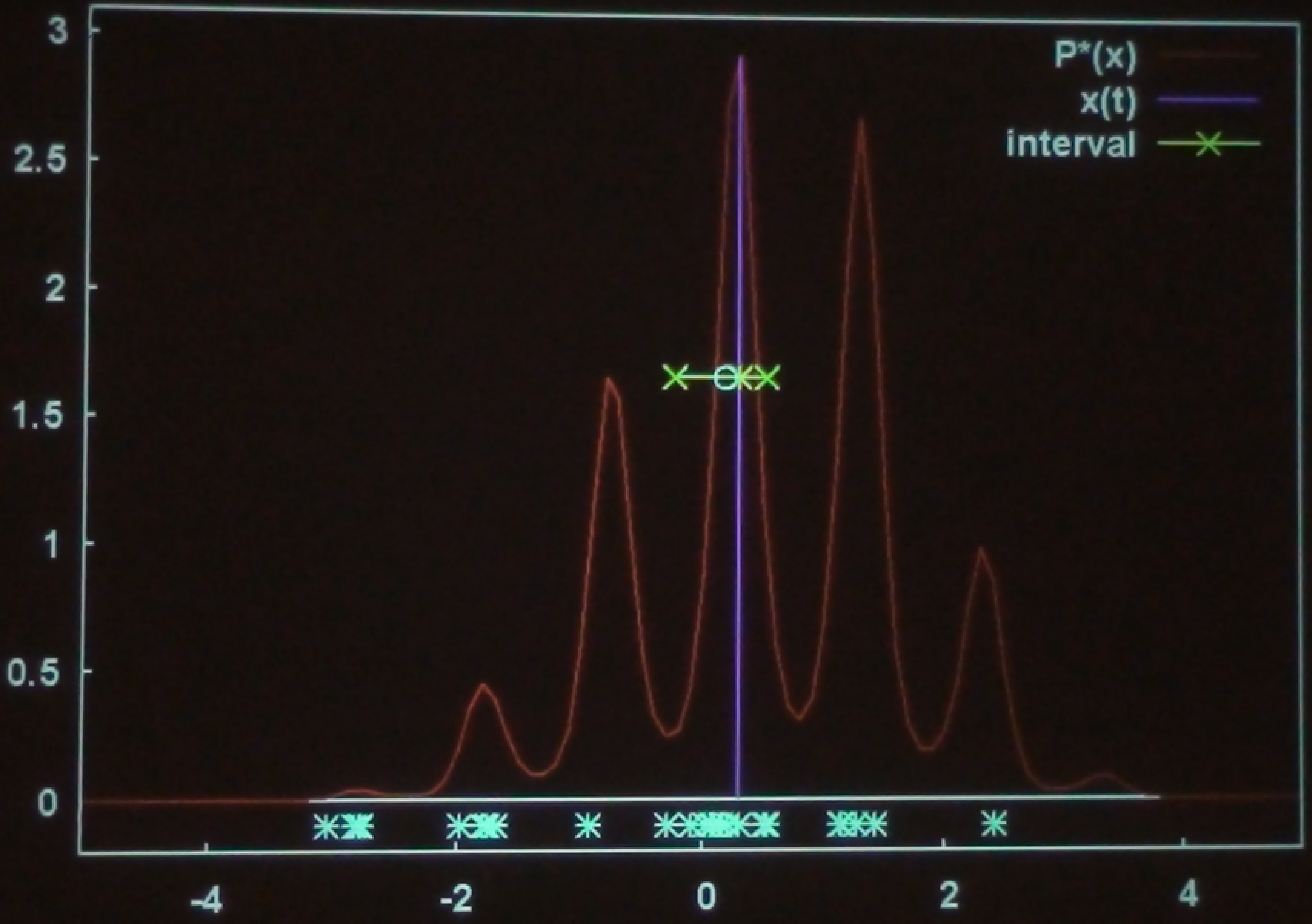
X_t


$P^*(x)$
w
w
?
✓
x_t
x_{t+1}

25 samples



25 samples

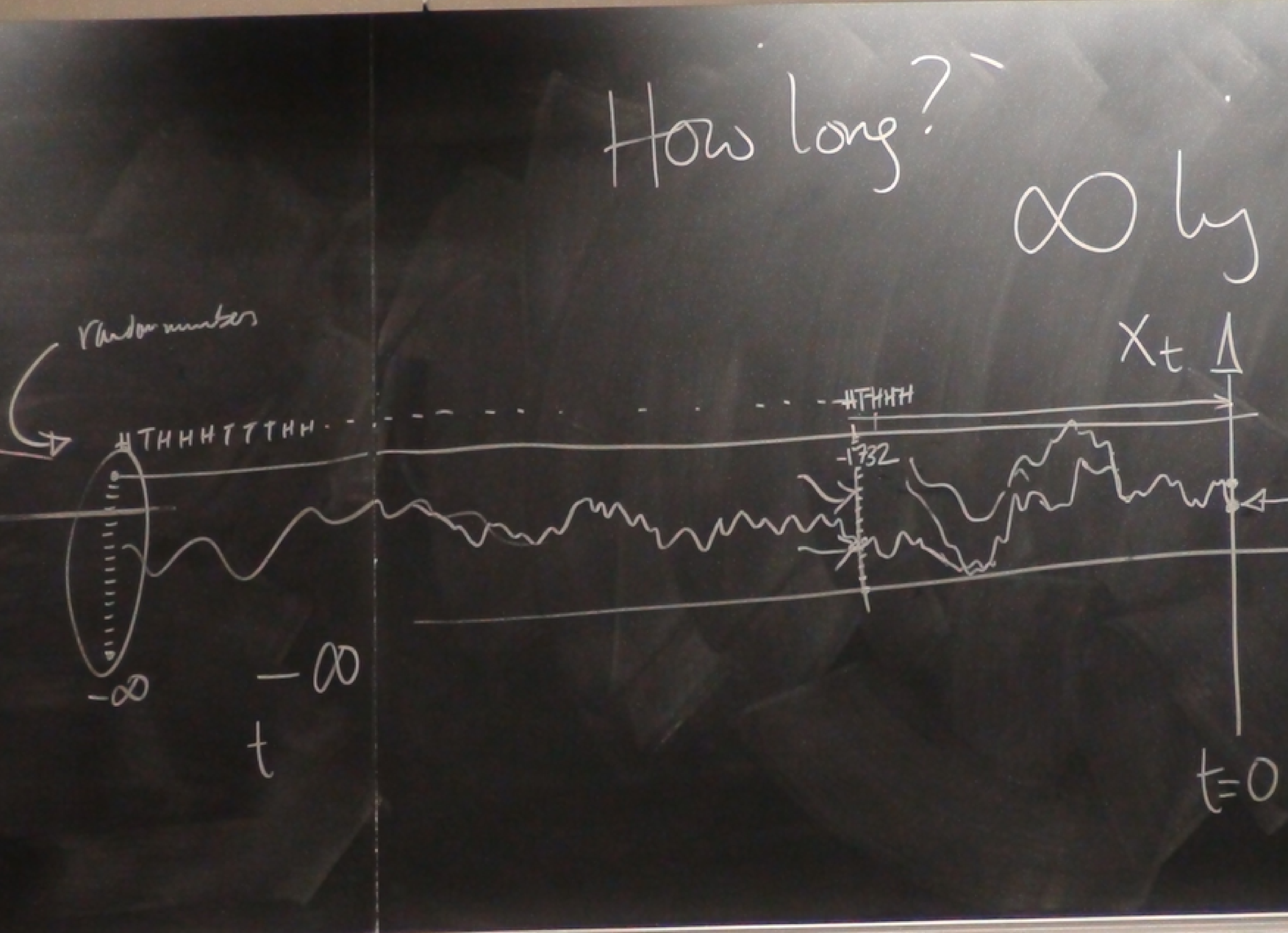


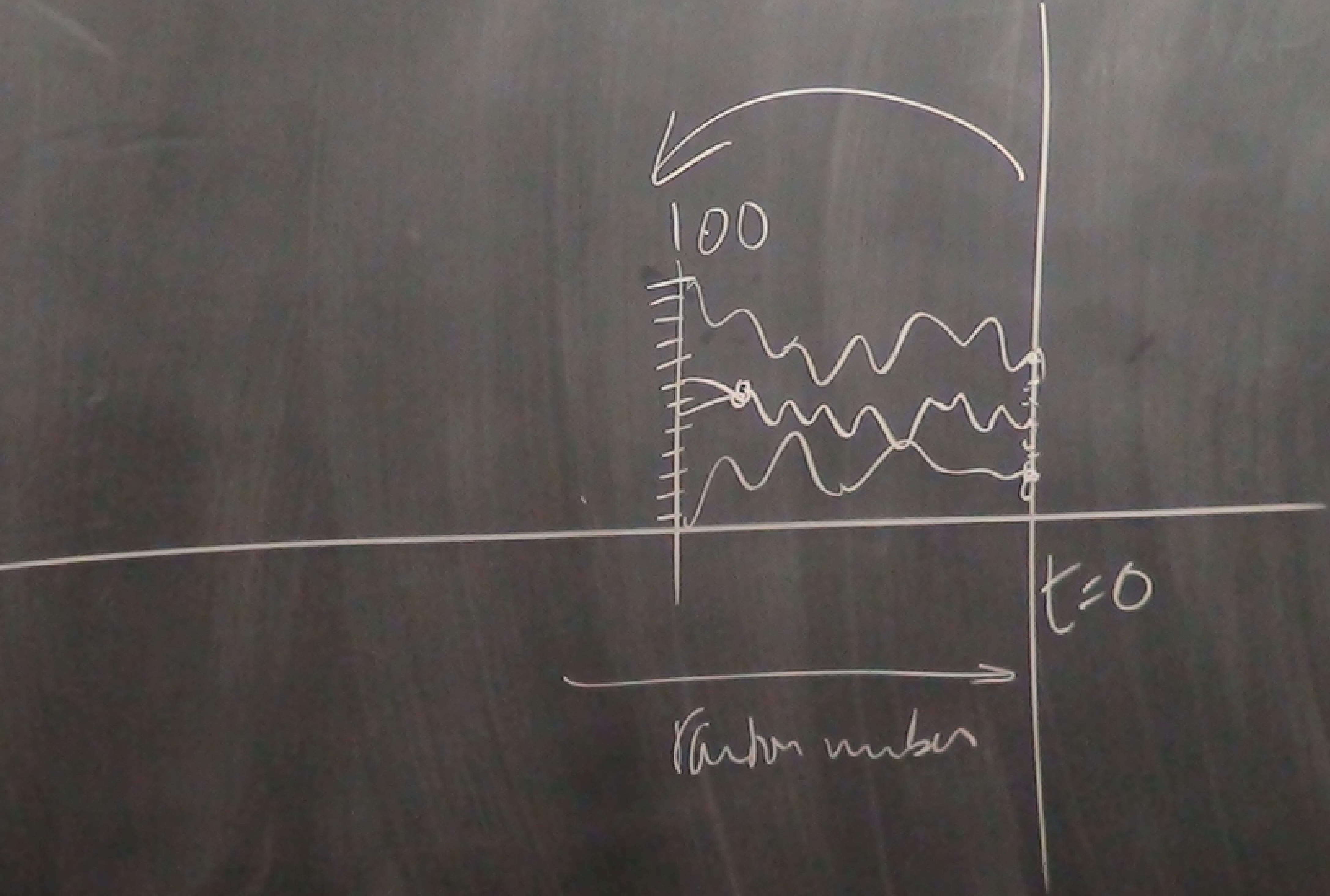
Self-terminating Monte Carlo methods

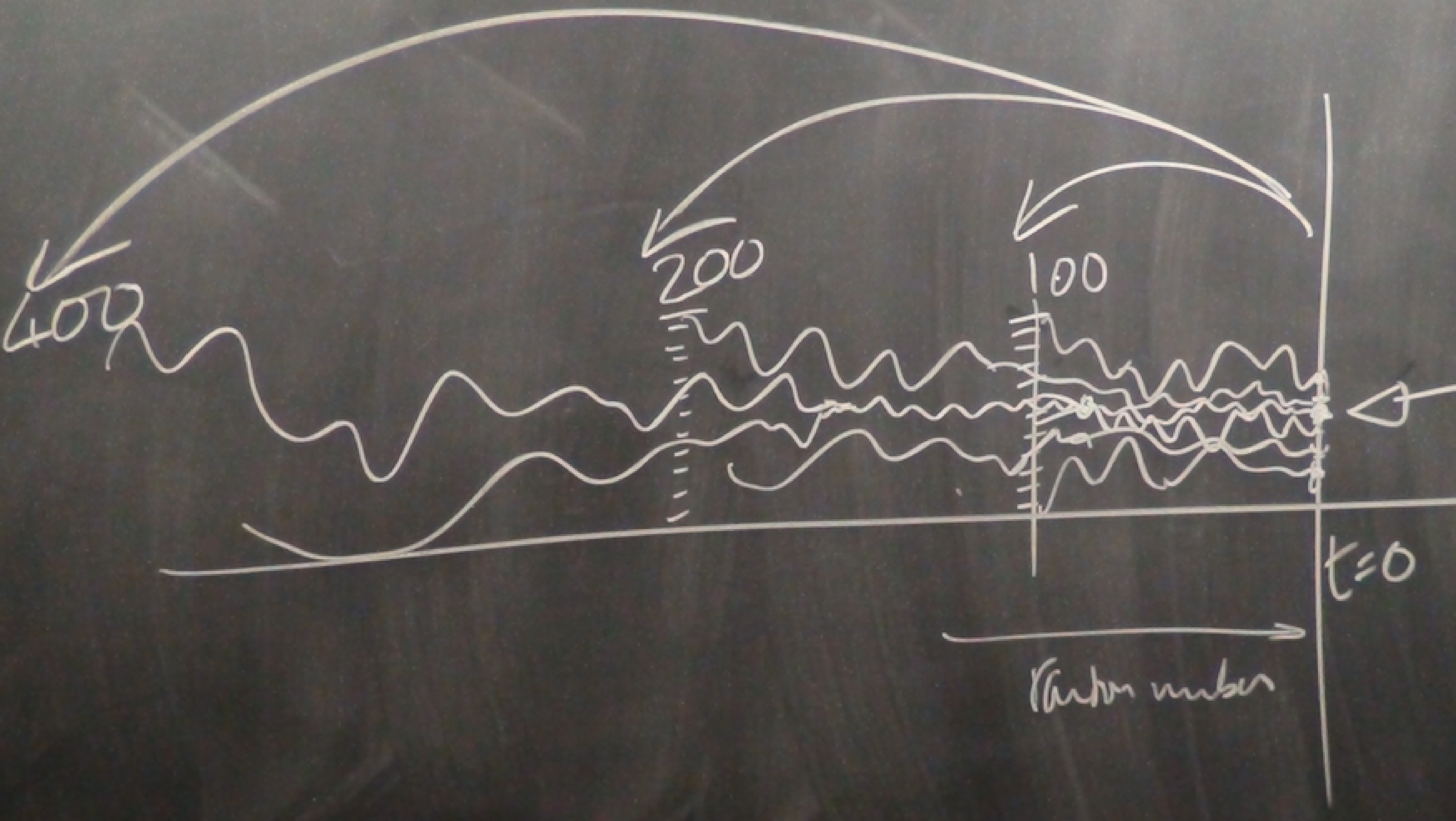
Exact sampling

How long?

∞ly long.

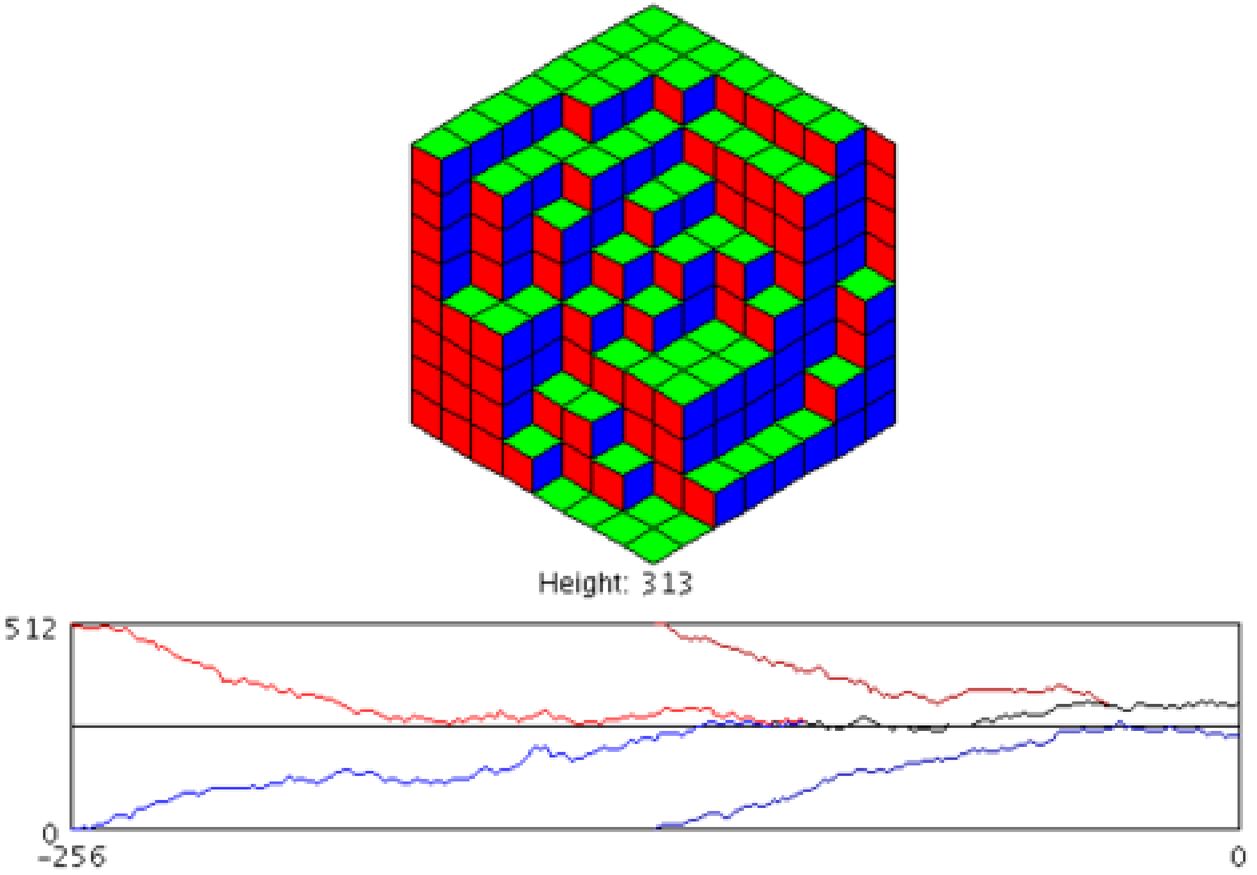

How long?
∞ ly
random numbers
HTHHHTTTHH
HTHHH
-1732
X_t
-∞
-∞
t
t=0


100
t=0
random number

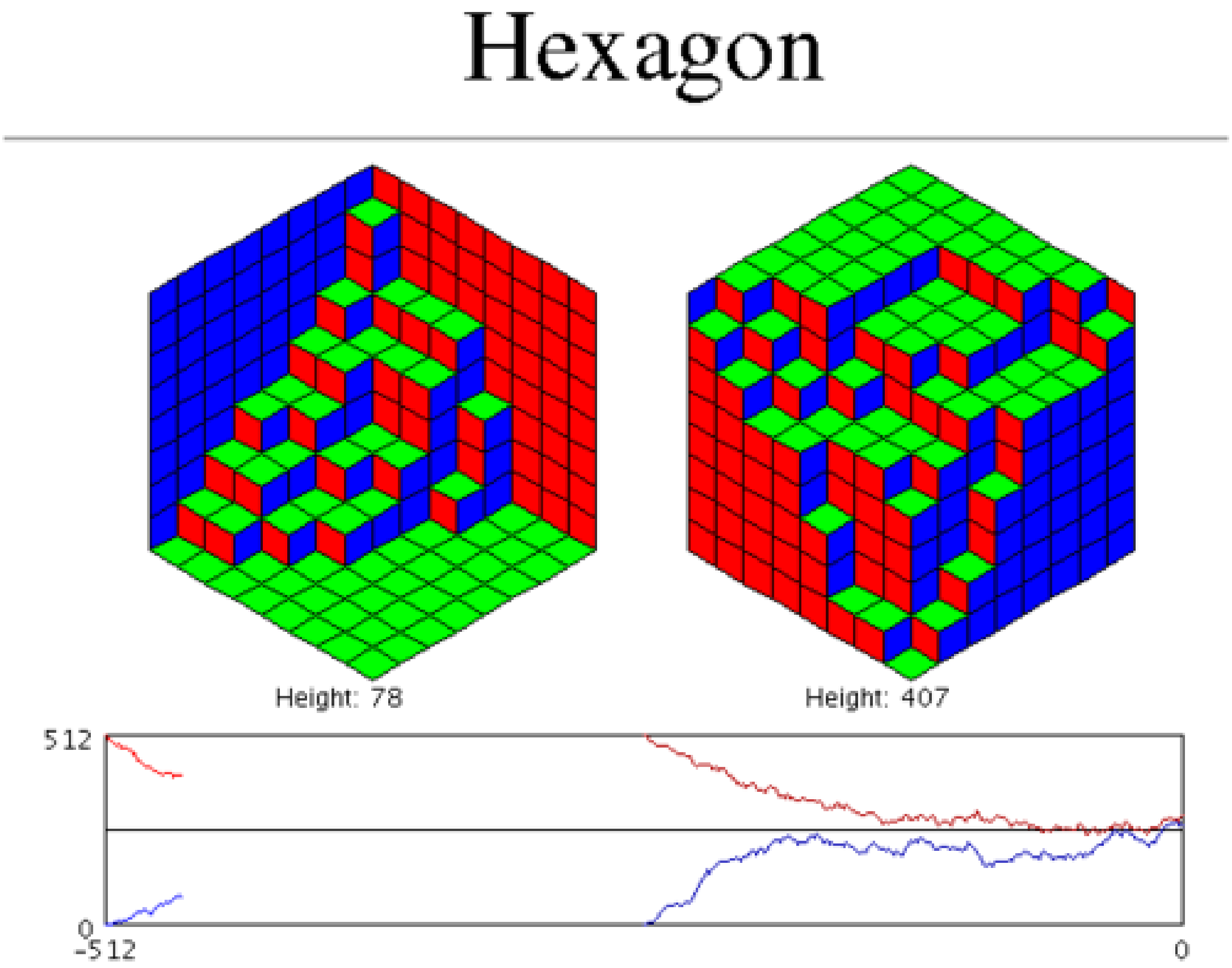

400
200
100
t=0
random number

Exact sampling

Hexagon

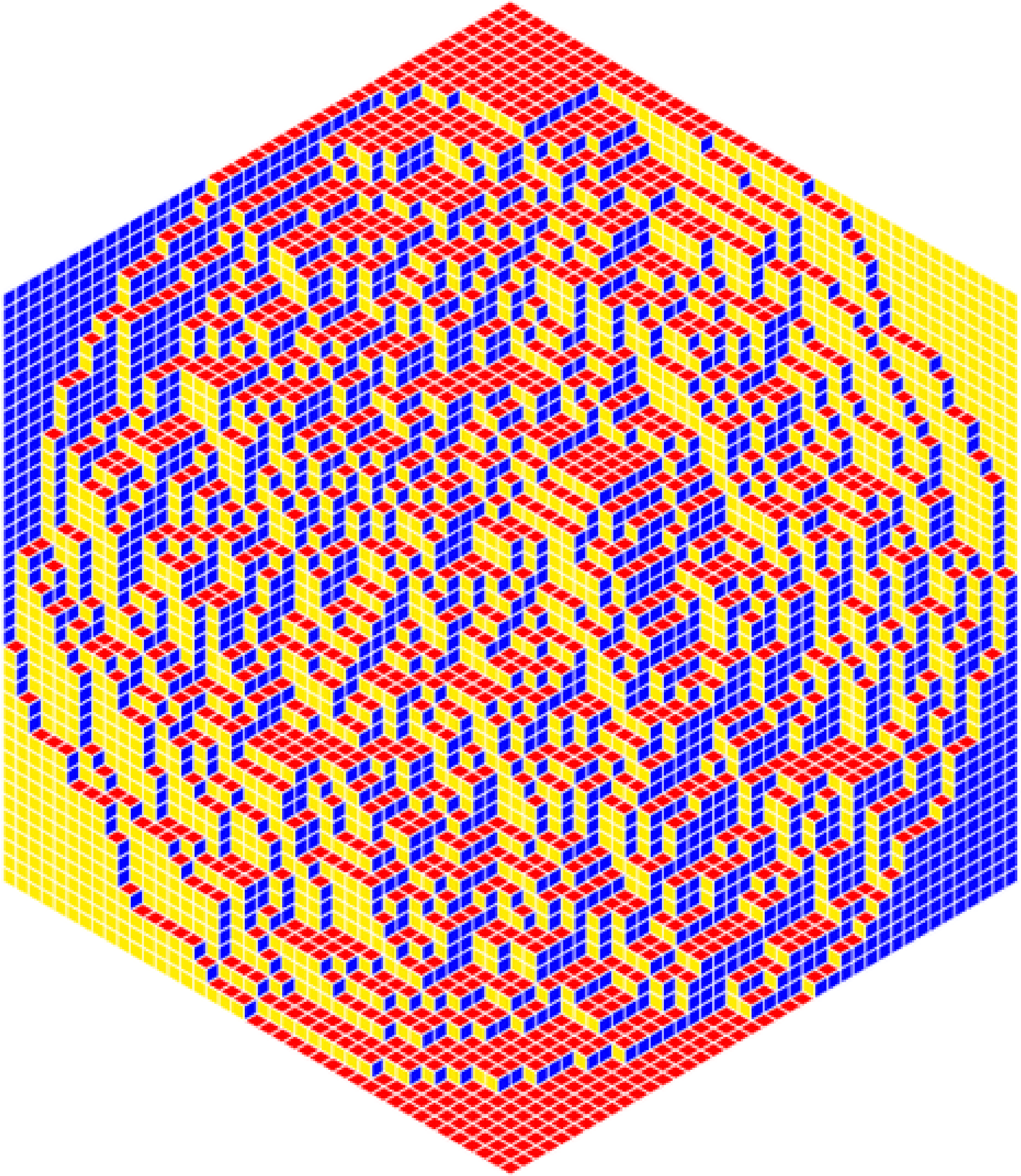


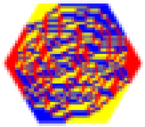
Exact sampling

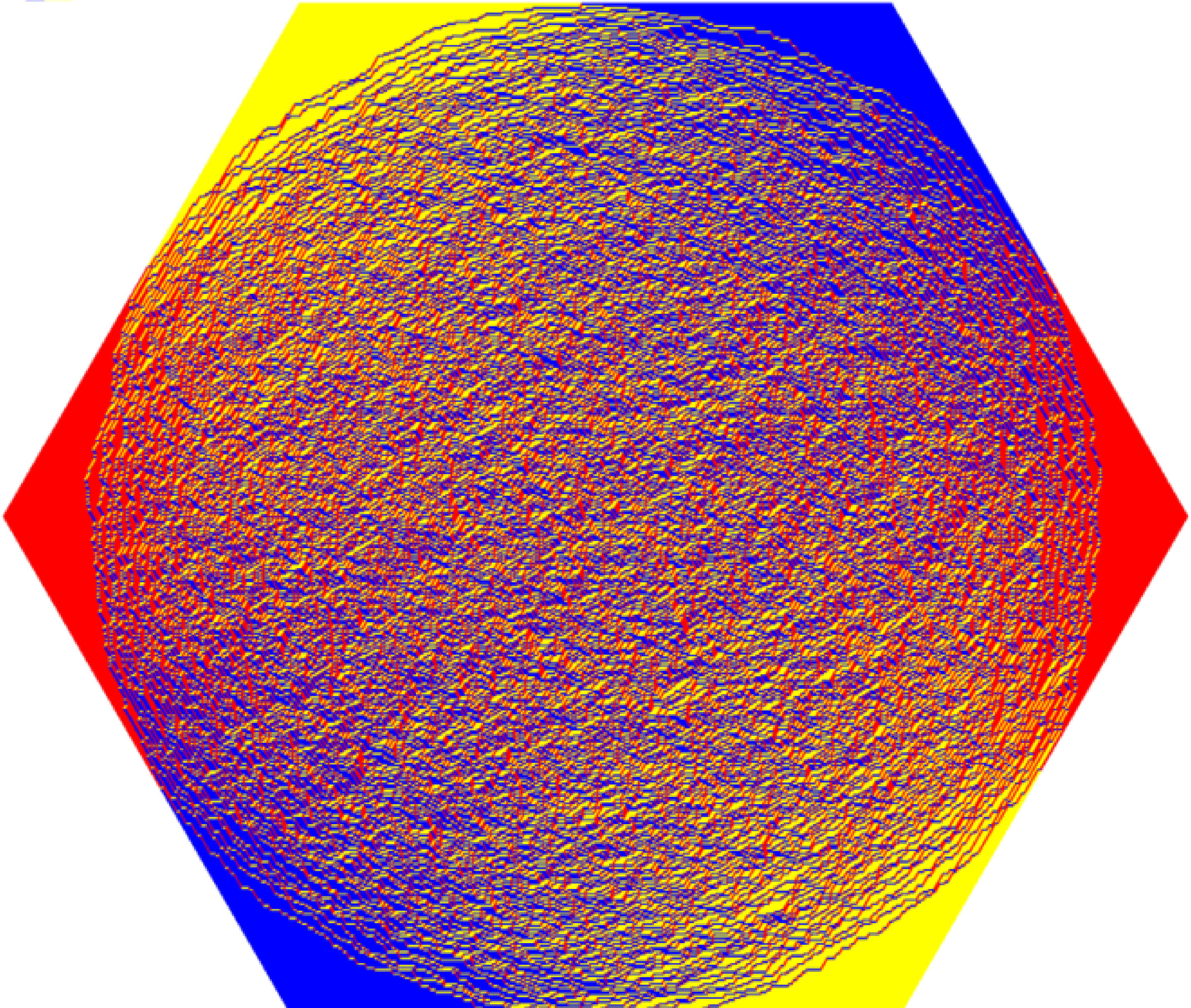


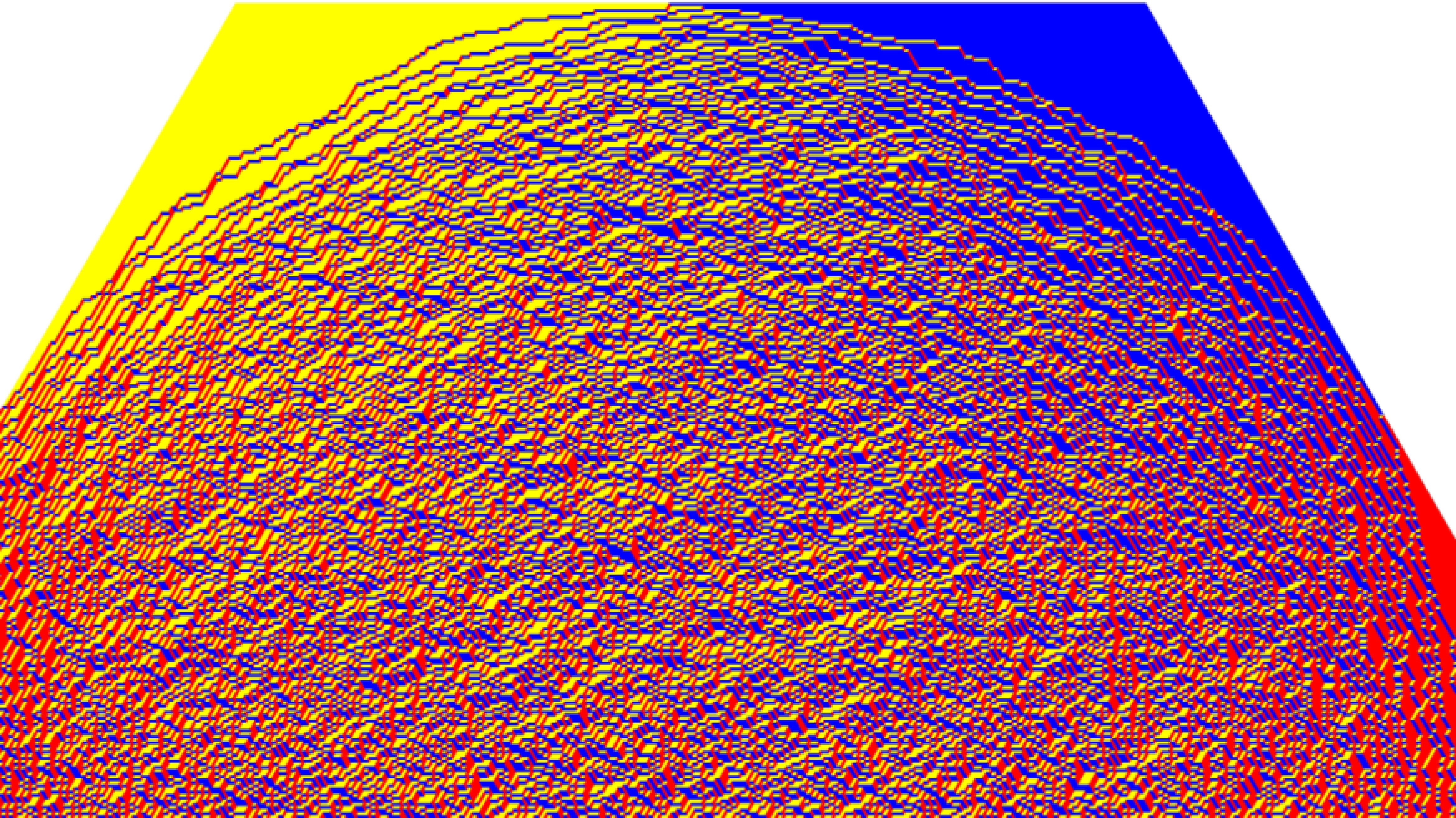
Hexagon

Exact sampling



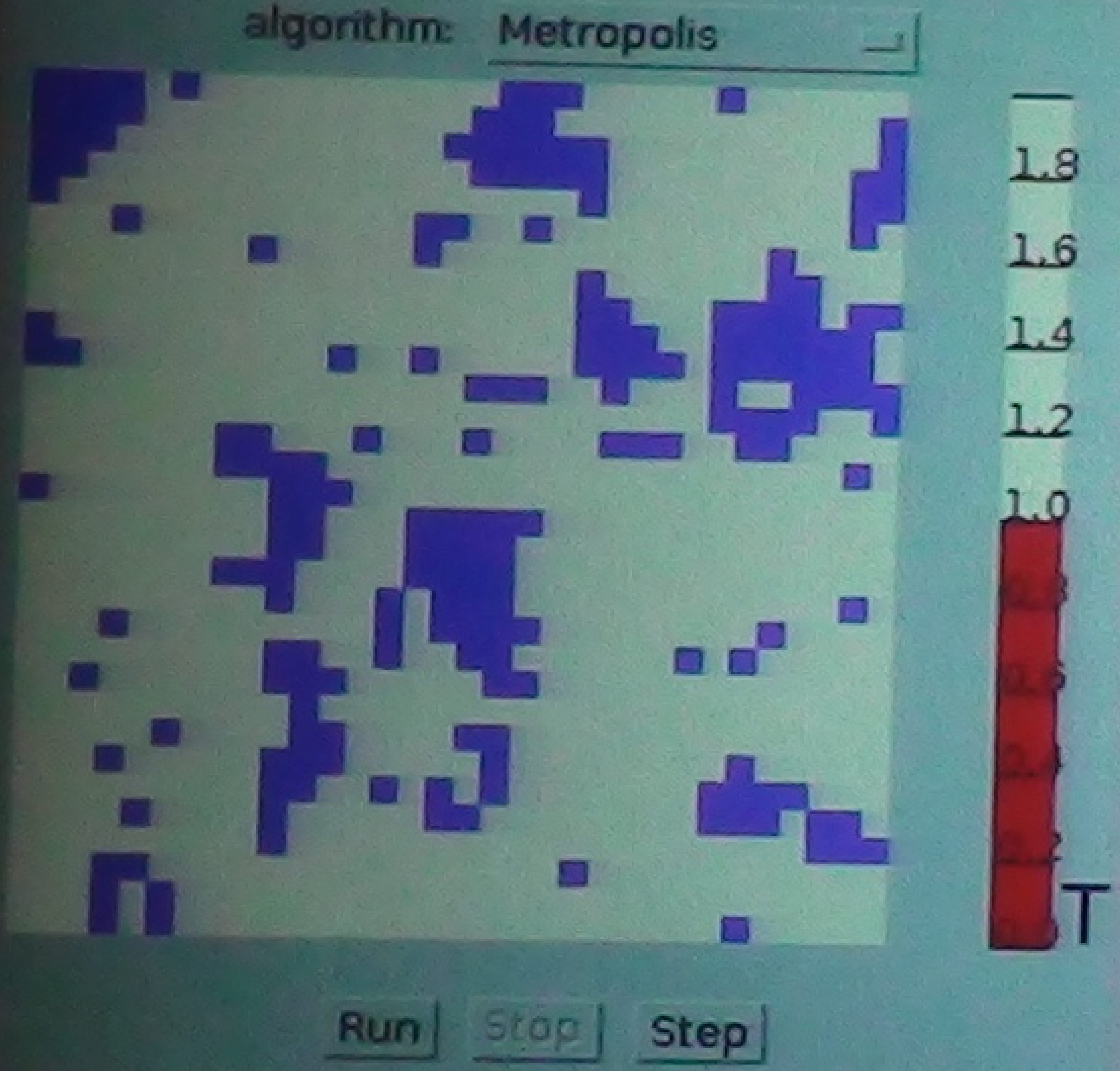






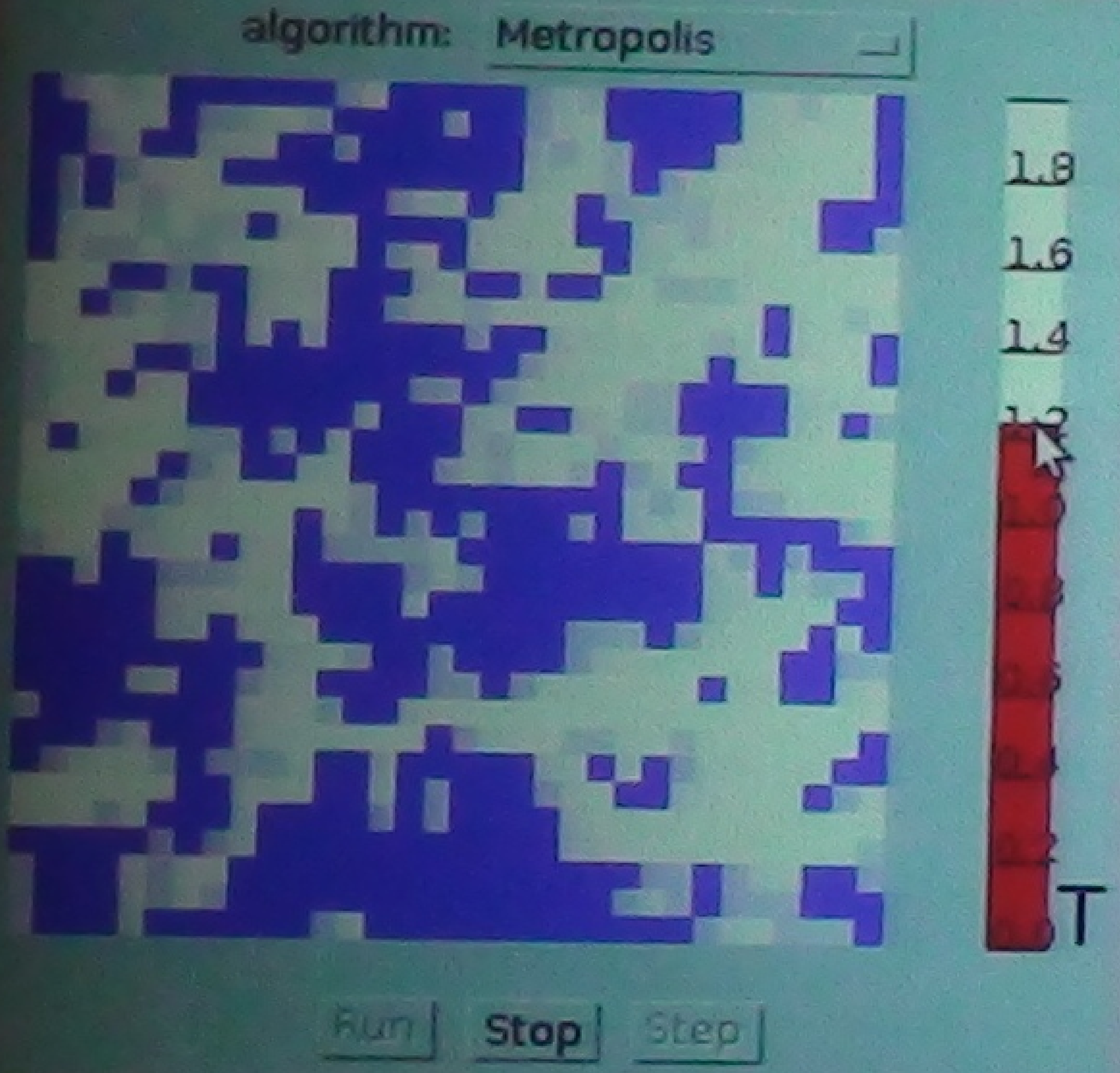
Exact sampling for the Ising model

controlled by the red bar, the spins prefer to be parallel.



algorithm: Metropolis

In the Delft computational physics group, Ising simulations controlled by the red bar, the spins prefer to be para



algorithm: Metropolis

Introduction to the Ising model

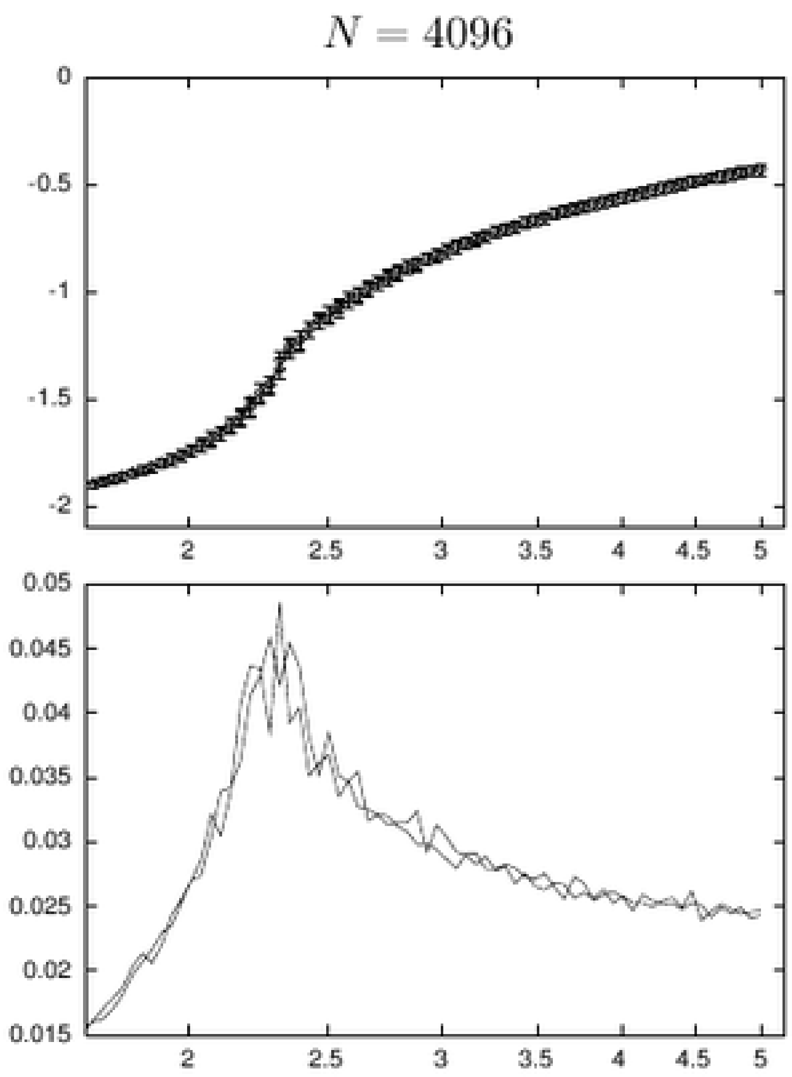
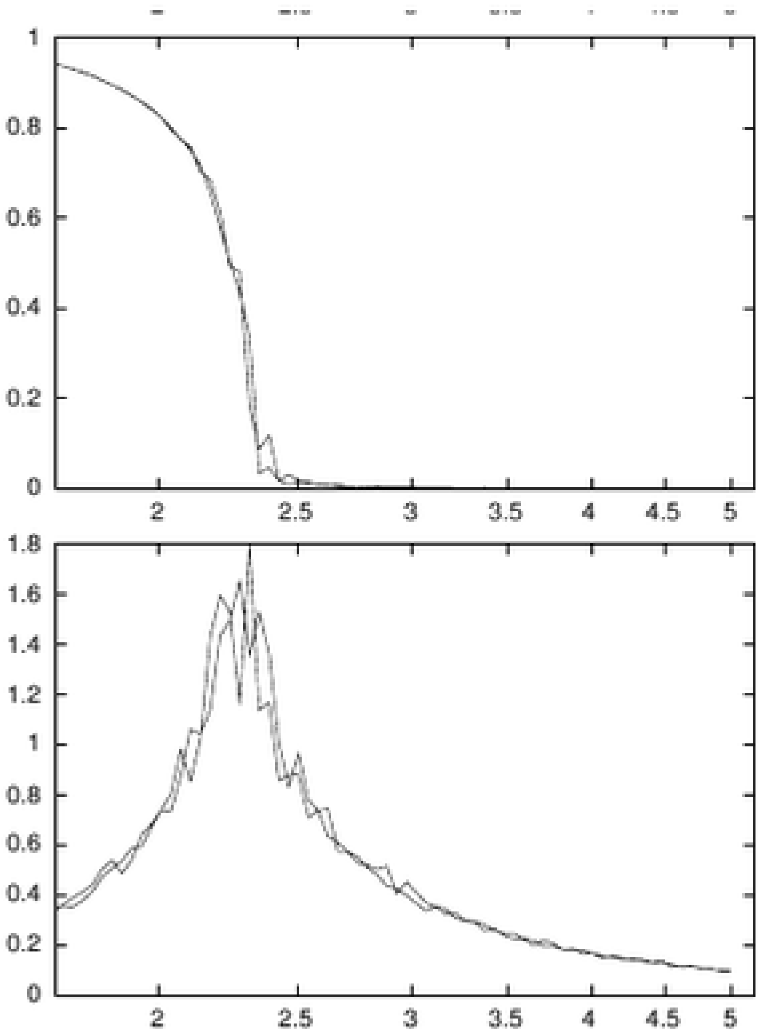
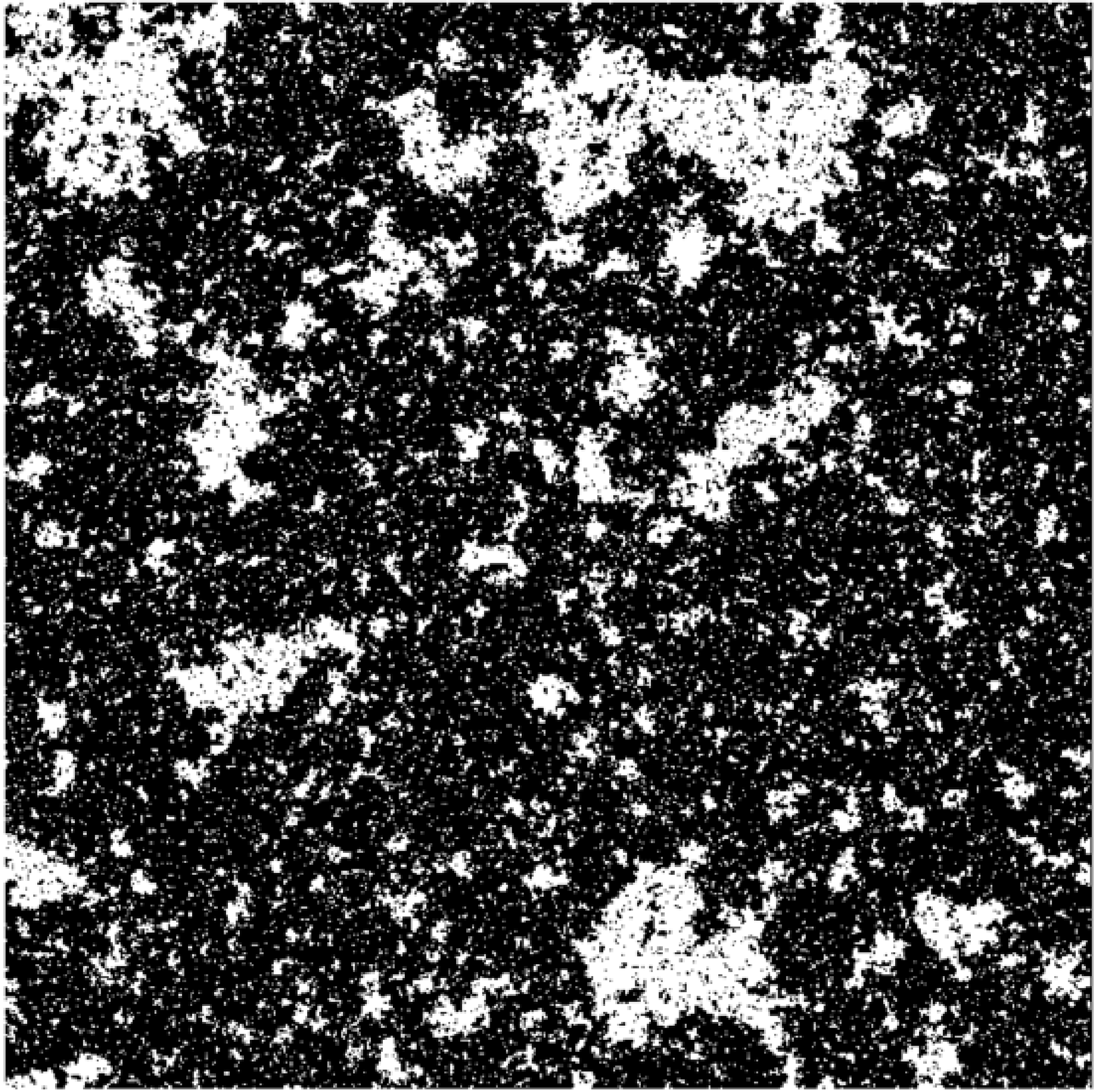


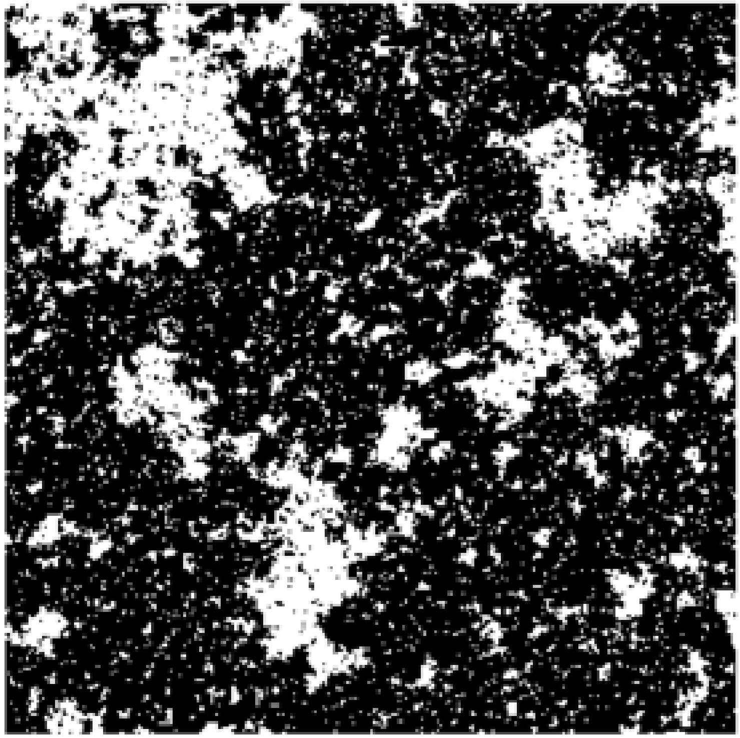
Figure 31.4. Detail of Monte Carlo simulations of rectangular Ising models with $J = 1$. (a) Mean energy and fluctuations in energy as a function of temperature. (b) Fluctuations in energy (standard deviation). (c) Mean square magnetization. (d) Heat capacity.

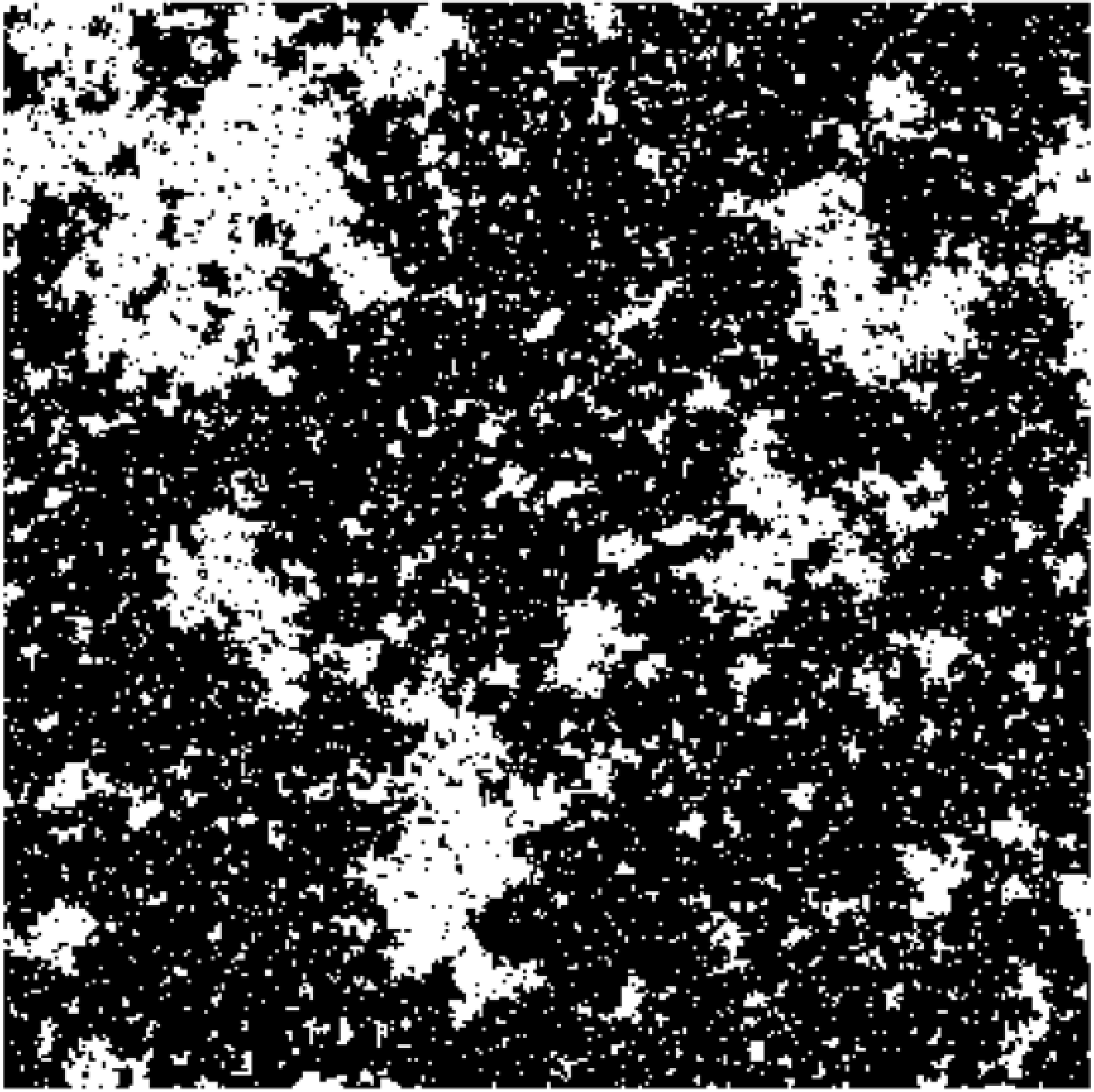


Exact sampling - Ising model at Tc

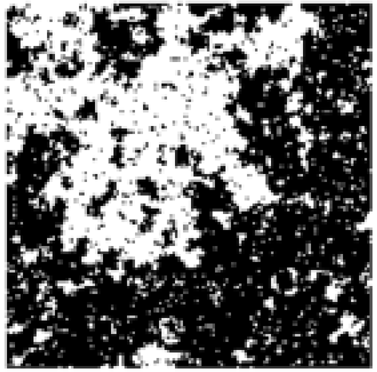


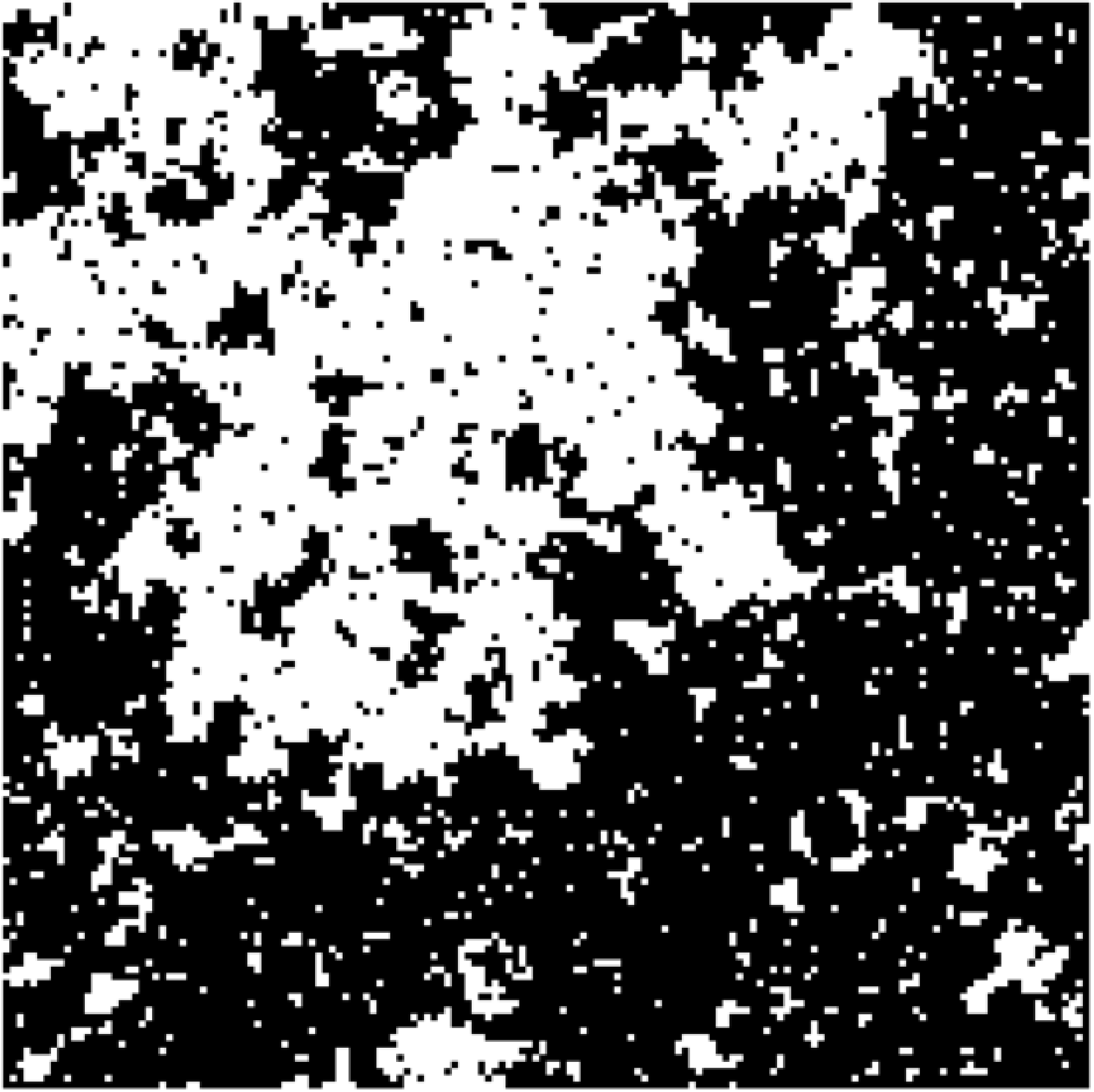
Exact sampling - Ising model at Tc



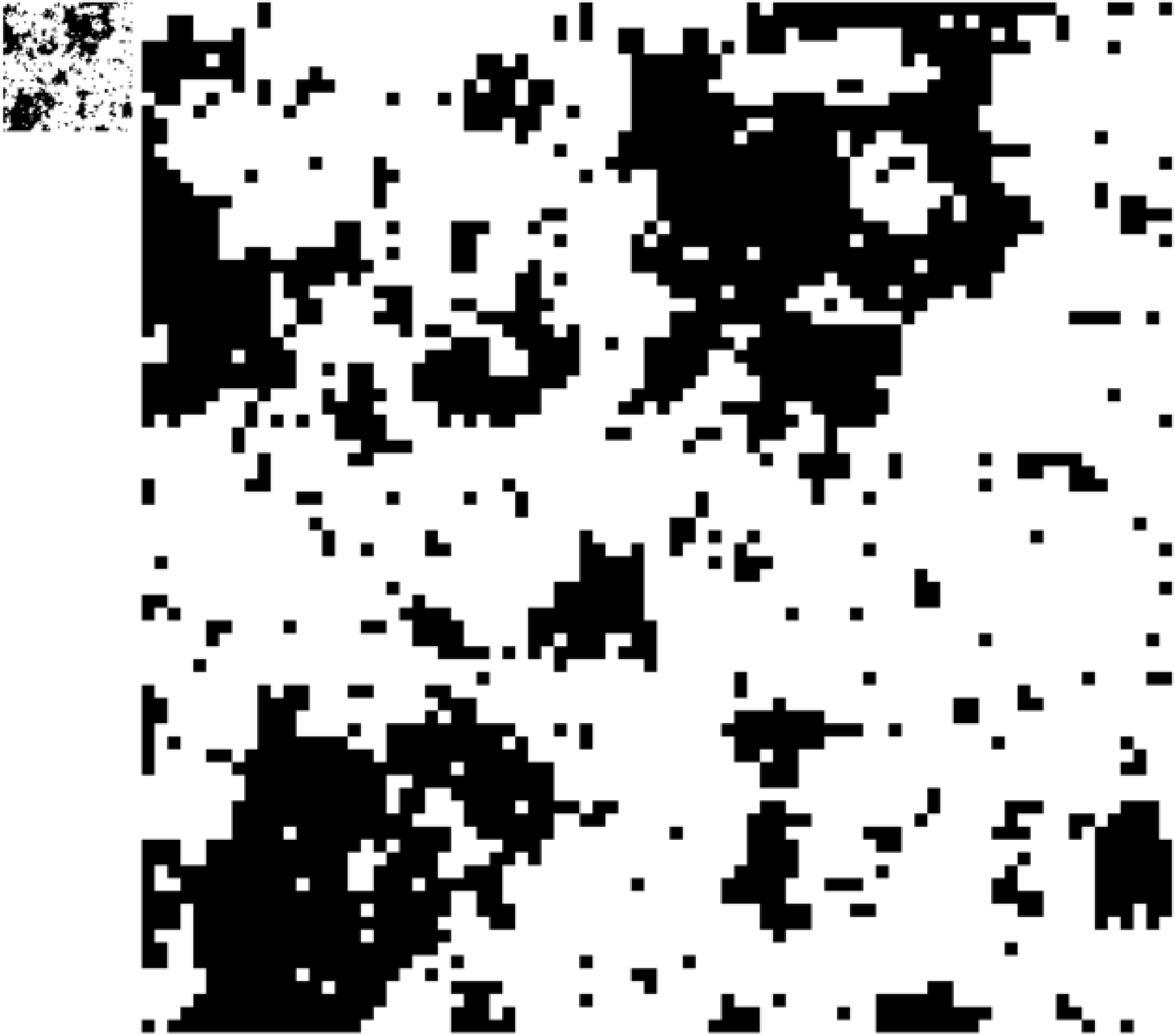


Exact sampling - Ising model at Tc





Exact sampling - Ising model at Tc



Problems with standard Monte Carlo methods

- Random walk behaviour

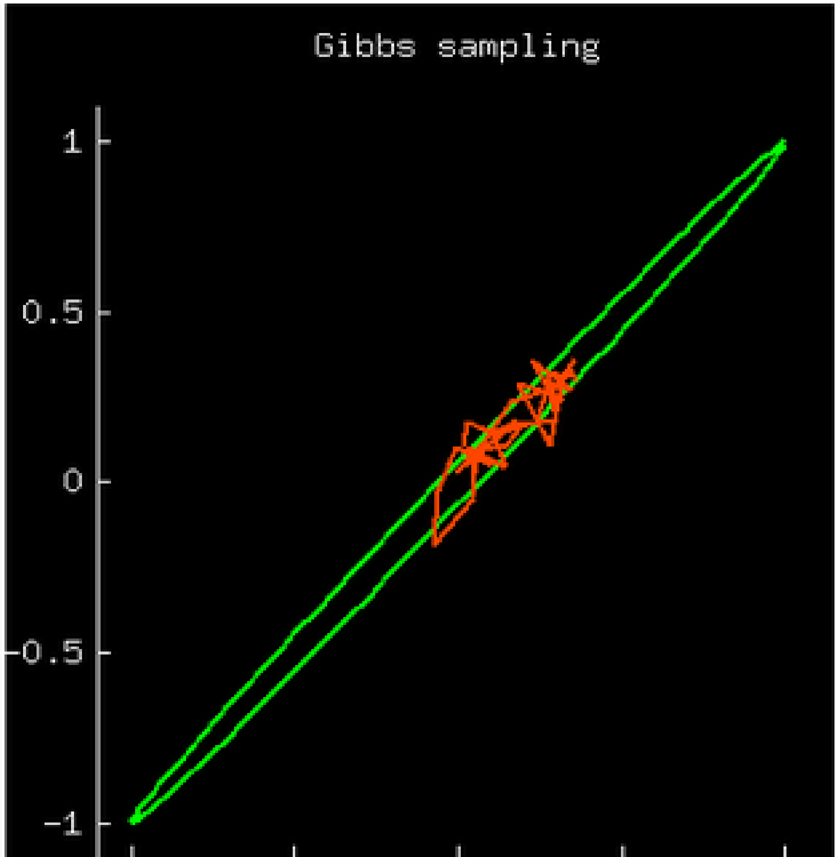
Efficient methods

- Sensitivity to step size

Slice sampling

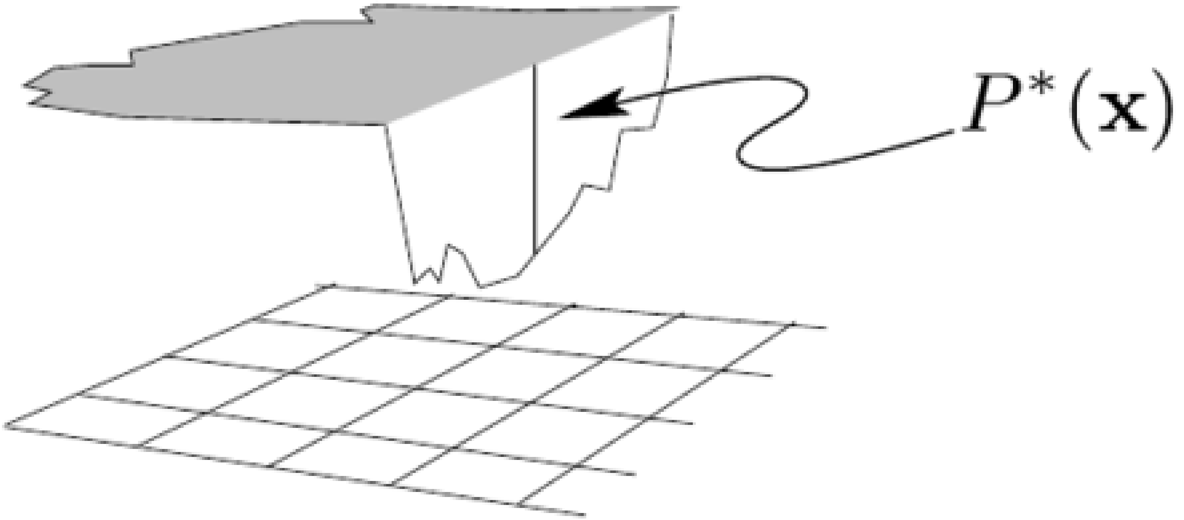
- When to stop

Exact sampling



- Not revealing the normalizing constant

  Thermodynamic integration
  Reversible-jump Markov chain Monte Carlo
  The acceptance ratio method
  Umbrella sampling
  Simulated tempering
  Tempered transitions (Radford Neal)
  Annealed importance sampling (Radford Neal)
  Linked importance sampling (Radford Neal)



$P^*(\mathbf{x})$

g

d Neal

Skilling